Feature Extraction of Hyperspectral Images With Image Fusion and Recursive Filtering

Xudong Kang, *Student Member, IEEE*, Shutao Li, *Member, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

***Abstract*—Feature extraction is known to be an effective way in both reducing computational complexity and increasing accuracy of hyperspectral image classification. In this paper, a simple yet quite powerful feature extraction method based on image fusion and recursive filtering (IFRF) is proposed. First, the hyperspectral image is partitioned into multiple subsets of adjacent hyperspectral bands. Then, the bands in each subset are fused together by averaging, which is one of the simplest image fusion methods. Finally, the fused bands are processed with transform domain recursive filtering to get the resulting features for classification. Experiments are performed on different hyperspectral images, with the support vector machines (SVMs) serving as the classifier. By using the proposed method, the accuracy of the SVM classifier can be improved significantly. Furthermore, compared with other hyperspectral classification methods, the proposed IFRF method shows outstanding performance in terms of classification accuracy and computational efficiency.**

***Index Terms*—Feature extraction, hyperspectral image, image classification, image fusion (IF), recursive filtering.**

I. Introduction

HIGH spectral resolution images are now available with hyperspectral satellite sensors, such as the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). Hyperspectral images provide detailed spectral information regarding the physical nature of the materials and thus can be used to distinguish different landscapes. However, developing efficient methods to process hyperspectral images with more than 100 channels is a difficult objective. Furthermore, in the context of supervised classification, the high dimensionality also brings the problem named as "Hughes phenomenon" which will influence the classification performance [1]. In order to solve these problems, feature selection and extraction [2] are known to be important techniques in hyperspectral image classification.

Manuscript received June 7, 2013; revised July 11, 2013; accepted July 25, 2013. This work was supported by the National Natural Science Funds for Distinguished Young Scholar, by the National Natural Science Foundation of China (No. 61172161), by the Hunan Provincial Innovation Foundation for Postgraduate, by the Rector Scholarship of Hunan University for Excellent Doctoral Student, and by the Chinese Scholarship Award for Excellent Doctoral Student.

X. Kang is with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China, and also with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland (e-mail: xudong_kang@163.com).

S. Li is with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: shutao_li@hnu.edu.cn).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland (e-mail: benedikt@hi.is).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2013.2275613

Feature selection aims at finding the best subset of hyperspectral bands that provide the highest class separability [3]–[5]. For example, different types of methods such as distance-based methods [4], [6] and information-theory-based method [7] have been proposed. The feature selection methods have an important advantage, i.e., preserving the physical meaning of the data. Furthermore, Bruzzone *et al.* discuss that the preservation of the spatial information is also quite important for feature selection [8]. As another interesting research topic, instead of performing selection on the hyperspectral image, Pedergnana *et al.* [9] perform selection of features obtained by other feature extraction methods. Although experimental results have shown the effectiveness of feature selection methods, these methods have a general theoretic limitation. Specifically, the "best" subset of features for classifier cannot be found until an exhaustive search of all the feature subset combinations has been carried out, which is an infeasible task because the combination of features to be examined increases exponentially with the number of dimension. Although a subset of features can be selected according to some prior assumptions, it will be not able to know whether the selected subset is indeed the best of all possible subsets.

Besides feature selection, feature extraction is also an effective way to reduce the dimension of the data. Specifically, a hyperspectral image is first projected into another feature space by applying a linear transformation. Then, only the significant components are retained for classification. Different types of dimensionality reduction techniques, including unsupervised approaches such as principal component analysis (PCA) [10] and independent component analysis (ICA) [11], as well as supervised approaches such as linear discriminant analysis [12]–[14], have been proposed. Among these methods, the PCA method can ensure that most information of the hyperspectral image can be preserved in a small amount of significant principal components (PCs), but it cannot ensure that the spectral signatures of interest are emphasized. By contrast, the ICA-based methods ensure that the transformed components are as independent as possible. However, the computational burden can also be increased due to the complexity of the ICA-based methods. Furthermore, another disadvantage of this type of feature extraction is that it just processes each pixel independently without considering the spatial context information.

Recently, many researchers have worked on spectral–spatial classification which can incorporate the spatial contextual information (the strong relationship between neighboring pixels) into the classification process [15]. For example, some spectral–spatial classification methods aim at postprocessing

the classification map obtained by pixelwise classifiers according to the spatial structures of the hyperspectral image [16]. For these types of methods, segmentation and optimization techniques such as watershed [17], minimum spanning forest [18], hierarchical segmentation [19], partitional clustering [20], and Bayesian model [21], [22] have been proposed to refine the classification results. However, most of these methods rely on the performance of automatic segmentation or optimization techniques and thus may be time-consuming.

Besides the methods described previously, an interesting alternative of spectral–spatial classification is integrating the spatial information in the feature extraction process. For example, Benediktsson *et al.* propose that the morphological transforms are able to model the spatial structures in hyperspectral images [23]–[26]. They first perform the PCA on the original hyperspectral image to obtain the significant PCs of the image. Then, the PCs are processed with iterative opening, closing, and reconstruction operations to model the spatial structures of different scales. Their approaches can be improved further by considering the attribute information [24]–[27]. Furthermore, Zhang *et al.* propose that the hyperspectral data can be represented using a tensor representation [28] to make full use of the spatial information. All of the aforementioned studies have verified that combining spectral and spatial information together in the feature extraction process is an effective way in increasing classification accuracy.

Recently, edge-preserving filtering [29]–[32] has been applied successfully in many applications such as high dynamic imaging [31], stereo matching [33], image fusion (IF) [34], [35], dehazing [36], and denoising [37] since it can smooth an image while preserving well its edge structures. In our previous work [38], joint edge-preserving filtering has been successfully applied for the postprocessing of support vector machine (SVM) classification. In this paper, one of the most widely used edge-preserving filters (EPFs), i.e., the transform domain recursive filter [39], is first applied for the feature extraction of hyperspectral images.

The proposed method is based on two simple assumptions: 1) the adjacent bands of the hyperspectral image usually contain redundant information, and 2) the neighboring pixels usually have quite strong correlations with each other. As shown in Fig. 1, the adjacent bands of the hyperspectral image look quite similar. Based on this observation, IF is adopted to combine the complementary information of adjacent bands for feature reduction. One advantage of IF is that it can effectively remove noise and can preserve well the structural information of the image in the fused bands. For the second assumption, transform domain recursive filtering is utilized to ensure that neighboring pixels on the same side of an edge have similar feature values. In other words, spatial context information is also well utilized in the feature extraction process. Experimental results demonstrate the outstanding performance of the proposed IF and recursive filtering (IFRF) based feature extraction method in terms of classification accuracy and computational efficiency. It means that the two basic assumptions, i.e., spectral redundancy and spatial consistency, are both very useful in feature extraction of hyperspectral images.

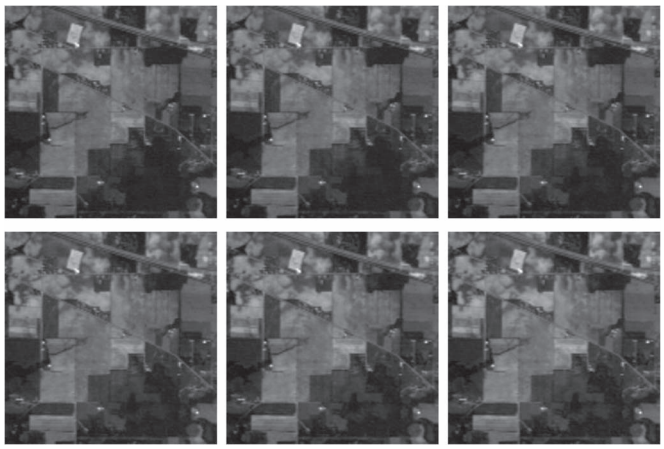

Fig. 1. From left to right and then top to bottom: the 11th to 16th bands of the Indian Pines image (each band has been normalized into the range between 0 and 1).

The remainder of this paper is organized as follows. In Section II, the transform domain recursive filter is reviewed. In Section III, the proposed feature extraction method is described. The experimental part is shown in Section IV, and finally, the conclusion is given in Section V.

II. Transform Domain Recursive Filter

Here, we will briefly describe the transform domain recursive filter. For a complete description of the transform domain recursive filter, we refer the reader to the original work [39]. Transform domain means that the input signal I is first transformed to the transform domain Ω_ω. Intuitively, the transformed coordinate $ct(x_m)$ is computed for each pixel such that the two pixels which lie on the same side of a strong edge have nearby coordinates, while pixels that lie on different sides of a strong edge are far apart. The transformed signal is then processed by recursive filtering as follows:

$$J[m] = (1 - a^b)I[m] + a^b J[m-1] \tag{1}$$

where $J[m]$ is the filtered result; $a = \exp(-\sqrt{2}/\delta_s) \in [0, 1]$ is a feedback coefficient, with δ_s as the spatial parameter; $I[m] = I(x_m)$ is the input discrete signal; and b is the distance between neighbor samples x_m and x_{m-1} in the transform domain (Ω_ω) which is estimated according to $b = ct(x_m) - ct(x_{m-1})$. The function $ct(u)$ which is used to compute the distance b defines the domain transform of a signal $I(x)$ as follows:

$$ct(u) = \int_0^u 1 + \frac{\delta_s}{\delta_r}|I'(x)|\mathrm{d}x, u \in \Omega_\omega \tag{2}$$

where $I'(x)$ is the derivative of the input signal $I(x)$ and δ_s and δ_r are the spatial and range parameters of the EPF, respectively. From (1), it can be seen that, as b increases, a^b will become close to zero, stopping the propagation chain. Therefore, edges are preserved, while pixels on the same side of the edge will tend to have similar filtering outputs. In the 2-D image case, the 1-D filtering operation is separately performed along each dimension of the image iteratively. In other words, 1-D filtering is first performed along each image row and then along each image column. In [39], it has been shown that three iterations

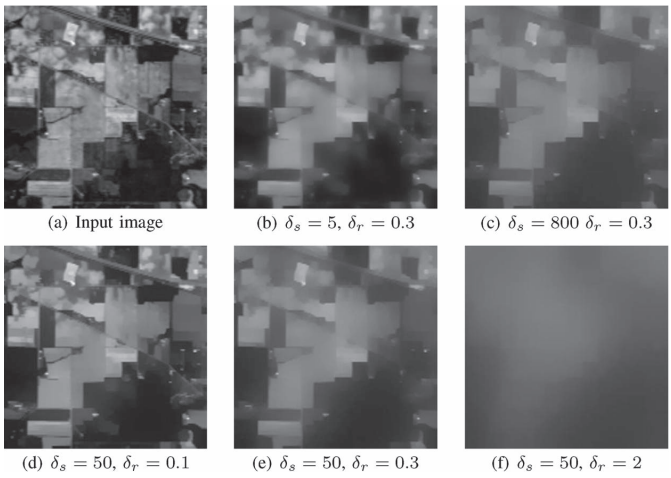

Fig. 2. Influence of the spatial and range standard deviations, i.e., δ_s and δ_r, to the filtering results. (a) Input image. (b)–(f) Filtered images with (b) $\delta_s = 5$ and $\delta_r = 0.3$, (c) $\delta_s = 800$ and $\delta_r = 0.3$, (d) $\delta_s = 50$ and $\delta_r = 0.1$, (e) $\delta_s = 50$ and $\delta_r = 0.3$, and (f) $\delta_s = 5$ and $\delta_r = 2$.

of 1-D filtering are able to obtain satisfactory filtering results for an image. Therefore, three iterations of 1-D filtering are adopted for the recursive filtering used in this paper.

Here, the influence of the two parameters δ_s and δ_r on the filtering results is analyzed in Fig. 2. From Fig. 2(b)–(e), it can be seen that the recursive filter can effectively remove the texture information and can also preserve the strong edge structures. As δ_r and δ_s increase, a more obvious smoothing effect will be produced on the filtering outputs. Moreover, when δ_r becomes relatively large, i.e., $\delta_r = 2$, the filtering output will tend to be extremely smooth, and only little useful information is then preserved. By contrast, when δ_s tends to approach infinity, e.g., $\delta_s = 800$, the recursive filter will not produce unbounded smoothing of the image. The reason has been clearly analyzed in [39]. In Section IV-B, the influence of the two parameters on the classification performance of the proposed method will be analyzed further.

III. Proposed Approach

As shown in Fig. 3, the proposed feature extraction and classification approach consists of four steps: 1) partition the hyperspectral image into multiple subsets of adjacent bands; 2) fuse the adjacent bands in each subset; 3) perform recursive filtering on the fused bands; and 4) perform classification on the filtered images. It can be seen that the feature reduction step with IF is before the recursive filtering step. The reason is that performing recursive filtering before the fusion means that each hyperspectral band should be filtered, which will be much more time-consuming.

1) Band partitioning: the hyperspectral image is spectrally partitioned into K subsets of hyperspectral bands. The kth ($k \in (1, \ldots, K)$) subset can be obtained as follows:

$$\mathbf{P}^k = \begin{cases} (\mathbf{x}_k, \ldots, \mathbf{x}_{(k+\lfloor D/K \rfloor)}), & \text{if } k + \lfloor D/K \rfloor \leq D \\ (\mathbf{x}_k, \ldots, \mathbf{x}_D), & \text{if } k + \lfloor D/K \rfloor > D \end{cases} \quad (3)$$

where $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_D) \in \mathbb{R}^{D \times J}$ denotes the original hyperspectral image with D-dimensional feature vectors and J pixels, and $\lfloor D/K \rfloor$ represents the floor operation which calculates the largest integer not greater than D/K.

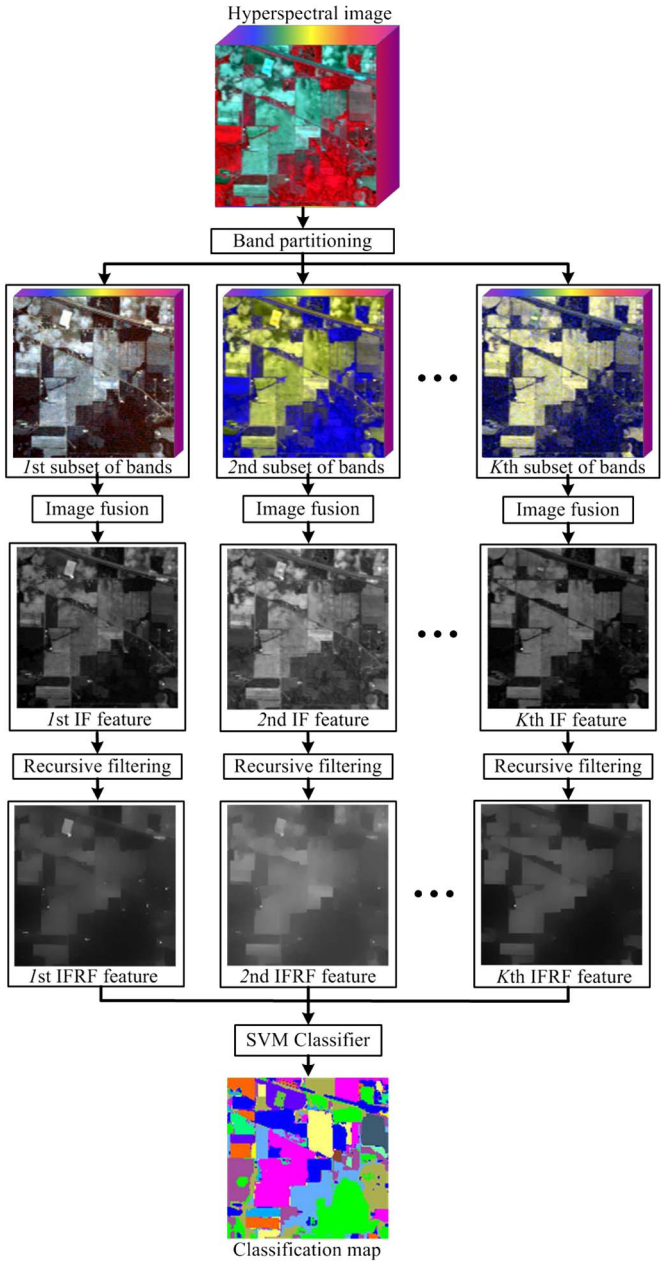

Fig. 3. Schematic of the proposed IFRF-based classification method.

2) IF: the adjacent bands in the kth subset are fused by one of the most simple IF methods, i.e., the averaging method. Specifically, the kth fused band, i.e., the kth IF feature F_k, is calculated as follows:

$$Q^k = \frac{\sum_{n=1}^{N_k} \mathbf{P}_n^k}{N_k} \quad (4)$$

where $\mathbf{P}_n^k$ refers to the nth band of the kth subset of hyperspectral bands and N_k refers to the total number of bands in the kth subset. This step actually calculates the average image of each subset which aims at removing the noisy pixels and the redundant information for each subset.

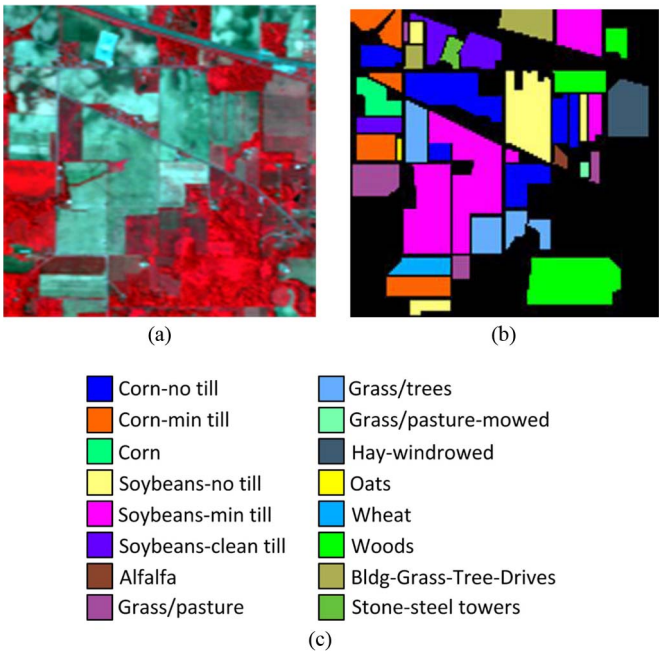

Fig. 4. Indian Pines data set. (a) Three-band color composite of the Indian Pines image. (b) and (c) Reference data for the Indian Pines image.

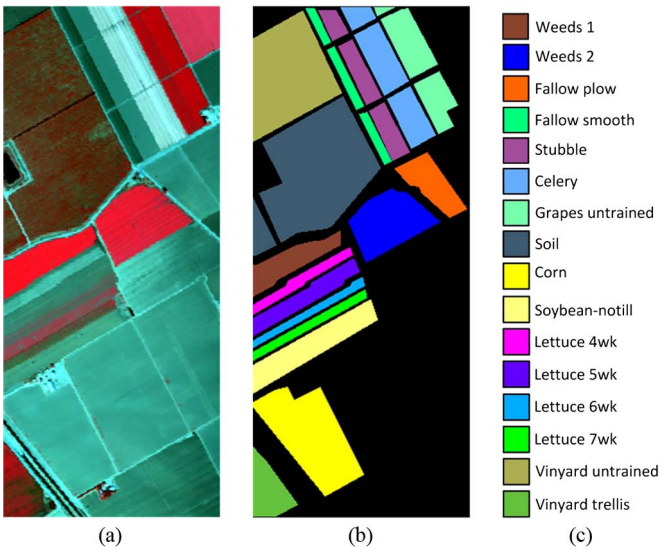

Fig. 5. Salinas data set. (a) Three-band color composite of the Salinas image. (b) and (c) Reference data for the Salinas image.

3) Recursive filtering: transform domain recursive filtering is performed on each fused band to obtain the kth feature

$$O_i^k = RF_{\delta_s,\delta_r}\left(Q_i^k\right) \tag{5}$$

where RF represents the transform domain recursive filtering operation, δ_s and δ_r are the spatial and range standard deviations of the filter [39], and $\mathbf{O} = (O^1, \ldots, O^K) \in \mathbb{R}^{K\times J}$ is the resulting feature image obtained by IFRF.

4) Classification: the SVM classifier is utilized for the classification of the IFRF features. The SVM classifier is one of the most widely used pixelwise classifiers and has, in particular, shown a good performance in terms of classification accuracy [40]. Furthermore, the SVM classifier has a major advantage, i.e., robust to the dimension of data sets [40]. Therefore, in this situation, the classification result of the SVM method for an original data set can be considered in providing the best pixelwise classification results [11].

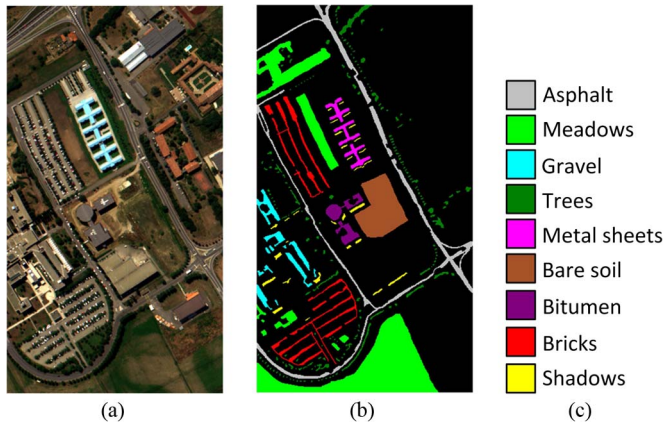

Fig. 6. University of Pavia data set. (a) Three-band color composite of the University of Pavia image. (b) and (c) Reference data for the University of Pavia image.

IV. Experiments

A. Experimental Setup

1) Data Sets: Three hyperspectral data sets, i.e., the Indian Pines image, the Salinas image, and the University of Pavia image, are utilized in our experiments. The Indian Pines image is for the agricultural Indian Pine test site of Northwestern Indiana. It was acquired by the AVIRIS sensor. The image has 220 bands of size 145 × 145, with a spatial resolution of 20 m per pixel and a spectral coverage ranging from 0.4 to 2.5 μm. Twenty water absorption bands (no. 104–108, 150–163, and 220) were removed. The color composite of the Indian Pines image and the corresponding ground truth data are shown in Fig. 4.

The Salinas image was captured by the AVIRIS sensor over Salinas Valley, CA, USA. The image has 224 bands and is of size 512 × 217, with a spatial resolution of 3.7 m per pixel. Similar to the Indian Pines image, 20 water absorption bands (no. 108–112, 154–167, and 224) were discarded. The color composite of the Salinas image and the corresponding ground truth data are shown in Fig. 5.

The University of Pavia image is from an urban area surrounding the University of Pavia, Pavia, Italy. It was recorded by Reflective Optics System Imaging Spectrometer with a spatial resolution of 1.3 m per pixel and a spectral coverage ranging from 0.43 to 0.86 μm. The image has 115 bands of size 610 × 340. The 12 most noisy channels were removed before experiments. The color composite of the University of Pavia image and the corresponding ground truth data are shown in Fig. 6. For the experiments of this paper, the number of training and test samples for each class used for classification is detailed in Tables I–III.

2) Quality Indexes: In order to evaluate the performance of the proposed method, three widely used quality indexes, i.e., overall accuracy (OA), average accuracy (AA), and kappa

TABLE I
CLASSIFICATION ACCURACIES (IN PERCENT) FOR THE SVM [40], SVMPCA [10], SVMICA [11], IF, EMP [26], LRML [21], EPF [38], AND IFRF METHODS (INDIAN PINES IMAGE)

Class	Train	Test	Pixel-wise methods				Spectral-spatial methods			
			SVM	SVMPCA	SVMICA	IF	EMP	LMLL	EPF	IFRF
Alfalfa	23	23	60.6	**91.7**	84.0	64.7	85.2	**100**	**100**	95.8
Corn_N	79	1349	71.0	74.9	70.9	**80.3**	84.4	88.0	96.0	**98.2**
Corn_M	81	749	71.1	63.0	64.7	**79.7**	87.4	87.6	96.9	**99.3**
Corn	66	171	45.2	50.8	55.6	**58.0**	78.1	**100**	69.8	94.5
Grass_M	71	412	89.0	86.4	84.7	**89.1**	92.6	97.6	98.5	**99.5**
Grass_T	78	652	93.6	**95.4**	95.0	93.3	96.9	**100**	98.3	97.5
Grass_P	15	13	92.3	80.0	**92.9**	92.3	70.6	**100**	**100**	**100**
Hay_W	72	406	98.3	**99.5**	**99.5**	99.0	99.8	**100**	**100**	**100**
Oats	10	10	**71.4**	34.8	42.9	55.6	71.4	**100**	**100**	**100**
Soybean_N	79	893	73.0	**75.6**	70.9	75.1	85.7	90.7	92.1	**96.6**
Soybean_M	111	2344	82.1	76.9	74.8	**88.0**	90.4	92.8	93.3	**98.8**
Soybean_C	74	519	75.0	73.4	77.0	**83.6**	94.4	98.1	93.3	**99.0**
Wheat	64	141	92.2	92.8	**97.9**	90.4	97.9	**100**	**100**	97.9
Woods	84	1181	95.6	**96.2**	95.7	95.6	99.9	96.4	99.9	**100**
Buildings	70	316	54.3	63.7	**71.7**	47.7	99.0	**100**	84.7	97.5
Stone	47	46	**95.7**	77.2	91.7	**95.7**	97.8	**100**	97.9	90.2
OA	-	-	79.30	79.05	78.51	**83.43**	91.30	93.82	94.70	**98.42**
AA	-	-	78.76	77.02	79.37	**80.50**	89.47	96.94	95.05	**97.80**
Kappa	-	-	76.33	76.02	75.37	**81.05**	89.99	92.89	93.90	**98.25**

TABLE II
CLASSIFICATION ACCURACIES (IN PERCENT) FOR THE SVM [40], SVMPCA [10], SVMICA [11], IF, EMP [26], LRML [21], EPF [38], AND IFRF METHODS (SALINAS IMAGE)

Class	Train	Test	Pixel-wise methods				Spectral-spatial methods			
			SVM	SVMPCA	SVMICA	IF	EMP	LMLL	EPF	IFRF
Weeds_1	68	1941	99.8	99.9	**100**	99.9	**100**	99.6	**100**	**100**
Weeds_2	68	3658	99.2	**99.9**	**99.9**	99.1	99.4	**100**	**100**	**100**
Fallow	67	1909	94.6	**97.6**	97.5	95.1	96.0	**99.8**	99.1	**99.8**
Fallow_P	68	1326	98.7	**99.5**	99.4	99.4	98.9	**99.8**	98.4	98.6
Fallow_S	67	2611	**99.4**	97.7	97.2	98.9	99.8	98.5	**100**	99.9
Stubble	67	3892	99.9	**100**	**100**	99.8	99.9	99.7	**100**	**100**
Celery	67	3512	99.8	**100**	99.9	98.3	99.6	99.5	**100**	99.9
Grapes	70	11201	78.7	81.6	**82.3**	81.1	92.4	89.3	94.2	**99.3**
Soil	69	6134	**99.4**	99.3	99.1	**99.4**	99.2	99.6	99.6	**99.9**
Corn	67	3211	85.1	94.9	**96.6**	92.2	91.2	94.0	92.4	**99.8**
Lettuce_4	67	1001	94.8	95.9	**96.0**	89.9	98.5	99.0	**100**	97.8
Lettuce_5	68	1859	97.6	**98.0**	97.1	97.7	98.7	**100**	**100**	99.7
Lettuce_6	67	849	95.7	**97.2**	96.7	94.6	97.9	98.6	**100**	99.9
Lettuce_7	67	1003	**96.2**	94.6	90.0	93.1	98.5	94.5	**98.7**	98.6
Vinyard_U	68	7200	59.6	66.1	**67.4**	64.4	84.8	76.8	80.8	**100**
Vinyard_T	67	1740	92.7	97.8	**98.8**	97.8	88.7	99.5	99.4	**99.9**
OA	-	-	87.75	90.34	**90.64**	89.40	94.88	93.90	95.29	**99.69**
AA	-	-	93.20	95.00	**94.86**	93.78	96.47	96.77	97.66	**99.57**
Kappa	-	-	86.39	89.24	**89.57**	88.21	94.30	93.16	94.76	**99.66**

TABLE III
CLASSIFICATION ACCURACIES (IN PERCENT) FOR THE SVM [40], SVMPCA [10], SVMICA [11], IF, EMP [26], LRML [21], EPF [38], AND IFRF METHODS (UNIVERSITY OF PAVIA IMAGE)

Class	Train	Test	Pixel-wise methods				Spectral-spatial methods			
			SVM	SVMPCA	SVMICA	IF	EMP	LMLL	EPF	IFRF
Asphalt	190	6641	**97.4**	95.4	95.6	97.2	**99.2**	93.7	98.8	98.0
Meadows	191	18458	**96.8**	95.8	96.5	96.6	98.7	98.2	99.7	**99.8**
Gravel	190	1909	**73.0**	68.5	67.5	72.1	97.8	84.9	95.8	**98.4**
Trees	190	2874	**89.7**	83.2	86.0	**89.7**	97.5	96.5	**99.7**	98.9
Sheets	190	1155	98.4	**100**	**100**	97.5	87.2	**99.9**	99.1	**99.9**
Soil	190	4839	72.0	71.5	72.9	**73.0**	79.8	99.8	94.0	**99.9**
Bitumen	190	1140	**73.2**	58.0	59.0	72.9	97.5	97.8	**100**	97.5
Bricks	190	3492	83.6	78.2	77.7	**84.2**	**98.6**	95.9	92.5	97.1
Shadows	190	757	**100**	99.9	99.7	99.9	**100**	**100**	99.5	97.2
OA	-	-	**91.12**	86.80	87.46	89.54	95.61	96.83	97.99	**99.05**
AA	-	-	**88.08**	83.37	83.88	87.14	95.13	96.75	97.67	**98.51**
Kappa	-	-	**88.23**	82.60	83.48	86.19	94.17	93.12	97.30	**98.72**

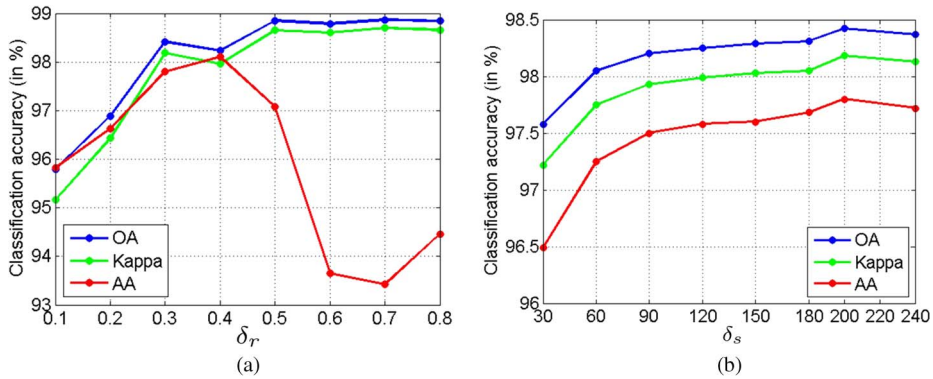

Fig. 7. Indian Pines image. Analysis of the influence of the parameters δ_s and δ_r.

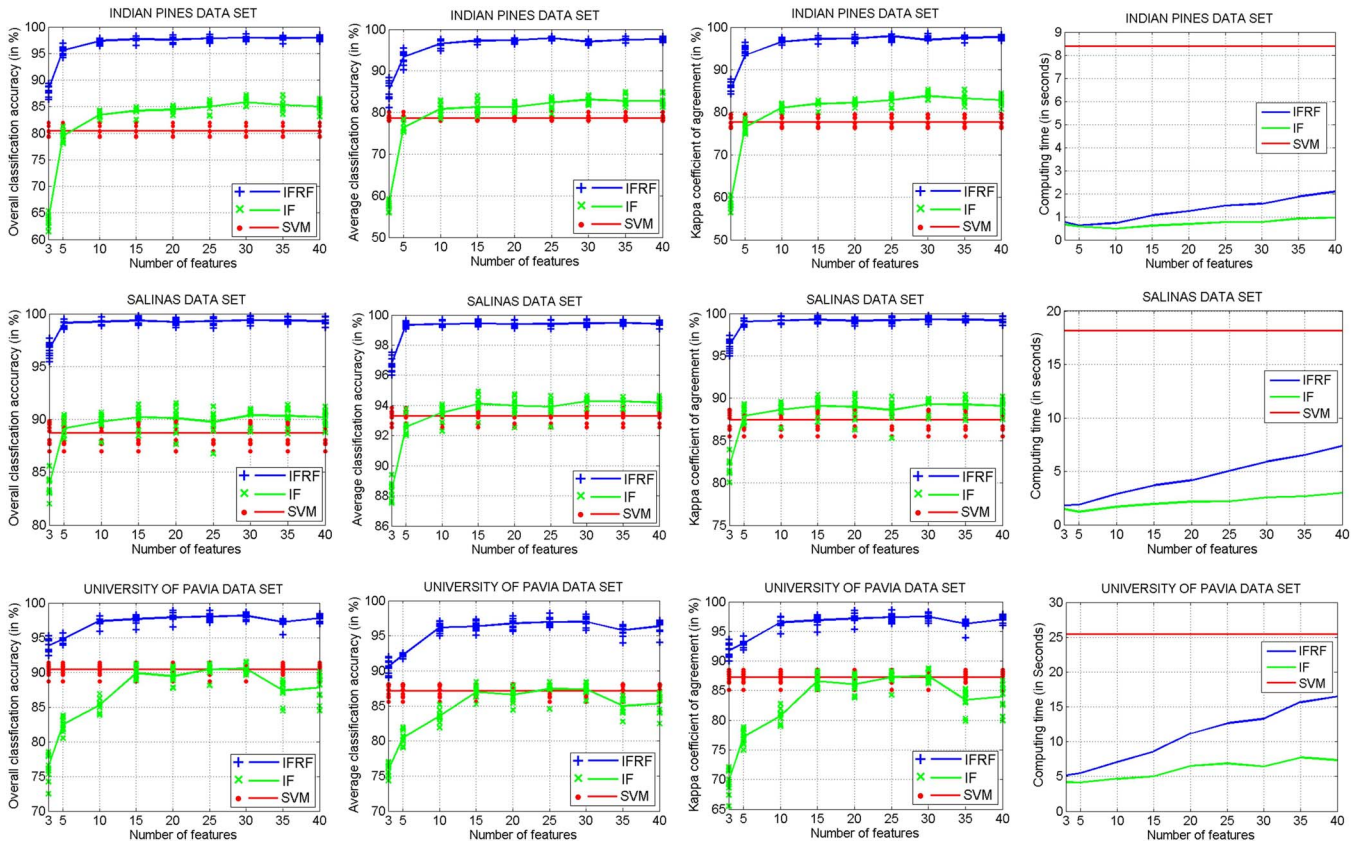

Fig. 8. Comparison of (first column) overall classification accuracy, (second column) average classification accuracy, (third column) kappa coefficient of agreement, and (fourth column) computing time obtained with SVM (full feature space), IF, and IFRF (with respect to different numbers of features). The rows from top to bottom, respectively, correspond to the Indian Pines, Salinas, and University of Pavia data sets.

coefficient [41], are adopted. Among the three metrics, OA measures the percentage of pixels that are correctly classified. AA refers to the mean of the percentage of correctly classified pixels for each class. In order to make the measurement more objective, the kappa coefficient estimates the percentage of correctly classified pixels corrected by the number of agreements that would be expected purely by chance.

B. Classification Results

1) Analysis of the Influence of Parameters: For the proposed IFRF-based feature extraction technique, the parameters δ_s and δ_r of the recursive filter need to be determined. The influence of the two parameters on the classification performance is analyzed in an experiment on the Indian Pines image (see Fig. 7). The training set which accounts for 10% of the ground truth was chosen randomly. The number of training and test samples for each class is detailed in Table I. The OA, AA, and kappa of the proposed method are measured with different parameter settings. When the influence of δ_s is analyzed, δ_r is fixed at 0.3. Similarly, δ_r is analyzed in the same way with δ_s fixed at 200. From Fig. 7, it can be seen that, when δ_r is very large, the average classification accuracy will decrease dramatically. The reason is that recursive filtering with a large δ_r will oversmooth some useful edge features. Therefore, some objects will not be classified accurately. For example, although the OA obtained with $\delta_r = 0.6$ is larger than the accuracy obtained with $\delta_r = 0.3$ [see Fig. 7(a)], the AA decreases dramatically when $\delta_r = 0.6$

because some classes are totally misclassified when δ_r is too large. Similarly, small values of δ_s and δ_r are also not good for the proposed method because it means that only very limited local spatial information is considered in the feature extraction process. In this paper, the default parameter setting of the proposed method is set as $\delta_s = 200$ and $\delta_r = 0.3$. The following experimental results show that good classification accuracies are obtained for different images with the default parameter setting.

2) Influence of the Number of IFRF Features: When applying the proposed method, the number of features has an influence on both the final classification accuracy and the computational efficiency. In this section, the influence of the number of features is analyzed in Fig. 8. The IF represents the proposed method with only IF. The IFRF refers to the proposed method with IFRF. Experiments are performed on three different data sets, i.e., the Indian Pines, Salinas, and University of Pavia data sets, respectively. The training samples are selected randomly, which account for 10% (Indian Pines), 2% (Salinas), and 4% (University of Pavia) of the reference data. Furthermore, experiments have been repeated ten times to estimate the mean value of the OA, AA, and kappa coefficient. From Fig. 8, it can be seen that, if the number of features is quite small, i.e., $k = 3$, the performance of the IFRF method will not be satisfactory. The reason is that, when the number of features is very small, the hyperspectral bands in each subset will contain a large amount of complementary information. In this situation, useful discriminative information will be lost in the IF process. From Fig. 8, it is observed that the proposed IFRF method can always obtain satisfactory classification results when the number of features is larger than 10. Furthermore, from the fourth column of Fig. 8, it can be seen that the computing time of the IFRF method increases linearly when the number of features is increasing. Therefore, in this paper, $k = 20$ is set to be the default number of features because it can give good classification accuracies and a relatively low computing burden. Furthermore, it can be seen that, although the number of training samples is fixed, the classification accuracy of the SVM and IF methods will vary in a small range (2%–5%) with different training samples obtained by random selection. Compared with the SVM and IF methods, the dynamic range of the accuracies obtained by the IFRF method is narrower (1%–3%), which means that the performance of the IFRF method is more stable.

3) Comparison of Different Classification Methods: In this section, the proposed IFRF method is compared with several widely used classification methods including the SVM method [40], the PCA-based SVM classification method (SVMPCA) [10], the ICA-based SVM classification method (SVMICA) [11], the extended morphological profile (EMP) based method [26], the logistic regression and multilevel logistic (LRML) based method [21], and the EPF-based method [8]. The SVM algorithm is implemented in the LIBSVM library [42] by using the Gaussian kernel with fivefold cross-validation. For the SVMPCA and SVMICA methods, 20 PCs or independent components (ICs) obtained by the ENVI software[1] are used as the inputs of the SVM classifier. In [11], it has been demonstrated that 20 ICs are able to obtain the best results for most of the test hyperspectral images such as the Indian Pines image and the University of Pavia image. For the PCA method, we also find that 20 PCs are also able to obtain nearly the best accuracies for SVM classifier. For the EMP method, in order to obtain very good accuracies for different images, the morphological profiles are constructed with the first three

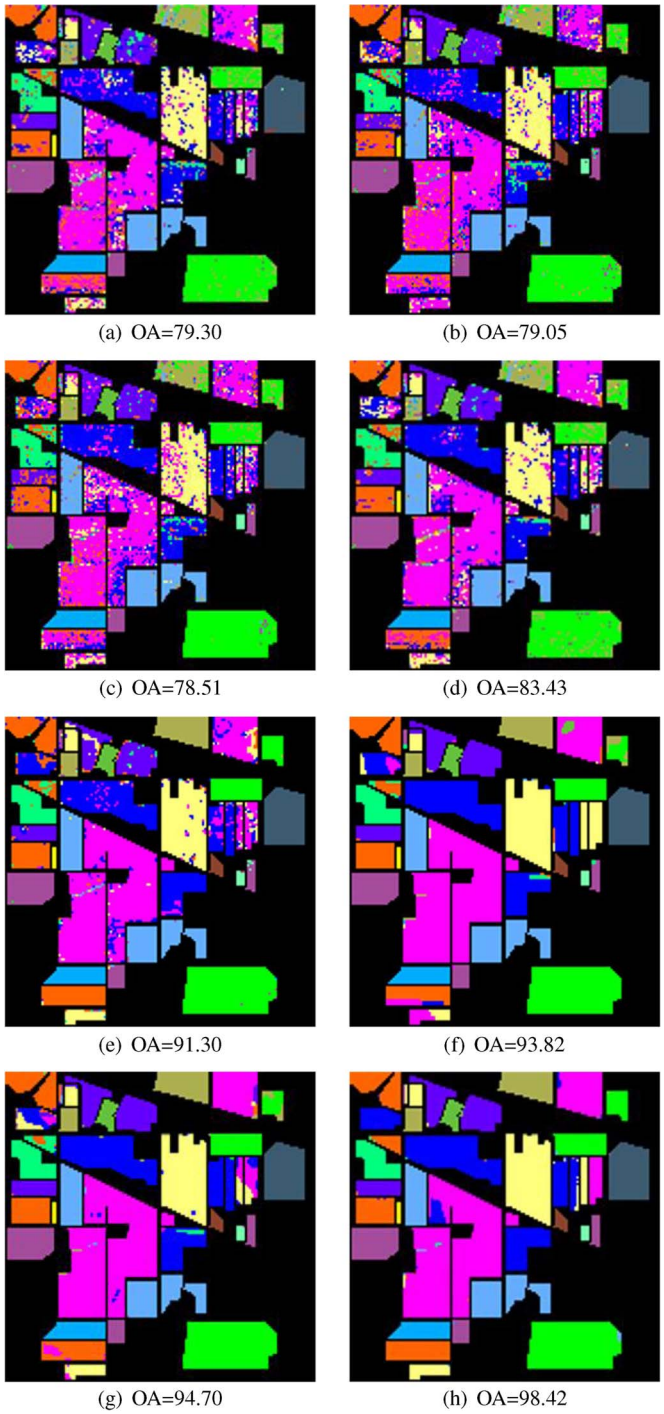

Fig. 9. Classification results (Indian Pines image) obtained by (a) the SVM method (full feature space), (b) the SVMPCA method, (c) the SVMICA method, (d) the IF method, (e) the EMP method, (f) the LRML method, (g) the EPF method, and (h) the IFRF method. The value of OA is given in percent.

[1]http://www.exelisvis.com/

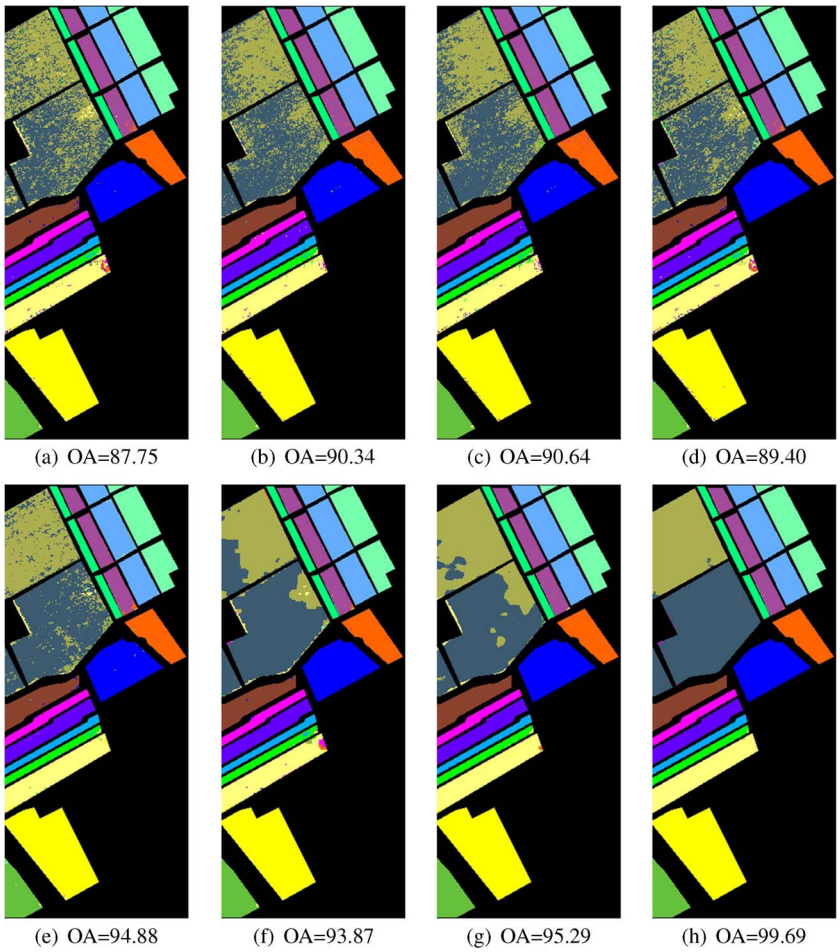

Fig. 10. Classification results (Salinas image) obtained by (a) the SVM method (full feature space), (b) the SVMPCA method, (c) the SVMICA method, (d) the IF method, (e) the EMP method, (f) the LRML method, (g) the EPF method, and (h) the IFRF method. The value of OA is given in percent.

PCs, a circular structural element, a step size increment of two, and four openings and closings for each PC. The code of the LRML method is available on Dr. Li's homepage.[2] The default parameters given in their implementation are adopted. Similarly, for the EPF method, the default parameters given in [38] are adopted. For the IFRF method, 20 IFRF features are used for classification. Furthermore, the default parameters given in the first subsection of Section IV are adopted for all images.

Experiment is first performed on the Indian Pines data set. Table I presents the number of training and test samples (the training set which accounts for 10% of the ground truth was chosen randomly). Fig. 9 shows the classification results obtained by different methods of the Indian Pines image associated with the corresponding OA values. As shown in this figure, compared with other pixelwise methods such as the SVMPCA and SVMICA methods, the proposed IF method performs better in terms of the highest classification accuracy. It means that the proposed method can both reduce the dimension and preserve well the useful information of the hyperspectral data. By combining the spatial information into feature extraction, the EMP method can effectively improve the classification accuracy of the SVM classier. However, some noisy pixels are still visible in the obtained classification results. By contrast, the LRML method [21], the EPF method [38], and the proposed IFRF method perform much better in removing "noisy pixels." Specifically, the proposed method increases the OA compared to the SVM method by about 19%. Compared with the LRML method and the EPF methods, the proposed IFRF method gives a much higher classification accuracy. Furthermore, Table I shows the classification accuracies of different methods. From this table, it can be observed that, by using the proposed IFRF method, the AA of SVM increased from 79.30% to 98.42% and the Kappa accuracy also increased significantly. More importantly, since the proposed method can reduce the dimension of the data into 20 channels, it also performs much more efficiently than the LRML and EPF methods which are developed as postprocessing methods of the pixelwise classifiers.

The second and third experiments are performed on the Salinas images and the University of Pavia image, respectively. Tables II and III present the number of training and test samples (for the Salinas image and the University of Pavia image, the training sets which, respectively, account for 2% and 4% of the ground truth were chosen randomly). Figs. 10 and 11 show

[2]http://www.lx.it.pt/~jun/

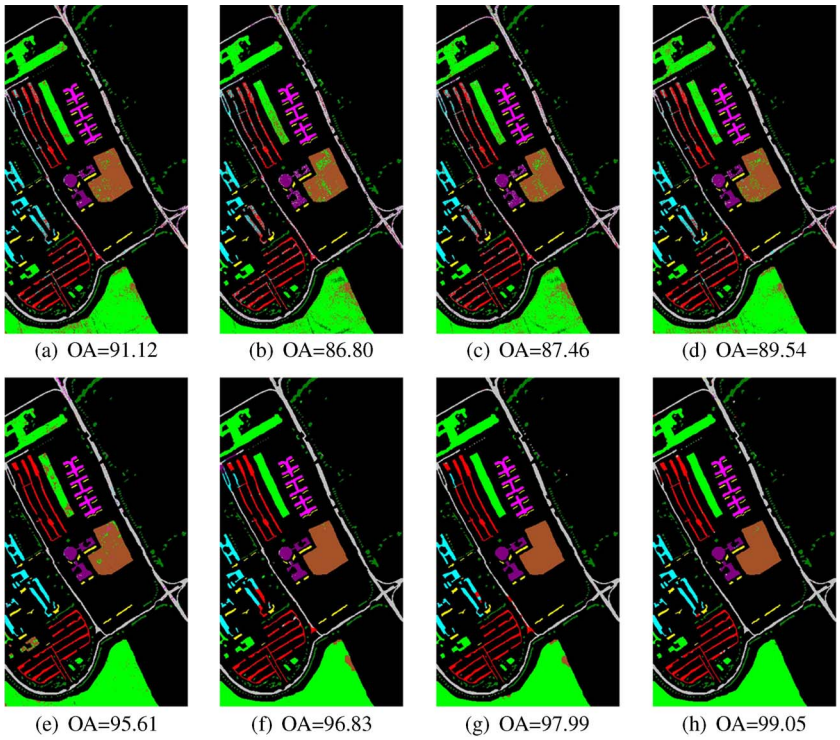

Fig. 11. Classification results (University of Pavia image) obtained by (a) the SVM method (full feature space), (b) the SVMPCA method, (c) the SVMICA method, (d) the IF method, (e) the EMP method, (f) the LRML method, (g) the EPF method, and (h) the IFRF method. The value of OA is given in percent.

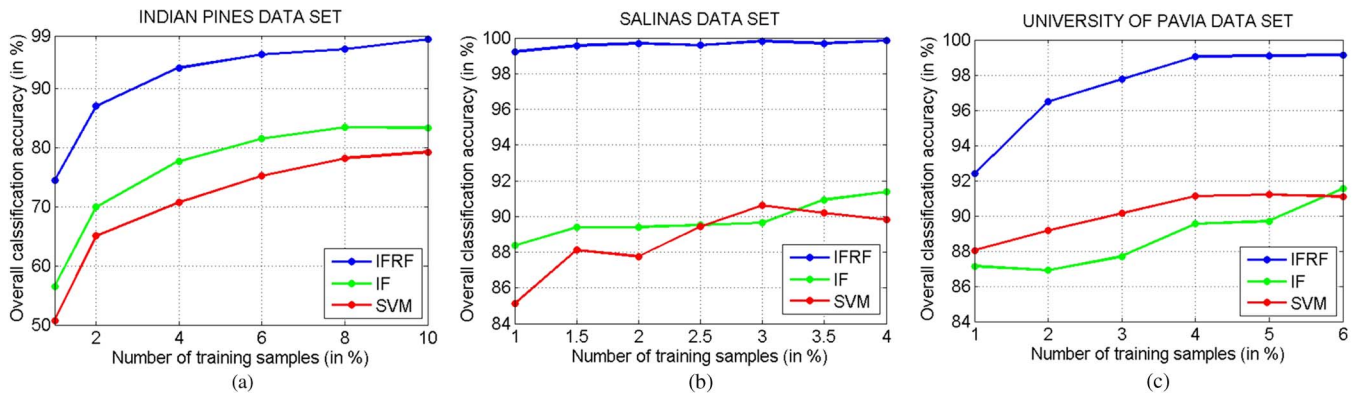

Fig. 12. OA of the proposed IFRF method with different numbers of training samples on different images. (a) Indian Pines. (b) Salinas. (c) University of Pavia.

the classification results obtained by different methods associated with the corresponding OAs. As shown in Fig. 10, only the proposed IFRF method is able to accurately discriminate the vinyard_U class and the soil class [the two classes with the largest area at the top-left of Fig. 10(h)]. Tables II and III show the classification accuracies for different methods. From the two examples, it can be seen that the proposed IFRF method always outperforms the EMP, LRML, and EPF methods in terms of OA, AA, and kappa. Compared with the SVM method, the proposed method can improve the classification accuracies significantly. For example, in Table II, the classification accuracy of the vinyard_U class increases from 59.6% to 100%. Similar improvements can be found in the experimental results of the University of Pavia example. As shown in Table III, the classification accuracy of the gravel class can be increased from 73.0% to 98.4%. The two sets of results are presented in Figs. 10 and 11 and further demonstrate the advantage of the proposed method.

4) Classification Results With Different Training and Test Sets: In this section, the influence of different training and test sets to the performance of the proposed method is analyzed. Experiments are performed on three images, i.e., the Indian Pines image, the Salinas image, and the University of Pavia image. Fig. 12 shows the classification results of the IFRF method with the number of training samples (in percent) increased from 1% to 10% for the Indian Pines image, from 1% to 4% for the Salinas image, and from 1% to 6% for the University of Pavia image. From this figure, it can be seen that the proposed method can always improve the classification accuracy significantly with different numbers of training samples. For example, regarding the Indian Pines image, when the OA of SVM is about 70% (4% ground truth samples are used as training samples),

TABLE IV
COMPUTING TIME (IN SECONDS) OF THE PROPOSED FEATURE EXTRACTION ALGORITHM

Images	India Pines		Salinas		University of Pavia	
Methods	IF	IFRF	IF	IFRF	IF	IFRF
Time	0.06	0.63	0.30	2.59	0.42	5.07

the IFRF method can obtain a classification accuracy near 94%. For the Salinas image, it can be seen that, with relatively limited training samples (1% of the ground truth), the proposed method still can obtain an OA near 99%.

5) Computational Complexity: Here, experiments are performed using MATLAB on a Laptop with 2.5-GHz CPU and 4-GB Memory. Table IV shows the computing time of the proposed IF and IFRF methods. From this table, it can be seen that the MATLAB implementation of the proposed method is already very fast (the IFRF method takes only 0.63 s for the Indian Pines image). The reason is that the proposed method only requires K averaging operations and recursive filtering to be done K times. The complexity of the recursive filtering operation is $O(N)$, and thus, the proposed method has a computational complexity of $O(KN)$, where N is the number of pixels and K is the number of features. As described in [39], the recursive filter can be easily implemented in real-time with C++ programming. It means that the proposed IFRF method will be quite suitable for real applications since it is indeed computationally efficient.

V. CONCLUSION

In this paper, a new approach for hyperspectral image feature extraction, the IFRF, has been proposed. The proposed approach is based on the application of IF to reduce the dimension of the data, the use of recursive filtering to combine spatial information into the resulting IFRF features. Experiments have been carried out on three different real hyperspectral images. The results of the experiments showed the effectiveness of the proposed method, which provided better results than those of the widely used pixelwise classifiers and the spectral–spatial classifiers. Moreover, the proposed method has presented several other advantages: 1) the feature can well preserve the physical meaning of the hyperspectral data. In other words, the pixel values in the feature image still reflect the spectral response of a pixel in a specific spectral range; 2) it is time efficient since it is based on a very fast EPF; and 3) although the classification accuracy obtained by the IFRF is influenced by the number of features and the parameters of the recursive filter, these choices are not critical. The reason is that there is a large region around the optimal number of features for which the proposed method has similar results and outperforms other classification methods in terms of accuracy. Further developments of this work include a comprehensive research of the adoption of other EPFs to process the fused bands and an investigation of the possibility of considering the correlation degree of adjacent bands in the band-partitioning process.

ACKNOWLEDGMENT

The authors would like to thank the Editor-in-Chief, the anonymous Associate Editor, and the reviewers for their insightful comments and suggestions which have greatly improved this paper and M. Pedergnana and Dr. J. Li for providing the software of the EMP and LRML methods.

REFERENCES

[1] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.
[2] X. Jia, B.-C. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proc. IEEE*, vol. 101, no. 3, pp. 676–697, Mar. 2013.
[3] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2297–2307, May 2010.
[4] R. Huang and M. He, "Band selection based on feature weighting for classification of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 156–159, Apr. 2005.
[5] S. B. Serpico and G. Moser, "Extraction of spectral channels from hyperspectral images for classification purposes," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 484–495, Feb. 2007.
[6] S. Backer, P. Kempeneers, W. Debruyn, and P. Scheunders, "A band selection technique for spectral classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 3, pp. 319–323, Jul. 2005.
[7] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson, "Band selection for hyperspectral image classification using mutual information," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 522–526, Oct. 2006.
[8] L. Bruzzone and C. Persello, "A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization capability," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 9, pp. 3180–3191, Sep. 2009.
[9] M. Pedergnana, P. R. Marpu, M. D. Mura, J. A. Benediktsson, and L. Bruzzone, "A novel technique for optimal feature selection in attribute profiles based on genetic algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3514–3528, Jun. 2013.
[10] S. Prasad and L. Mann Bruce, "Limitations of principal components analysis for hyperspectral target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 625–629, Oct. 2008.
[11] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.
[12] W. Li, S. Prasad, J. E. Fowler, and L. Mann Bruce, "Locality-preserving discriminant analysis in kernel-induced feature spaces for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 894–898, Sep. 2011.
[13] W. Li, S. Prasad, J. E. Fowler, and L. Mann Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.
[14] W. Liao, A. Pizurica, P. Scheunders, W. Philips, and Y. Pi, "Semisupervised local discriminant analysis for feature extraction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 184–198, Jan. 2013.
[15] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. S1, pp. S110–S122, Sep. 2009.
[16] G. Moser and S. B. Serpico, "Combining support vector machines and Markov random fields in an integrated framework for contextual image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2734–2752, May 2013.
[17] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, Jul. 2010.
[18] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 5, pp. 1267–1279, Oct. 2010.
[19] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral–spatial classification approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.
[20] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[21] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011.
[22] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.
[23] J. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.
[24] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, Mar. 2005.
[25] D. M. Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542–546, May 2011.
[26] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
[27] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A spatial–spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognit.*, vol. 45, no. 1, pp. 381–392, Jan. 2012.
[28] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Tensor discriminative locality alignment for hyperspectral image spectral–spatial feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 242–256, Jan. 2013.
[29] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. Int. Conf. Comput. Vis.*, Jan. 1998, pp. 839–846.
[30] S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach," *Int. J. Comput. Vis.*, vol. 81, no. 1, pp. 24–52, Jan. 2009.
[31] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 67:1–67:10, Aug. 2008.
[32] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[33] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 504–511, Feb. 2013.
[34] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.
[35] S. Li and X. Kang, "Fast multi-exposure image fusion with median filter and recursive filter," *IEEE Trans. Consum. Electron.*, vol. 58, no. 2, pp. 626–632, May 2012.
[36] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
[37] C. H. Lin, J. S. Tsai, and C. T. Chiu, "Switching bilateral filter with a texture/noise detector for universal noise removal," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2307–2320, Sep. 2010.
[38] X. Kang, S. Li, and J. A. Bendiktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, Jun. 2014, to be published. [Online]. Available: http://ieeexplore.ieee.org
[39] E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 69:1–69:12, Jul. 2011.
[40] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
[41] J. A. Richards, *Remote Sensing Digital Image Analysis*. New York, NY, USA: Springer-Verlag, 2012.
[42] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, Apr. 2011.

Xudong Kang (S'13) received the B.Sc. degree from Northeast University, Shenyang, China, in 2007. He is currently working toward the Ph.D. degree in electrical engineering at Hunan University, Changsha, China.

In 2012–2013, he is a visiting Ph.D. student in electrical engineering at the University of Iceland, Reykjavik, Iceland. He is engaged in image fusion, image superresolution, pansharpening, and hyperspectral image classification.

Shutao Li (M'07) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

In 2001, he joined the College of Electrical and Information Engineering, Hunan University. From May 2001 to October 2001, he was a Research Associate with the Department of Computer Science, Hong Kong University of Science and Technology, Kowloon, Hong Kong. From November 2002 to November 2003, he was a Postdoctoral Fellow with the Royal Holloway College, University of London, Egham, U.K., working with Prof. J.-S. Taylor. From April 2005 to June 2005, he was a Visiting Professor with the Department of Computer Science, Hong Kong University of Science and Technology. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He is the author or coauthor of more than 160 refereed papers. His professional interests are information fusion, pattern recognition, and image processing.

Jón Atli Benediktsson (S'84–M'90–SM'99–F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984 and the M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

He is currently the Pro Rector of Academic Affairs and a Professor of electrical and computer engineering with the University of Iceland. His research interests are in remote sensing, image analysis, pattern recognition, biomedical analysis of signals, and signal processing, and he has published extensively in these fields. He is a cofounder of the biomedical startup company Oxymap.

Prof. Benediktsson was the 2011–2012 President of the IEEE Geoscience and Remote Sensing Society (GRSS), and he has been on the GRSS Administrative Committee since 2000. He is a Fellow of SPIE and a member of Societas Scientiarum Islandica and Tau Beta Pi. He was the Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2003 to 2008, and he has served as an Associate Editor of TGRS since 1999 and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003. He received the Stevan J. Kristof Award from Purdue University in 1991 as outstanding graduate student in remote sensing. In 1997, he was the recipient of the Icelandic Research Councils Outstanding Young Researcher Award; in 2000, he was granted the IEEE Third Millennium Medal; in 2004, he was a corecipient of the University of Iceland's Technology Innovation Award; in 2006, he received the yearly research award from the Engineering Research Institute of the University of Iceland; and in 2007, he received the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society. He was a corecipient of the 2012 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Best Paper Award and the corecipient of the GRSS Highest Impact Paper Award in 2013. In 2013, he received the joint IEEE/VFI Award as Electrical Engineer of the Year.