

# A Bilateral Filter Based Post-Processing Approach for Supervised Spectral-Spatial Hyperspectral Image Classification

BUSHRA NAZ SOOMRO\*, LIANG XIAO\*\*, MUHAMMAD MOAZZAM JAWAID\*, SHAHZAD HYDER SOOMRO\*, AND NAVEED AHMED JAFFARI\*

RECEIVED ON 02.08.2017 ACCEPTED ON 12.02.2018

## ABSTRACT

To effectively improve the performance of representation based classifier, a spatial spectral joint classification post-processing approach is proposed, based on the application of edge preserving BF (Bilateral Filtering) method. The proposed framework includes two key processes: (1) the classifier (such as SRC, CRC, or KSRC) based on sparse representation of each pixel is used to obtain soft-classified probabilities belonging to each information class for each pixel; (2) spatial spectral joint BF for the soft-classified probabilities map. It is aimed to integrate context-aware information for each pixel class labels. Under the spatial guidance image, extracted from the three principle component, a BF is employed to get the refined probability maps. The BF considers not only the spatial distance but it also considers the image context-aware distance which significantly improves the classification results. Finally, the class label is obtained by choosing the maximum probability criteria. The experimental results on three benchmark hyperspectral data sets showed that the “local smoothing” is efficient and has a potential to achieve high classification accuracy. All the algorithms are implemented with equal number of labeled samples and comparative results are presented in terms of visual classification map and numerical classification results. The major advantages of proposed method are: it is simple, non-iterative and easy to implement. Hence, the advantages lead to significant usage in real applications.

**Key Words:** Hyperspectral Image Classification, Bilateral Filter, Representation Based Classifier, Probability Optimization, Principal Component Analysis.

## 1. INTRODUCTION

The rich spectral information available in remotely sensed hyperspectral images allows for the possibility to distinguish between spectrally similar materials [1]. The use of hyperspectral images brings in new capabilities along with some difficulties in their processing and analysis. Although spectral (pixel-wise) classifiers have the advantages of conceptual simplicity and computational effectiveness, their limitations are also obvious. In order to

achieve higher classification accuracy, an increasing number of spatio-contextual analysis techniques have been developed recently to complement the spectral classification approaches. Accounting for spatial smoothness which: (1) reduces the mis-classification errors, (2) reveals the size and shape of the structure the pixel belongs to, and (3) allows the discrimination between structures made of the same materials, but belonging to different land-use types.

Authors E-Mail: (bushra.naz@faculty.muet.edu.pk, xiaoliang@mail.njust.edu.cn, moazzam.jawaid@faculty.muet.edu.pk, shahzadhyder95@yahoo.com, naveed.faffari@faculty.muet.edu.ok)

\* Department of Computer Systems Engineering, Mehran University of Engineering & Technology, Jamshoro.

\*\* School of Computer Science & Engineering, Nanjing University of Science & Technology, China.

Generally, the spectral-spatial classification methods can be divided into following categories.

- (1) Pre-processing methods are the first category. In these methods the dimensionality reduction techniques are adopted initially which produces the image attributes in the spectral region to bring the information in the original space (which in the case of hyperspectral data is almost empty) to the right subspace which allows separating the classes by discarding information that is useless for classification purposes. The most applicable methods are PCA (Principal Component Analysis) [2] and KPCA (Kernel Principal Component Analysis) [3]. And then spatial feature extraction techniques such as morphological profiles (EMPs) [2] and EMAPs (Extended Multi-Attribute Profiles) [4] are applied to produce the spatial profiles. Once these spatial-spectral features are obtained, benefitted with the CKs (Composite Kernels) [5], GCKs (Generalized CKs) [6], or MFL (Multiple Feature Learning) [7], these spectral-spatial features are integrated.
- (2) Second group belongs to the post-processing approach, in which first pixel-wise spectral only classifier generates the classification map. And then by adding the spatial content the classification maps are post-processed to get the accurate classification results. The most common post-processing methods are majority voting [8], MRF (Markov Random Field) [9, 10], graph regularization [11] or ERWs (Extended Random Walkers) [12].

For HIS (Humane Society International), to collect the training samples expensive human labor is required and mostly we have limited training samples for experiment.

Thus, to lighten this problem and further improve the classification accuracy, spectral-spatial classification post processing approaches are under consideration [13]. Benediktsson et. al. [14] adopted the EMPsto integrate the spectral-spatial attributes via the adaptive neighborhood. Moreover, nowadays kernel based spectral-spatial features, e.g. composite kernels in [8], morphological kernels [15], and graphic kernels [16] have been adopted to improve the SVM (Support Vector Machines) pixel-wise classifier accuracy.

Among them the kernel-based and morphological approaches results showed the good accuracies to classify HSI [5,14]. In our previous work [11], a graph based spectral and spatial self-similarity of local and non-local neighbors using the ELN (Elastic Net)-coding neighborhood graph is incorporated. Furthermore, Kang et. al. [17] a two stage post-processing classification method based on ERWs is proposed. Initially it adopted a SVM for the pixel-wise classification probability maps. And then the probability maps are optimized with ERWs by using weighted graph to add the spatial information. Zhang et. al. [18], a graph based regularized post-processing algorithm for a KSRC-based HSI classifier is introduced, which exploits the spatial and spectral information. Xu et. al. [19], we have post-processed the ELN-regularized multinomial logistic regression with the gradient profile ELN-based MRF prior.

A simple and well known approach to classify pixel entities using the spatial and the spectral information is the filtering based post-processing approach. In this approach simply an adaptive filters or window are applied to the spectral bands. In HSI filtering requires more intensive work than in multispectral images. It is a simple and effective way to regularize for spatial smoothness to meliorate the input space with features accounting for the neighborhood of the pixels. For HSI classification the BF is well suited, because it performs the weighted average of the neighborhood samples. BF is a non-linear filter and it considers the spectral

and spatial distances between the pixels, which enables the BF to preserve the image details. The spatial variability presents due to the noise can be smoothed out with the help of BF. Benefited with the weighted averaging approach of BF, for the post-processing step it is well suited to add the geometric information with the pixel-wise spectral-only classifier. Previous research [9,11,19] have indicated that post-processing classification is an important step in improving the quality of classifiers. In last decades' researchers have been proposed the different kinds of EPFs, i.e. Joint bilateral [20], weighted least square [21], guided [22], domain transform [23], local linear Steins unbiased risk estimate [24], and  $\ell_0$ -gradient [25] filters. All filters have the joint filtering approach which smoothen the given image on the basis of a guidance image. During the filtering process the spatial information is well considered. Moreover, nowadays image processing researchers keenly noticed the EPF (Edge Preserving Filtering) [20-22]. A lots of applications in image processing field is introduced such as high dynamic imaging [21], stereo matching [26], image fusion [27,28], enhancing [29], de-hazing [30], and de-noising [31] have been developed. Furthermore, for HSI visualization [32] and classification [33] EPFs drawn the remarkable results.

A BF is a non-linear, edge-preserving and noise-reducing smoothing filter for images. The intensity value at each pixel in an image is replaced by a weighted average of intensity values from nearby pixels. The BF is defined as a weighted average of nearby pixels which takes into account the difference in value with the neighbors to preserve edges while smoothing. The key idea of the BF is that for a pixel to influence another pixel, it should not only occupy a nearby location but also have a similar value. The technique of joint BF is as an extension of the BF [20]. This is a traditional edge-preserving approach while smoothing. Compare to the Guassian filter, the BF is embedded with a similarity function between the central pixel (where the filter is applied) and its neighborhoods (used in blurring), and a function of intensity value difference with the neighborhoods.

Considering the above description, the application that we are interested in this paper is post-processing supervised classification framework. Keeping such a view, we proposed a BF based edge-preserving spectral-spatial classification model depicted in Fig. 1. Our proposed method includes three main steps. For the initial classification representation based pixel-wise classifier (i.e. SRC, CRC, or KSRC) is first represented by the multiple of probability maps. Benefitted with the spatial guidance image obtained from the very well-known PCA (i.e. the first three principal components), the obtained probability map is further post-processed with edge-preserving BF. At the end the optimized spatio-spectral classification is obtained which approaches the maximum probability scheme. Considering the neighborhood information, BF not only smooths the probabilities but it also ascertaining the alignment of authentic object boundaries of the smoothed probabilities. By experimental analysis it is proved that the classification accuracy of the spectral only representation based classifier is improved. Because of the less computer resources requirements the proposed model will be very helpful for the real applications.

The remnant paper is arranged in the following manner. The general model for representation based classification model is detailed in section 2. The details of the proposed two stage post-processing framework are presented in Section 3. Section 4 gives the detail of the comparative experimental analysis in terms of classification accuracies and influence of the key parameters. Finally, the paper is concluded with remarkable comments in section 5.

## **2. GENERAL MODEL FOR REPRESENTATION BASED CLASSIFICATION**

Representation based classification is of great interest in the context of remote sensing data processing recently because of the CS (Compressed Sensing) [34]. These approaches intend to represent most observations or image

pixels with linear combinations of a small number of elementary (dictionary) samples, known as atoms, which are chosen from an over complete labeled dictionary. Formally, an over complete dictionary is a collection of atoms, such that the number of atoms exceeds the dimension of the image space, and any image pixel can be represented by more than one combination of different atoms. Under given circumstances, an image pixel can be recovered with a minimal number of atoms by solving a representation regression problem.

Let  $X = \{x_i\}_{i=1}^N \in R^{L \times N}$  represents a HSI dataset, where L is showing the spectral bands having class labels  $k = \{1, 2, \dots, K\}$  (K represents the classes number). Let  $M_k$  represents the labeled samples of the  $k^{\text{th}}$  class, and  $\sum_{k=1}^K M_k = M$ . A representation based approximation for  $x_i$  is retrieved by combining all available sampled  $\Phi$  linearly which is also known as dictionary. The representation coding coefficient vector  $\hat{a}_i$  can be recovered from  $x_i$  and  $\Phi$  by solving the following optimization problem:

$$\hat{a}_i = \arg \min_{a_i} \frac{1}{2} \|x_i - \Phi a_i\|_2^2 + \lambda \|a_i\|_{\ell_q} \quad (1)$$

Where  $\lambda$  is a regularization parameter and by setting  $\ell_q = 1, 2$  different forms of representation based model can be obtained.

### 2.1 Sparse and Collaborative Representation Classification

For  $\ell_q = 1$ , SR (Sparse Representation) model is obtained while for  $\ell_q = 2$ , a CR (Collaborative Representation) model is obtained. For representation based classification, a successful classification is based on the weight vector (i.e.  $a_i$ ) that is discriminatively sufficient to reflect fully the particular characteristics of the classes (features).

In SRC, labeled data is used to sparsely represent the testing samples by  $\ell_2$ -norm regularization, thus the SRC can be optimized given as:

$$\arg \min_{a_i^{(SR)}} \|x_i - \Phi a_i^{(SR)}\|_2^2 + \lambda_1 \|a_i^{(SR)}\|_1 \quad (2)$$

For classification, after getting  $\hat{a}_i$ , the SRC computes the class label  $y_i$  of  $x_i$  directly with the obtained sparse coding vector  $\hat{a}_i$  by calculating the minimal residual between  $x_i$  and its approximation from each class sub-dictionary.

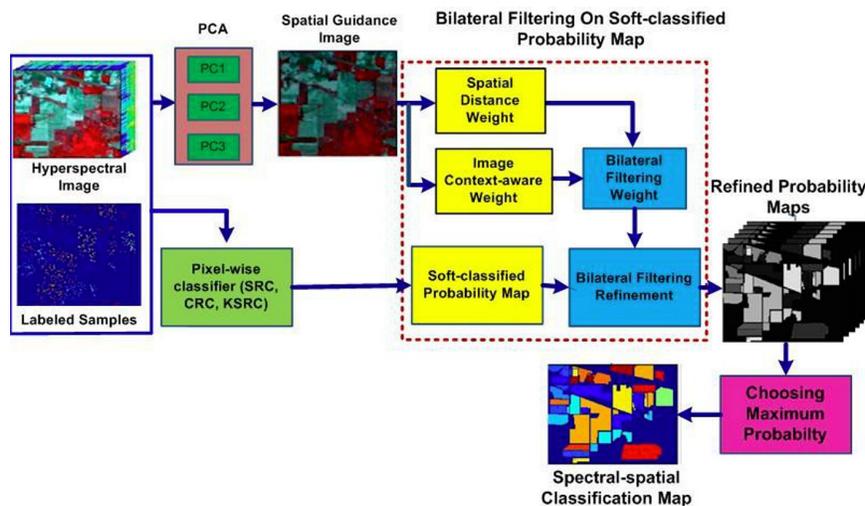


FIG. 1. TWO STAGE FRAMEWORK COMBINING PIXEL-WISE CLASSIFIER WITH POST-PROCESSING

$$r_k^{SRC}(x_i) = \left\| \Phi_k a_{ik}^{(SR)} - x_i \right\|_2 \quad (3)$$

and class label is given as:

$$SRC(x_i) = \arg \min_{k=1, \dots, K} r_k^{SRC}(x_i) \quad (4)$$

Although SRC(Collaborative Representation based Classifications) requires high time complexity as it necessities exclusive coding for each test sample. In previous work [35-37], authors have a debate that the classification accuracy is improved by the “collaborative” representation rather the “competitive” approach which is presented as sparseness constraint. Works in [18,35] also showed that the  $\ell_2$ -norm have the same competence with the less resources and computation. The approach which regularizes the representation based coding using the  $\ell_2$ -norm known as CRC. In CRC, for a single pixel all atoms are collaborated by which all atoms get the equal chance to participate for the representation. The optimization problem for CRC can be rewritten as follows:

$$\arg \min_{a_i^{(CR)}} \left\| x_i - \Phi a_i^{(CR)} \right\|_2 + \lambda_2 \left\| a_i^{(CR)} \right\|_2 \quad (5)$$

The biggest benefit of this replacement is the dramatic reduction of computational cost, because this new convex optimization problem has a closed-form solution given as:

$$a_i^{(CR)} = (\Phi^T \Phi + \lambda_2 \mathbf{I})^{-1} \Phi^T \mathbf{x}_i \quad (6)$$

Though the  $\ell_2$ -norm based regularizer in Equation (5) is no longer a sparsity constraint on the coefficients, it still has the potential to induce the competition among labeled samples from all candidate classes, which may cause the right class to have relatively smaller reconstruction error

and larger  $\ell_2$ -norm values of coefficients. For  $\ell_2$ -norm  $\left\| \hat{a}_i^k \right\|_2$  which features for discriminative information, the classification approach is modified as normalized residuals:

$$y_i = \text{class}(x_i) \arg \min_{k=1, 2, \dots, K} \frac{\left\| x_i - \Phi^k \hat{a}_i^k \right\|_2}{\left\| \hat{a}_i^k \right\|_2} \quad (7)$$

Till today, obtaining an accurate weight coefficient for representation based classification models is a challenge. The resulting coefficients hold the meaning as they represent the importance of each labeled sample. So, choosing or getting the accurate weight vector is the most critical element for the representation based classification. Keep in view,  $\ell_1$ -norm representation also called as sparse choose very few samples which cannot give the within class variations. Concerning the  $\ell_2$ -norm representation, it uses all atoms which may causes the between-class interference. The parameter  $\lambda$  adjustment affects the weight vector and residual norm which eventually leads to the classification accuracy.

$$CRC(x_i) = \arg \min_{k=1, \dots, K} r_k^{CRC}(x_i) \quad (8)$$

## 2.2 Kernelized SRC

Although representation based classifier SRC and CRC gets good performance in HSI classification, it is hard to classify the data that is not linearly separable. If the data set is not linearly separable, the kernel trick could be used into a feature space, in which the data become linearly separable [6]. As an extension of SRC, Kernel SRC (KSRC) [6] can be constructed using only the kernel function, without having to consider the mapping explicitly. The objective function can be described as:

$$\arg \min_{\alpha_i^{(KSR)}} \left\| \phi(\mathbf{x}_i) - \Phi \alpha_i^{(KSR)} \right\|_2^2 + \lambda_1 \left| \alpha_i^{(KSR)} \right|_1 \quad (9)$$

Where  $\phi = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_N)] \in \mathbb{R}^{L \times N}$  Equation- (9) can be modified to:

$$\arg \min_{\alpha_i^{(KSR)}} \left\| \Phi^T \phi(\mathbf{x}_i) \Phi^T \Phi \alpha_i^{(KSR)} \right\|_2^2 + \lambda_1 \left| \alpha_i^{(KSR)} \right|_1$$

Which equals

$$\arg \min_{\alpha_i^{(KSR)}} \left\| \mathbf{K}(\mathbf{x}_i) - \mathbf{K} \alpha_i^{(KSR)} \right\|_2^2 + \lambda_1 \left| \alpha_i^{(KSR)} \right|_1$$

Where  $\mathbf{K}(\mathbf{k}, \mathbf{x}) = [k(\mathbf{X}_1, \mathbf{X}_i), k(\mathbf{X}_2, \mathbf{X}_i), \dots, k(\mathbf{X}_N, \mathbf{X}_i)]^T \in \mathbb{R}^{N \times 1}$ , and  $\mathbf{K} = \Phi^T \Phi \in \mathbb{R}^{N \times N}$  is the gram matrix with  $\mathbf{K}_{ij} = k(\mathbf{X}_i, \mathbf{X}_j)$ .

After getting the weight vector in the kernel space,  $\Phi_k = [\Phi(\mathbf{X}_{k,1}), \Phi(\mathbf{X}_{k,2}), \dots, \Phi(\mathbf{X}_{k,M_k})]$  represents the kernel sub-dictionary for class k, and it can be further expressed as:

$$\begin{aligned} r_k^{KSRC}(\mathbf{X}_i) &= \left\| \Phi_k \alpha_k^{KSRC} - \phi(\mathbf{X}_i) \right\|_2 \\ &= \sqrt{(\phi(\mathbf{X}_i) - \Phi_k \alpha_k^{KSRC})^T (\phi(\mathbf{X}_i) - \Phi_k \alpha_k^{KSRC})} \\ &= \sqrt{k(\mathbf{X}, \mathbf{X}) + \alpha_k^{KSRC} \mathbf{K}_k \alpha_k^{KSRC} - 2 \alpha_k^{KSRC} K_k(\mathbf{X}, \mathbf{X})} \end{aligned} \quad (10)$$

Finally, class label of the testing pixel is determined as:

$$KSRC(\mathbf{X}_i) = \arg \min_{k=1, \dots, K} r_k^{KSRC}(\mathbf{X}_i) \quad (11)$$

### 3. SPECTRAL-SPATIAL CLASSIFICATION USING BILATERAL FILTERING AND REPRESENTATION BASED CLASSIFIER

Here, we proposed a new spectral-spatial approach for HSIs using the BF based post-processing approach as a probability optimization process. We have elaborated a flow

chart of our proposed approach in Fig. 1. with the Indian Pines image. Initially, a representation oriented classifier (such as SRC, CRC, or KSRC) is applied to the original HSI and obtained the probability map which specifies the membership of each class. And then to integrate the spatial-contextual information, a BF based post-processing is applied to reduce the mis-classification error. To calculate the image-context aware weights we have used the principal components of the original HSI as a spatial guidance image. On the basis of maximum probability criteria, in the last step of our proposed approach the optimized spatio-spectral classification map is acquired. Concretely, when the k<sup>th</sup> probability map at location i has an immense high intensity which leads to the high probability, thus the HSI i<sup>th</sup> pixel inclines to belong the k<sup>th</sup> class. Comparing with the available spatio-spectral HSI classification methods the proposed framework has three main steps:

- (1) Calculate the initial pixel-wise probability maps
- (2) Refine the soft-classified probability maps via the BF
- (3) Classification predicted on the maximum probability scheme

#### 3.1 INITIAL PROBABILITY MAP VIA REPRESENTATION BASED PIXEL-WISE CLASSIFIER

In this work, we have employed the representation based classifier (SRC, CRC, or KSRC) on the HSIs to perform the pixel-wise classification. The different state-of-the-art pixel-wise classifier can also be adopted to test the different flavor of the proposed framework. The detail of the RC is given in section 2.

The SRC (CRC or KSRC) computes the class labels  $y_i$  for the testing samples  $\mathbf{x}_i$  directly with the characteristics of the recovered coding coefficient vector  $\alpha_i$  via the minimal

residual aforementioned in section 2. For instance, for SRC the class labels  $y_i$  for the testing samples  $x_i$  is obtained using the Equation (4).

Initially, by applying the SRC (CRC or KSRC) the following results can be obtained:

- (1) A classification map with unique class labeling for each pixel.
- (2) and a  $\mathbf{P}$  (Probability Maps) for each pixel having the probability defined in Equation (13).

Here the initial spectral only classification map  $y_i$  is given as a probability maps, i.e.  $\mathbf{P} = (\mathbf{P}^1, \dots, \mathbf{P}^k)$ , in which  $P_i^k \in [0,1]$  is the pixel-wise probability of a pixel  $i$  which is in  $k^{\text{th}}$  class. In addition,  $P_i^k$  is given as:

$$P_i^k = \begin{cases} 1 & \text{if } y_i = k \\ 0 & \text{Otherwise} \end{cases} \quad (12)$$

According to Equation (13), the probabilities we have obtained are the hard label probabilities such as [0,1]. We transformed the hard probability into soft-classified probability maps by averaging the number of k-classes of 8-neighbor's probability.

### 3.2 Refinement of the Soft-classified Probability Maps via the BF

Initially, in the obtained soft-classified probability maps the spatial information is not considered. Consequently, causes the noisy probability maps with not aligned authentic object boundaries. However, for the HSI, the edges and spatial image details is very important to preserve while performing the filtering for the refined classification results based on BF. Thus, to effectively integrate the spatial contextual information, these points must be in consideration.

- (1) For a given spatial window, the spectral similarity and geometric closeness should be under consideration while taking out the weighted average.
- (2) The details of the HSI image i.e. spatial and spectral values and edges must be preserve.

Keep in view these issues, to solve the above mentioned problem, the probability maps are optimized by the BF which is capable to do the weighted average of the pixels for a given spatial window within a band. Thus, probabilistic BF not only removes the artifacts but it also preserves the edges. We have presented an effective BF based post-processing approach to smoothen the probability maps. The BF is consisting of Gaussian filter which is the integration of distances in image plane (the spatial domain) and intensity axis (the image context-aware distance).

Concretely, by taking the weighted average with neighborhood probabilities the optimized refined probabilities can be modeled as:

$$\hat{P}_i^k = \sum_{j \in \Omega_i} W_{i,j}(\mathbf{I}) P_j^k \quad (13)$$

Here the  $i^{\text{th}}$  and  $j^{\text{th}}$  pixels are represented with  $i$  and  $j$ . The filtering weight  $W$  keep aligned the edges of the reference image  $\mathbf{I}$ , while smoothing the image. Thus, here we have two main problems:

- (1) Choosing the weighted edge preserving filtering, and
- (2) Selecting the spatial guidance (reference) image.

To resolve this issue, keeping in view the aforementioned problem, in our work benefitted from the BF characteristics, we have employed the joint BF to improve the misclassification error with a post-

processing spatial contextual information under the spatial guidance image. The BF is an integration of the soft-classified maps with the original hyperspectral image as spatial guidance which computes the spatial and image context-aware weights, instead of the probability maps given as:

$$BF(\hat{P}_i^k) = \frac{1}{R_i} \sum_{j \in \Omega_i} G_{\delta_s}(\|i - j\|) G_{\delta_r}(\|I_i - I_j\|) P_j^k \quad (14)$$

and the normalization term  $R_i$  ensures that pixel weights sum to 1 and defined by:

$$R_i = \sum_{j \in \Omega_i} G_{\delta_s}(\|i - j\|) G_{\delta_r}(\|I_i - I_j\|) \quad (15)$$

The spatial and image context-aware are based on the Gaussian decreasing distances, respectively. i.e.

$$G_{\delta_s}(\|i - j\|) = \exp\left(-\frac{\|i - j\|}{\delta_s^2}\right) \quad (16)$$

and

$$G_{\delta_r}(\|I_i - I_j\|) = \exp\left(-\frac{\|I_i - I_j\|}{\delta_r^2}\right) \quad (17)$$

where  $i$  is the central pixel of Gaussian kernel while  $j$  represents the pixel in, a local window with size  $(2\delta_s + 1) \times (2\delta_s + 1)$ . The input image is  $\hat{P}_i^k$  and  $I$  is the spatial reference image. The local window size is controlled by the  $\delta_s$  which will be helpful in filtering a pixel, while the  $\delta_r$  determine the decreasing weight of a pixel. In the Fig. 1, we have explained the concrete process to calculate the BF weight in joint bilateral filtering on soft-classified probability maps block. From Equation (15) it is clearly understandable that the spatial and spectral information are integrated for the proposed BF post-processing framework. Based on Equation (15), it can be concluded that when the pixel  $i$  neighborhood in the reference image

are of same intensities, i.e.  $I_i \approx I_j$ , the pixel  $j$  weight will be large, specifically when it is proximate to  $i$ , i.e.  $\|i - j\|$  is very small. While, the neighboring pixels which have different intensities in the reference image, the scenario will be quite antithesis. So, the filtering outcome will be same for the adjacent pixels which having the similar intensity or color in the spatial reference image. This characteristic reveals the Gaussian smoothing for homogenous zones of the image, across boundaries of image there will be no filtering. Benefitted with the BF features, i.e. to avoid the blurring while removing the noise between the homogenous zones, we can get the more accurate and effective classification accuracy. Moreover, BF provides the  $\delta_s$  and  $\delta_r$  parameters modification in non-iterative way.

We have adopted the PCA for the spatial reference (guidance) image. The initial three principal components are utilized as shown in Fig. 1.

Probability optimization with BF has two key advantages.

- (1) Noise probabilities which obtained from the initial pixel-wise classifier can be smoothed effectively.
- (2) Spatial-contextual information added refined probabilities are always coordinated with the boundaries of authentic object.

With these advantages it is concluded that the spatial guidance image is efficiently used in the BF post-processing for the spatial contextual information.

### 3.3 Choosing Maximum Probability

According to the Fig. 1, after post-processed the probability maps with BF, the class label for each pixel is based on the choosing maximum probability criteria. This

last step aims to transform the obtained optimized probability maps  $\hat{P}_i^k$  into the final spectral-spatial classification result  $\hat{y}_i$ . Once the optimized spatial-context aware probability maps are obtained, by choosing a maximum probability of pixel  $i$ , the class label can be obtained as:

$$\hat{y}_i = \arg \max_k \hat{P}_i^k \quad (18)$$

#### 4. COMPARATIVE EXPERIMENTAL ANALYSIS

##### 4.1 Experimental Parameter Setting

In this experiment, we have evaluated the performance of aforementioned discussed post-processing classification methods using three very well-known hyperspectral data sets i.e. Indian Pines, University of Pavia and Centre of Pavia images. The detail of these data sets is described in **Table 1**. For the fair comparison of these methods, we have chosen the same number of data samples. To perform the supervised classification, in Indian Pines data set, for each class we

have chosen the 10% of samples from the ground truth data as training samples and the remaining 90% samples are utilized as the test samples. For the University of Pavia data set, we have selected the 9% of samples from the reference data will be used for the training the classifier and the remaining will be helpful for the testing. In addition, concerning the Centre of Pavia data set we utilized 5% training samples while remaining will be used as the testing samples, are mentioned in Tables 2-4, respectively. The classification results are demonstrated with the well-known quality metrics i.e. OA, AA, and Kappa coefficient ( $\kappa$ ), which are widely used quality metrics for the HSI classification. For the comparative analysis of our proposed framework, we have considered the traditional spectral-spatial classifiers. We have considered the very well-known LORSAL with the multilevel logistic spatial prior denoted as LOR-MLL [10]. Another method composite kernel with SVM denoted as SVM-CK [5] has recently showed the excellence performance in hyperspectral image classification. Furthermore, a recently proposed semi-local graph based post-processing spectral-spatial classification method named as KSRC-SSG [38] is also considered for the experimental analysis

TABLE 1. DESCRIPTION OF DATA SETS

	Indian Pines Image	University of Pavia Image	Centre of Pavia Image
Instruments	Airborne Visible/Infrared Imaging Spectrometer	Reflective Optics System Imaging Spectrometer	Reflective Optics System Imaging Spectrometer
Scenes	Indian Pines region of North-west Tippecanoe Country, Indi- and, USA	Urban area of Pavia University, Italy	Urban area of Pavia City, Italy
Image Spatial size	145×145	610×340	1096×492
Spectral Bands	200(220 in total, and 20 noise and water absorption bands are removed)	103(115 in total, and 12 noise bands are removed)	102(115 in total, and 13 noise bands are removed)
Wavelength range	0.4-2.5µm	0.43-0.86 µm	0.43-0.86 µm
Spatial resolution	20 m	1.3 m	1.3 m
Spectral resolution	10nm	4 nm	4 nm
Ground-truth classes	16	9	9
Labeled pixels	10366	42776	103476

**TABLE 2. CLASSIFICATION ACCURACY (%) FOR THE INDIAN PINES IMAGE USING 1036 (ABOUT 10%) LABELED SAMPLES**

Class Type	Train	Test	SRC	CRC	KSRC	LOR- MLL	SVM-CK	KSRC-SSG	SRC- BF	CRC- BF	KSRC-BF
Alfalfa	6	48	35.58	33.33	75.00	70.42	95.83	82.17	94.62	98.48	98.87
Corn no till	144	1290	78.53	74.96	81.50	92.96	96.67	96.23	86.51	90.82	89.88
Corn-min till	84	750	51.87	46.53	74.63	86.65	90.93	95.45	87.07	98.37	96.72
Corn	24	210	28.57	30.00	69.57	79.38	85.71	94.24	91.72	98.91	100.0
Grass/Pasture	50	447	80.76	88.14	91.41	94.25	93.74	93.71	86.27	76.92	97.59
Grass/Tree	75	672	99.40	97.62	96.09	98.66	97.32	98.72	99.52	97.34	99.17
Grass/Pasture-Mowed	3	23	17.39	8.70	76.96	50.00	69.57	84.35	85.54	96.44	98.53
Hay-Windrowed	49	440	99.32	100.0	98.89	99.39	98.41	99.18	90.00	97.54	99.17
Oats	2	18	16.67	55.56	73.89	50.00	55.56	74.44	97.88	97.56	98.88
Soybeans-no till	97	871	63.95	36.17	77.21	90.20	93.80	91.73	93.16	100.0	100.0
Soybeans-min till	247	2221	86.04	86.90	85.93	93.89	94.37	96.11	98.35	99.54	91.55
Soybeans-Clean till	62	552	57.79	50.00	81.32	94.37	93.66	94.98	96.42	100.0	97.01
Wheat	22	190	100.0	99.47	99.11	99.58	99.47	99.68	99.18	99.38	98.41
Woods	130	1164	97.94	97.94	96.07	97.66	99.14	97.52	99.86	99.04	99.12
Building-Grass-Tree-Drives	38	342	35.96	43.7	60.80	79.30	87.43	95.56	99.67	79.70	95.25
Stone-Steel Towers	10	85	90.59	90.69	94.00	73.06	100.0	90.59	98.43	100.0	98.17
Overall Accuracy (%)			77.99	74.70	85.12	92.70	94.86	95.81	93.66	91.83	96.87
Average Accuracy (%)			65.27	61.84	83.23	84.36	90.73	92.83	93.95	95.69	97.38
k Statistic			0.726	0.715	0.830	0.917	0.941	0.952	0.926	0.907	0.963

**TABLE 3. CLASSIFICATION ACCURACY (%) FOR THE UNIVERSITY OF PAVIA IMAGE USING 3921 (ABOUT 9%) LABELED SAMPLES**

Class Type	Train	Test	SRC	CRC	KSRC	SVM-CK	LOR-MLL	KSRC-SSG	KSRC-BF	CRC-BF	SRC-BF
Asphalt	548	6304	80.65	54.03	76.09	88.48	79.85	77.17	82.23	89.79	89.20
Meadows	540	1814	64.74	92.06	69.61	76.22	84.68	81.61	72.47	85.07	88.82
Gravel	392	1815	1815	87.79	72.12	73.56	81.87	82.42	82.26	82.37	90.21
Trees	524	291	2912	89.73	98.11	98.76	96.36	95.46	98.56	97.33	95.99
Metal Sheets	265	1113	1113	100.0	99.73	99.70	99.37	99.03	99.82	99.87	97.53
Bare-Oil	532	4572	4572	59.68	87.66	97.47	93.55	96.94	93.92	94.05	99.50
Bitumen	375	981	981	99.22	88.07	94.74	90.21	93.83	92.46	92.51	94.34
Bricks	514	3364	3364	45.69	89.51	96.66	92.81	94.65	78.78	93.03	97.28
Shadows	231	795	795	99.80	97.21	96.20	95.35	97.47	96.98	97.85	98.17
Overall Accuracy (%)			76.87	78.01	78.93	85.69	87.18	86.16	90.26	86.63	88.73
Average Accuracy (%)			86.32	80.89	86.10	91.66	90.47	90.95	94.86	93.41	94.03
k Statistic			0.709	0.704	0.725	0.819	0.833	0.824	0.887	0.851	0.893

For SVM-CK algorithm the data set must be normalized before the experiment. The RBF kernel is used both for KSRC and SVM-CK. The code for LOR-MLL is given on Dr. Li's homepage. The default parameters for KSRC-SSG is adopted. In our proposed model it should be noticed to keep the blur degree and filtering size not very small, if so the filtering will only consider the limited local spatial information. Furthermore, selecting the large window and blur degree may cause the over smoothing. For example, a class with contains 20 scale is totally mis-classified when the filtering size is very large thus the average classification accuracy reduces dramatically. Although for different images, the classification accuracy can further be improved with the adjustment of BF parameters  $\delta_s$  and  $\delta_r$ . For our experimental analysis we set the parameters  $\delta_s = 3$  and  $\delta_r = 0.03$ .

#### 4.2 Visual and Numeric Comparison

The comparison of visual classification map for different methods are illustrated in Figs. 2-4. These visual maps illustrated that the pixel-wise SRC, CRC, or KSRC classification accuracy are not very satisfactory as the

obtained classification map have a lot of mis-classified estimates, while comparing the LOR-MLL, KSRC-SSG, SVM-CK and our method yields the better performance. More precisely with our proposed method the OA is increased 14-15% while comparing the pixel-wise classifier. In the comparison of a KSRC-SSG method, our proposed bilateral filtering based post-processing model gives the competitive classification results. Form the visual classification maps shown in Figs. 2-4, and by observing the numeric results presented in Tables 2-4, we can conclude that the combination of post-processed spatial information with a pixel-wise classification approach yield higher accuracies as compare to the with pixel-wise (spectral only) classification models.

We have represented the results in boldface that are significantly better than others. From the experimental results, we have the following outcomes: First, from Tables 2-4 we can easily observe that our proposed BFbased post-processing framework gains the competitive performances among all of the state-of-the-art spectral-spatial classifiers such as SVM-CK, LOR-MLL, and

TABLE 4. CLASSIFICATION ACCURACY (%) FOR THE CENTRE OF PAIVA IMAGE USING THE 5% LABELED SAMPLES

Class Type	Train	Test	SRC	CRC	KSRC	LOR-MLL	SVM-CK	KSRC-SSG	SRC-BF	CRC-BF	KSRC-BF
Asphalt	745	6453	99.35	97.46	98.13	99.41	99.34	100.0	100.0	100.0	100.0
Meadows	785	5722	83.52	93.08	92.76	94.04	94.39	95.49	96.18	94.74	95.32
Gravel	797	2094	97.09	97.23	97.03	96.80	95.65	95.89	97.72	95.31	98.20
Trees	485	1667	73.67	77.02	88.84	89.62	84.70	89.90	83.60	84.88	89.83
Metal Sheets	820	5729	95.92	98.39	94.89	97.85	99.04	98.50	98.15	96.04	98.99
Bare-Soil	820	5729	97.01	94.32	96.13	97.49	96.39	96.67	95.27	96.99	97.85
Bitumen	808	6479	92.65	97.50	95.40	98.12	98.41	97.31	97.35	91.69	97.41
Bricks	223	2899	99.45	99.83	99.34	99.83	99.97	98.41	99.93	99.72	98.62
Shaows	195	1970	99.09	99.95	99.39	100.0	98.73	99.49	99.23	99.83	99.39
Overall Accuracy (%)			97.13	96.81	97.19	98.73	98.47	98.32	98.24	98.15	98.94
Average Accuracy (%)			93.08	94.96	95.77	97.20	96.43	96.29	96.19	95.94	97.86
k Statistic			0.948	0.943	0.949	0.977	0.972	0.980	0.962	0.953	0.980

KSRC-SSG. It is also worth-full to notify that the proposed BF framework achieves significant improvements over the representation based pixel-wise classifiers, such as SRC, CRC and KSRC, which demonstrates the potential

and advantage of the introduced filtering based post-processing approach. In contrast to the LOR-MLL and SVM-CK method, the classification map of the post-processed KSRC-SSG is smoother.

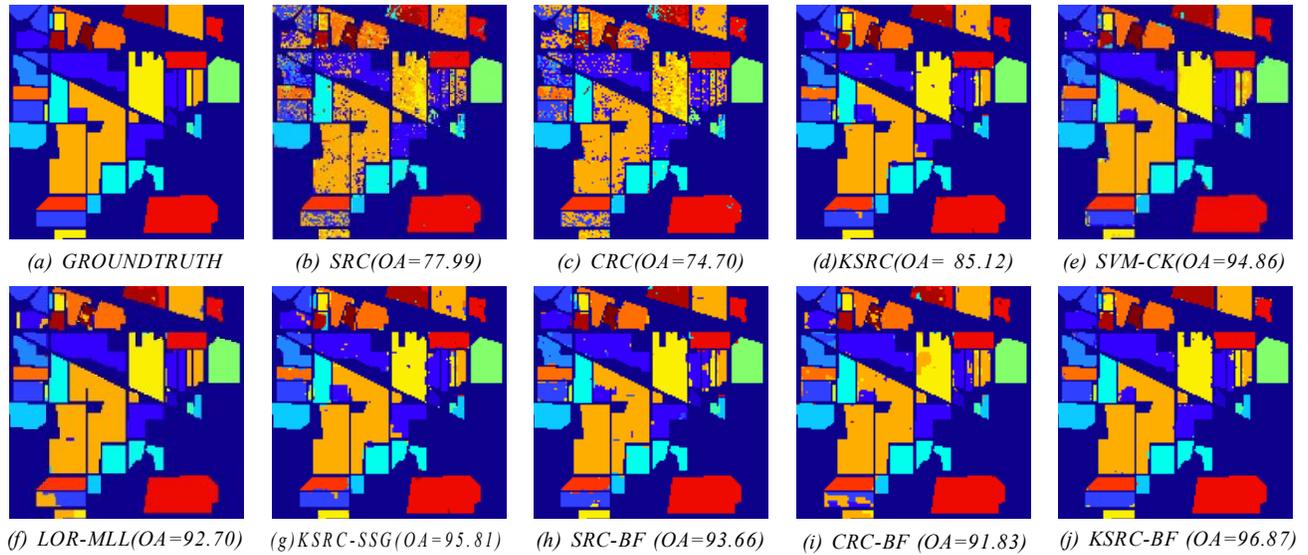


FIG. 2. INDIAN PINES IMAGE: REFERENCE DATA AND VISUAL CLASSIFICATION RESULTS

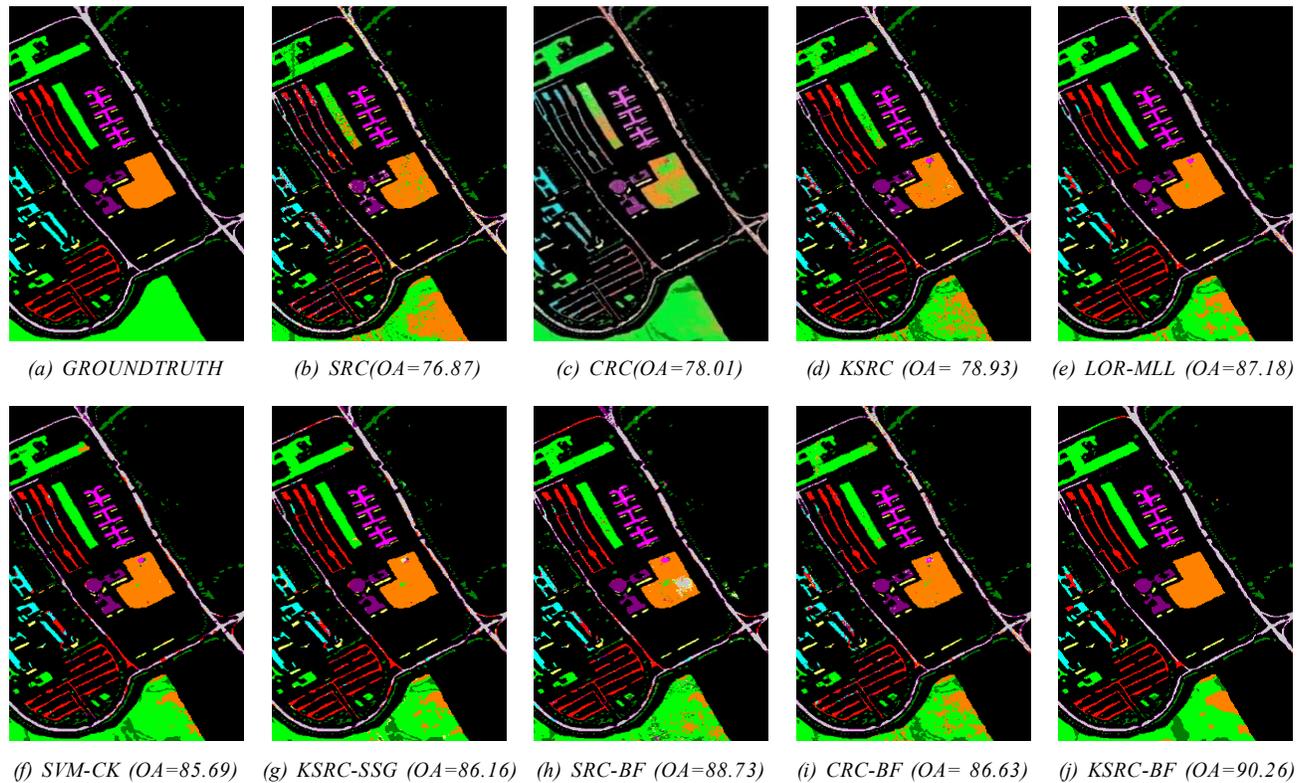


FIG. 3. UNIVERSITY OF PAVIA IMAGE: REFERENCE DATA AND VISUAL CLASSIFICATION RESULTS

From Figs. 2-4 it can be seen that the performance of our proposed BF based method are better than the others, and the KSRC-BF gives the best class-specific and most accurate overall accuracy compared with SRC-BF and CRC-BF. From Figs. 2-4 it can be seen that the classification maps of the post-processing approaches contains more homogeneous regions and the region edges of KSRC-BF are protected well, when compared with the other state-of-the-art spatial-spectral classification map. It is also important to notify that the optimized and soft-classified probability map do not have a big difference, which is because of the fact that the pixel-wise information is well considered in filtering process. For instance, for the University of Pavia data set (Fig. 3), for the LOR-MLL in Fig. 3(e) the “Bricks” were misclassified with “Asphalt”. In addition, “Grave” region is labeled with “Bricks” and some part is labeled a s “Asphalt”. When discussing the Figs. 3(f-g), in case of the SVM-CK, KSRC-SSG approaches mostly region of “Bricks” mis-classified as

“Gravel” and also as “Asphalt”. Moreover, we can clearly visualize that in Figs. 3(f-g), some amount of salt-and-pepper classification noise is available.

Similarly, we can also see the same findings for the Indian pines image. such as, in LOR-MLL the, at the top of the map “Bldg-grass-tree-drives” is mis-classified with the “Soybeans-clean till”. Furthermore, it is also facing the noise in the “soybean-min-till” and ‘Alfalfa” class. Also the SVM-CK and KSRC-SSG having some mis-classification, i.e. “Alfalfa”, “stone-steel-tower” and “corn-no-till”. It is worth to mention here that our proposed framework reduces these mis-classification results and preserves the edge boundaries well. Similar finding, we have for the Indian Pines (Fig. 2) and Centre of Pavia results (Fig. 4). The proposed KSRC-BF method showed the most similar classification map as the reference data. To further improve the classification results, different pixel-wise classifiers can be integrated with our proposed BF framework.

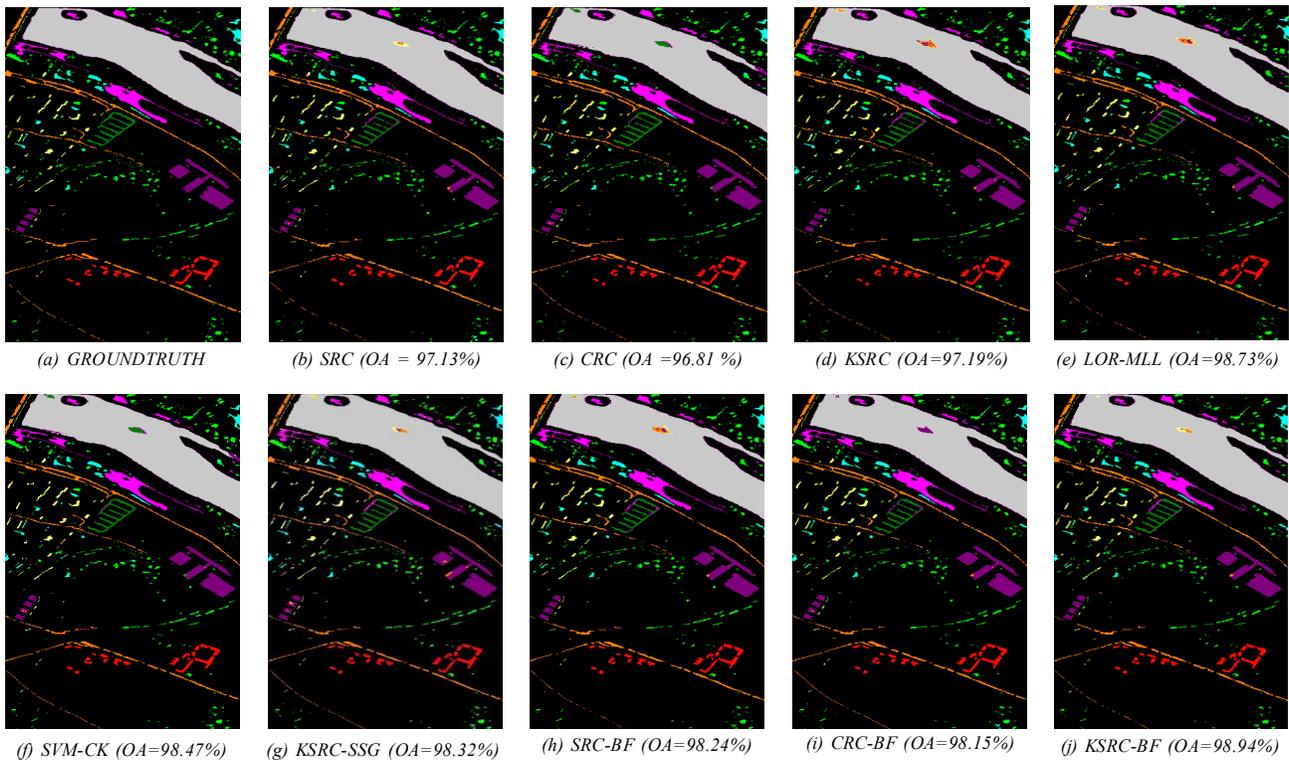


FIG. 4. CENTER OF PAVIA IMAGE. GROUND-TRUTH AND THE OVERALL ACCURACY OF THE METHODS WITH ABOUT 5% LABELED SAMPLES

### 4.3 Influence of the $\delta_r$ Parameters

For the proposed method the blur degree ( $\delta_r$ ) and the filtering size ( $\delta_s$ ) are the key parameters. The influence of these parameters are depicted in Fig. 5 using the Indian Pines image. We have chosen the 10% training samples as shown in Table 2. To test the impact of the  $\delta_r$  parameter we have fixed the parameter  $\delta_s = 3$ , while for testing the impact of the  $\delta_s$  parameter the blur degree parameter  $\delta_r$  is fixed to 0.03. From the Fig. 5(a), it is worth to notify that the setting the  $\delta_r$  parameter is vitally very important for the proposed method. We set the  $\delta_r$  as 0.005-0.04 with variation of 0.005 to analyzed the influence of the  $\delta_r$  parameter keeping all other parameters are fixed. We can observe in the Fig.5(a) that the classification accuracies (OA, AA and  $\kappa$ ) increases in the case when the  $\delta_r = 0.03$  while for  $\delta_r > 0.03$  the tendency of accuracies have decreased gradually. For the  $\delta_r = 0.03$ , we get the best classification result.

To analyze the influence of the filtering size  $\delta_s$  parameter, we set filtering size from 0-6 with step size of 1. It is also worth to recall that the spatial local window size is directly related with the filtering size as we define previously, i.e.  $\Omega = (2\delta_s + 1) \times (2\delta_s + 1)$ , thus setting the  $\delta_s$  parameter is significantly important to the performance of our proposed BF based method. The influence of  $\delta_s$  parameter is illustrated in the Fig. 5(b). The findings we have that the tendency of the classification accuracies are varied as we observed in the analysis of the  $\delta_r$  parameter. With the  $\delta_s \leq 3$ , the classification accuracies are increases gradually while for the filtering size parameter  $\delta_s > 3$ , there is a dramatically reduction in the average accuracy, while the OA and  $\kappa$  decreases gradually. Although the AA for  $\delta_s = 4$  is not reduced so much but for the  $\delta_s = 5$  and  $\delta_s = 6$  it is showing the very low AA, the reason is that because for a small-scale class like ‘‘Oats’’ which contains only 20 samples (Table 2) can be totally mis-classified when the filtering size is very large. It can be concluded that  $\delta_s$  should not be set to be too small or large. For the best result we set the filtering size  $\delta_s = 3$  which leads to the spatial window size of  $7 \times 7$ .

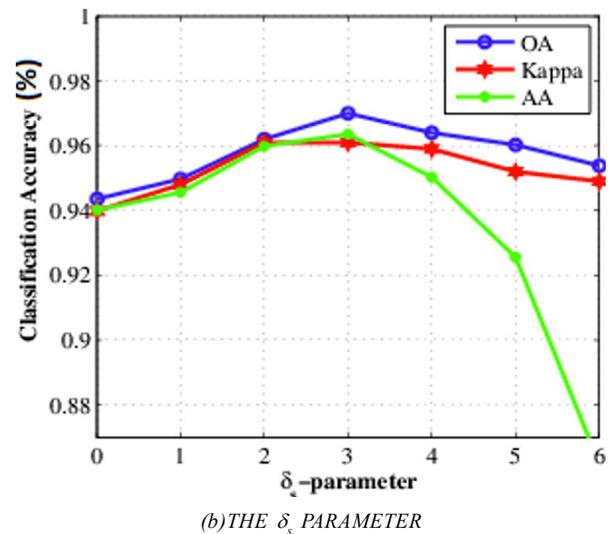
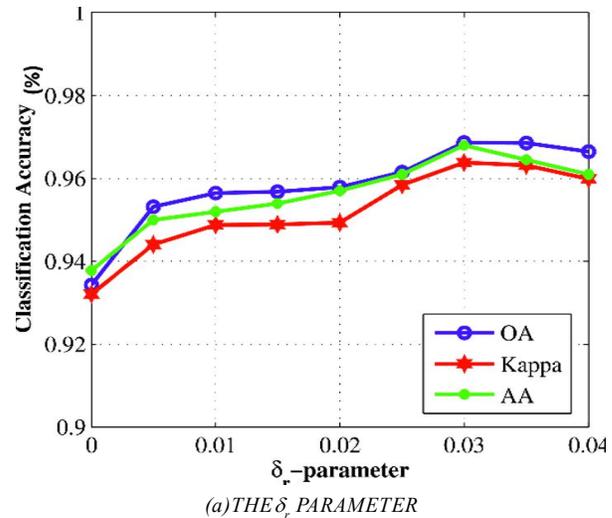


FIG. 5. IMPACT OF THE BILATERAL FILTERING PARAMETERS

## 5. CONCLUSION

Post-processing spatial regularization has been widely used during the past few years to improve the HSI classification. The BF which jointly exploit the spectral and spatial information causes either the loss, regularizer is based on the spatial window that defines the neighborhood of a pixel. We proposed a generic but powerful post-processing filtering approach which includes both spatial and spectral information as a post-processing for the HSI classification is presented. The proposed BF based method needs edge-preserving filtering many times while it takes once a PCA

decomposition. Therefore, the major advantage of the BF based post-processing methods is simplicity, non-iterative approach and easier implementation. Thus, in real application the proposed approach is very useful. In future by tuning the filtering and blur degree size of the BF, the performance can be further enhanced. Experimental results, tested using three widely used HSI data sets, validate the best classification results in comparison of traditional methods.

## ACKNOWLEDGMENT

The author would like to pay the thanks and sincere gratitude to the Prof. D. Landgrebe, TU Chemnitz, for making the AVIRIS Indian Pines hyperspectral dataset available to the community, Prof. P. Gamba, University of Pavia, for providing the ROSIS data over Pavia, Italy, along with the training and test set. And also want to pay thanks to the anonymous reviewer for their valuable comments and suggestions.

## REFERENCES

- [1] Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Ualquieri, A., Marconcini, M., Tilton, J. C., and Trianni, G., "Recent Advances in Techniques for Hyperspectral Image Processing", *Remote Sensing Environment*, Volume 113, No. 1, pp. 110-122, 2009.
- [2] Benediktsson, J.A., Palmason, J.A., and Sveinsson, J.R., "Classification of Hyperspectral Data from Urban Areas Based on Extended Morphological Profiles", *IEEE Transaction on Geoscience and Remote Sensing*, Volume 43, No. 3, pp. 480-491, 2005.
- [3] Fauvel, M., Chanussot, J., and Benediktsson, J.A., "Kernel Principal Component Analysis for the Classification of Hyperspectral Remote Sensing Data Over Urban Areas", *Journal of Advance Signal Processing*, Volume 1, pp. 1-4, 2009.
- [4] Dalla, M.M., Atli, B.J., Waske, B., and Bruzzone, L., "Extended Profiles with Morphological Attribute Filters for the Analysis of Hyperspectral Data", *International Journal of Remote Sensing*, Volume 31, No. 22, pp. 5975-5991, 2010.
- [5] Camps-Valls, G., Gomez-Chov, L.A., Munoz-Mari, J., Vila-Frances, J., and Calpe-Maravilla, J., "Composite Kernels for Hyperspectral Image Classification", *IEEE Geoscience and Remote Sensing Letter*, Volume 3, No. 1, pp. 93-97, 2006.
- [6] Li, J., Marpu, P.R., Plaza, A., Bioucas-Dias, J.M., and Benediktsson, J.A., "Generalized Composite Kernel Framework for Hyperspectral Image Classification", *IEEE Transaction on Geoscience and Remote Sensing*, Volume 51, No. 9, pp. 4816-4829, 2013.
- [7] Li, J., "Multiple Feature Learning for Hyperspectral Image Classification", *IEEE Transaction on Geoscience and Remote Sensing*, Volume 53, No. 3, pp. 1592-1606, 2015.
- [8] Tarabalka, Y., Benediktsson, J.A., and Chanussot, J., "Spectral-Spatial Classification of Hyperspectral Imagery Based on Partitional Clustering Techniques", *IEEE Transaction on Geoscience and Remote Sensing*, Volume 47, No. 8, pp. 2973-2987, 2009.
- [9] Soomro, B.N., Xiao, L., Huang, L., Soomro, S.H., and Molaei, M., "Bilayer Elastic Net Regression Model for Supervised Spectral-Spatial Hyperspectral Image Classification", *IEEE Journal of Selected Topics in Earth Observations and Remote Sensing*, Volume 9, No.9, pp. 4102-4116, 2016.
- [10] Li, J., Bioucas-Dias, J.M., and Plaza, A., "Hyperspectral Image Segmentation Using a New Bayesian Approach with Active Learning", *IEEE Transactions on Geoscience and Remote Sensing*, Volume 49, No. 10, pp. 3947-3960, 2011.
- [11] Soomro, B.N., Xiao, L., Molaei, M., Huang, L., Lian, Z., and Shahzad, S., "Local and Non-Local Context-Aware Elastic Net Representation Based Classification for Hyperspectral Images", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Volume 10, No.6, pp. 2922-2939, 2017.
- [12] Paris, S., and Durand, F., "A Fast Approximation of the Bilateral Filter Using a Signal Processing Approach", *International Journal of Computer Vision*, Volume 81, No.1, pp. 24-52, 2009.
- [13] Soomro, B.N., Xiao, L., Soomro, S.H., and Molaei, M., "Spatial-Aware Supervised Learning for Hyperspectral Image Classification: Comprehensive Assessment", *Journal of Donghua University*, Volume 33, No.6, pp. 954-960, 2016.
- [14] Benediktsson, J., Pesaresi, M., and Amason, K., "Classification and Feature Extraction for Remote Sensing Images from Urban Areas Based on Morphological Transformations", *IEEE Transactions on Geoscience and Remote Sensing*, Volume 41, No. 9, pp. 1940-1949, 2003.
- [15] Fauvel, M., Chanussot, J., and Benediktsson, J.A., "A Spatial-Spectral Kernel Based Approach for the Classification of Remote Sensing Images", *Pattern Recognition*, Volume 45, No.1, pp. 381-392, 2012.

- [16] Camps-Valls, G., Shervashidze, N., and Borgwardt, K.M., "Spatio-Spectral Remote Sensing Image Classification with Graph Kernels", *IEEE Geoscience and Remote Sensing Letter*, Volume 7, No. 4, pp. 741-745, 2010.
- [17] Kang, X., Li, S., Fang, L., Li, M., and Benediktsson, J.A., "Extended Random Walker-Based Classification of Hyperspectral Images", *IEEE Transactions on Geoscience and Remote Sensing*, Volume 53, No. 1, pp. 144-153, 2015.
- [18] Zhang, L., Yang, M., Feng, X., Ma, Y., and Zhang, D., "Collaborative Representation Based Classification for Face Recognition", *arXiv preprint arXiv:1204.2358*, 2012.
- [19] Xu, L. and Li, J. "Bayesian Classification of Hyperspectral Imagery Based on Probabilistic Sparse Representation and Markov Random Field", *IEEE Transaction on Geoscience and Remote Sensing*, Volume 11, No. 4, pp. 823-827, 2014.
- [20] Tomasi, C., and Manduchi, R., "Bilateral Filtering for Gray and Color Images", *Proceedings of International Conference on Computer Vision*, pp. 839-846, 1998.
- [21] Farbman, Z., Fattal, R., Lischinski, D., and Szeliski, R., "Edge-Preserving Decompositions for Multi-Scale Tone and Detail Manipulation", *ACM Transactions on Graphics*, Volume 27, No.3, pp. 67-76, 2008.
- [22] He, K., Sun, J., and Tang, X., "Guided Image Filtering", *IEEE Transactions on Pattern Analysis and Machine Intelligent*, Volume 35, No.6, pp. 1397-1409, 2013.
- [23] Gastal, E.S L., and Oliveira, M.M., "Domain Transform for Edge-Aware Image and Video Processing", *ACM Transactions on Graphics*, Volume 30, No.4, pp. 69-78, 2011.
- [24] Qiu, T., Wang, A., Yu, N., and Song, A., "LLSURE: Local Linear Sure-Based Edge-Preserving Image Filtering", *IEEE Transactions on Image Processing*, Volume 22, No.1, pp. 80-90, 2013.
- [25] Xu, L., Lu, C., Xu, Y., and Jia, J., "Image Smoothing Via  $\ell_0$ -Gradient Minimization", *ACM Transactions on Graphics*, Volume 30, No. 6, pp. 174-185, 2011.
- [26] Hosni, A., Rhemann, C., Bleyer, M., Rother, C., and Gelautz, M., "Fast Cost-Volume Filtering for Visual Correspondence and Beyond", *IEEE Transactions on Pattern Analysis and Machine Intelligent*, Volume 35, No. 2, pp. 504-511, 2013.
- [27] Li, S., Kang, X., and Hu, J., "Image Fusion with Guided Filtering", *IEEE Transactions on Image Processing*, Volume 22, No. 7, pp. 2864-2875, 2013.
- [28] Li, S., and Kang, X., "Fast Multi-Exposure Image Fusion with Median Filter and Recursive Filter", *IEEE Transactions on Consumer Electronics*, Volume 58, No. 2, pp. 626-632, 2012.
- [29] Zhang, B., and Allebach, J.P., "Adaptive Bilateral Filter for Sharpness Enhancement and Noise Removal", *IEEE Transaction on Image Processing*, Volume 17, No. 5, pp. 664-678, 2008.
- [30] He, K., Sun, J., and Tang, X., "Single Image Haze Removal Using Dark Channel Prior", *IEEE Transactions on Pattern Analysis and Machine Intelligent*, Volume 33, No. 12, pp. 2341-2353, 2011.
- [31] Lin, C.H., Tsai, J.S., and Chiu, C.T., "Switching Bilateral Filter with a Texture/Noise Detector for Universal Noise Removal", *IEEE Transactions on Image Processing*, Volume 19, No. 9, pp. 2307-2320, 2010.
- [32] Kotwal, K., and Chaudhuri, S., "Visualization of Hyperspectral Images Using Bilateral Filtering", *IEEE Transactions on Geoscience and Remote Sensing*, Volume 48, No. 5, pp. 2308-2316, 2010.
- [33] Soomro, S.H., Xiao, L., and Soomro, B.N., "Hyperspectral Image Classification Via Elastic Net Regression and Bilateral Filtering", *Proceedings of IEEE International Conference on Progress in Information and Computing*, Volume 1, No. 1, pp. 56-60, Nanjing, China, 2015.
- [34] Zhang, L., Zhou, W.D., Chang, P.C., Liu, J., Yan, Z., Wang, T., and Li, F.Z., "Kernel Sparse Representation Based Classifier", *IEEE Transactions on Signal Process*, Volume 60, No. 4, pp.1684-1695, 2012.
- [35] Zhang, L., Yang, M., and Feng, X., "Sparse Representation or Collaborative Representation: Which Helps Face Recognition?", *Proceedings of IEEE International Conference on Computer Vision*, [DOI: 10.1109/ICCV.2011.6126277], pp. 471-478, 2011.
- [36] Li, W., and Du, Q., "Joint Within-Class Collaborative Representation for Hyperspectral Image Classification", *IEEE Journal of Selected Topics and Applied Earth Observation in Remote Sensing*, Volume 7, No. 6, pp. 2200-2208, 2014.
- [37] Waqas, J., Yi, Z., and Zhang, L. "Collaborative Neighbor Representation Based Classification Using  $\ell_2$  Minimization Approach", *Pattern Recognition Letter*, Volume 34, No. 2, pp.201-208, 2013.S
- [38] Liu, J., Wu, Z., Sun, L., Wei, Z., and Liang, X., "Hyperspectral Image Classification Using Kernel Sparse Representation and Semilocal Spatial Graph Regularization", *IEEE Geoscience and Remote Sensing Letters*, Volume 11, No. 8, pp. 1320-1324, 2014.