A NOVEL LOCAL BINARY PATTERNS AND WAVELET TRANSFORM-BASED APPROACH FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Andreia Valentina MICLEA, Romulus TEREBES
Technical University of Cluj-Napoca, Communications Department, Cluj-Napoca, Romania
Andreia.Miclea@com.utcluj.ro, Romulus.Terebes@com.utcluj.ro

Abstract: In this paper we propose a novel approach for the classification of hyperspectral images based on a parallel architecture. The model is constructed based on different representations of the hyperspectral data along the spatial and the spectral dimensions. On the spatial domain we use Local Binary patterns-based histograms whereas along the spectral dimension of the hyperspectral dataset Wavelet filters are employed. These two descriptors are concatenated in a unique feature vector and a Support Vector Machine classifier is further used for classification. We show in the experimental section that the proposed method improves the accuracy on standard datasets (Indian Pines and Salinas) when compared to similar designs that consider Fourier transform-based descriptors.

Keywords: *SVM, hyperspectral, local binary patterns, wavelet, spatial-spectral information*

I. INTRODUCTION

Hyperspectral imaging systems can collect data along different regions of the electromagnetic spectrum, acquiring hundreds of narrow spectral bands. This will define the near-continuous electromagnetic spectrum information stored in each pixel which can provide significant features. The spectral information represented by each individual pixel is dependent on the chemical and physical properties of the respective material. These significant features enable improved performance for many applications. The concept of high-resolution hyperspectral images (HSI) if found in a large variety of processing systems, such as astronomy, agriculture, imaging, or geoscience [1],[2],[3].

Good classification results for the hyperspectral data sets always depend on the good quality of the hyperspectral set, on the type of feature extraction method employed and on the type of classifier. An accurate classification system must insure the extraction of discriminative features from the HSI dataset and preservation of low spatial resolution details. The inherent tradeoff between spectral and spatial resolution can be solved by designing classification systems taking into account both the spatial and spectral information [4].

Hyperspectral images are known to contain rich spectral information. Based on this fine representation for different materials, classification methods have been successfully applied to different types of high-resolution images, such as multispectral data and hyperspectral data. The major cause for a poor performance in classification tasks is represented by the size of training data set correlated with the large dimensionality of hyperspectral data. One of the simplest, but effective way of dealing with hyperspectral data is to reduce its dimensionality. This can be realized by applying different feature extraction methods to the hyperspectral data, when confronted with a limited set of training samples [5].

Several studies showed the immense impact in classification due to the correlation between the spatial information and the spectral dimension [5].

In this paper, we propose a novel approach that uses local binary patterns (LBPs) for extracting texture-like features in the spatial domain and wavelet-based multi-resolution and multi-multiscale analysis for representing the spectral signature associated to the HSI pixels. The obtained feature vectors are concatenated and used in a Support Vector Machine (SVM)-based supervised learning framework for the classification of HSI data. The method is described in section 2 and is evaluated experimentally in section 3. The final section of the paper includes conclusive remarks and also indicates some future research directions.

II. PROPOSED METHOD

A. *Local Binary Pattern for the spatial dimension*

The basic idea behind the LBP method is to capture texture information by quantifying local structures in a image by comparing each pixel with its neighborhood. The method takes a pixel as the center, t_c and for that corresponding pixel uses a thresholding operation for comparing its luminance value with those of its neighbors. If the intensity of the center pixel is greater or equal to its neighbor, then the neighbor takes the "1" value, respectively the "0" value. This is obtained using the function:

$$S(t_i - t_c) = \begin{cases} 1 \; if \; t_i - t_c \geq 0 \\ 0 \; if \; t_i - t_c < 0 \end{cases} \quad (1)$$

The LBP code is obtained using a predefined number of neighbors (P), equally distributed on a circle of radius R [6], as illustrated below:

$$LBP_{P,R}(t_c) = \sum_{i=0}^{P-1} S(t_i - t_c) 2^i \, . \quad (2)$$

In (2) t_c represents the gray value of the center pixel having

Manuscript received February 15 2021; revised April 28 2021

the coordinates (x_c, y_c). This value is compared with the intensities of the neighboring pixels (t_i). As illustrated in [6], several versions of the LPB operator do exist and, based on the analysis carried out on the same paper, we use the $LBP^{ri}_{P,R}$ rotation invariant operator for quantifying the spatial content on each spectral band.

As in [7] we further use square image patches to aggregate and compute histograms associated with these LBP codes.

B. *Wavelet Transform for the Spectral Dimension*

The wavelet transform of a 1D hyperspectral pixel along the spectral dimension can provide a multiscale analysis of its content allowing detection of the locations and scales at which significant properties of the spectral signature do exist.

In a classic formulation, the wavelet transform is designed to analyze at multiple scales the information in a signal. Processes called scaling and translation are used in order to obtain frequency content quantification and time localization. The scaling and translation of a function $\psi(x)$ is denoted as $\psi_{s,\tau}(x)$, with s representing the scaling coefficient and τ the translation coefficient:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-\tau}{s}) \qquad (3)$$

In (3), $\psi(t)$ is called wavelet or mother function. Essentially the wavelet transform decomposes a signal *x(t)* on the set of basis functions $\psi_{s,\tau}(x)$, producing a time-scale representation of *x(t)* [8]:

$$W_x(s,\tau) = \int_{-\infty}^{\infty} \psi_{s,\tau}(t)x(t)dt \qquad (4)$$

In order to approximate in a discrete setup the Continuous Wavelet Transform, we use the classical Discrete Wavelet Transform (DWT), implemented through iterated filter banks allowing the decomposition of the original signal through operations of filtering and decimation [9]. The DWT analysis is carried out on the spectral signature associated to a hyperspectral pixel and produces for each level of decomposition a representation of the spectral signature spectral content in two low-pass (L) and high –pass components (H). The procedure is followed by decimation and is iterated on the low-pass component, producing new low-pass and high-pass sub-bands [9]. Such a procedure is illustrated in Figure 1 for a spectral signature takes from the Indian Pines HSI dataset.

The properties of the resulting decomposition depend strongly on the type of the mother wavelet function. For investigating if such a choice has a major influence on the performance of the classification chain shown in Figure 2 we used two different classes of wavelet functions: Haar and Daubechies. Daubechies wavelets are orthonormal in nature. In the rest of the paper we employ the notation *dbN* for Daubechies wavelets, whith *N* depicting the order of the filter and *db* representing the name of the wavelet.

C. *Parallel design using LBP spatial features and wavelet-based coefficients spectral features*

The HSI classification system we are proposing is depicted in Figure 2. It is based on the concatenation of the two feature vectors: the feature vector obtained after the wavelet transform is applied on the spectral signatures,

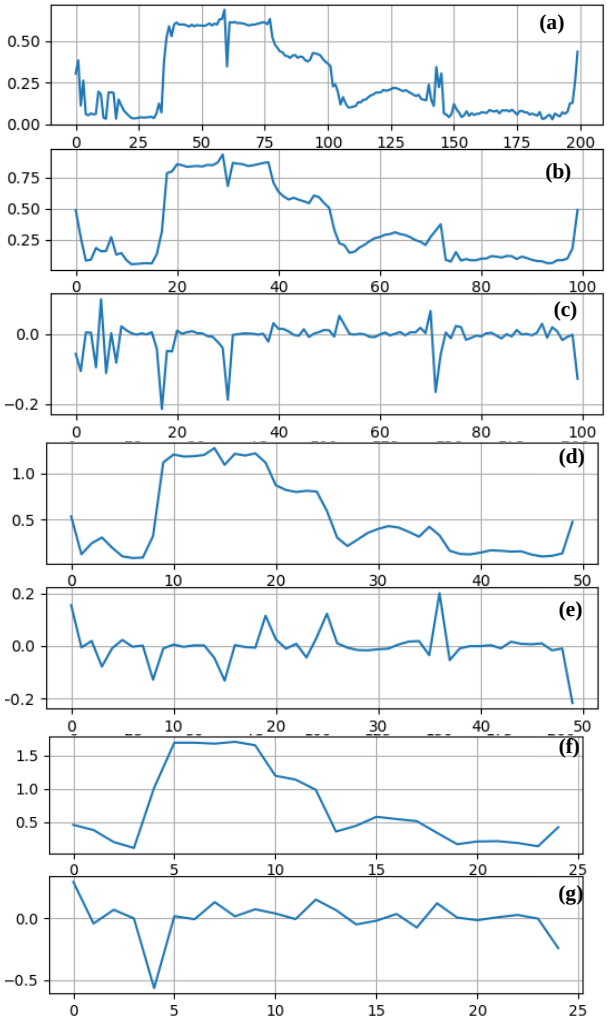


Figure 1. An example of wavelet coefficients based on the Haar wavelet for a hyperspectral pixel in Indian Pines: (a) hyperspectral signature, (b) and (c): the low (L) and high coefficients(H) for the first level decomposition (N=1), (d) and (e): the LL and LH coefficients for N=2, (f) and (g): the LLL and LLH coefficients for N=3

and, respectively a feature vector representing LBP histograms.

The resulting feature vector encodes both spectral features and intensity, texture-like information in selected bands obtained using principal component analysis (PCA) for selecting the most informative bands capable to offer high classification rates [10].

The proposed model is based on a similar parallel architecture as presented in [10]. With respect to this approach we encode the spectral information using the wavelet transform on a different number of levels of decomposition. The influence of the choice of the mother wavelet, of the coefficients associated to a particular sub band are investigated in the experimental section.

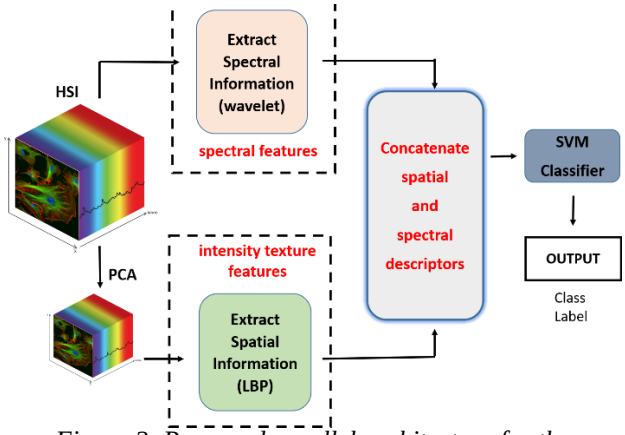


Figure 2. Proposed parallel architecture for the classification of hyperspectral data based on LBP and wavelet feature extraction and SVM classifiers

III. EXPERIMENTAL RESULTS

A. *Datasets*

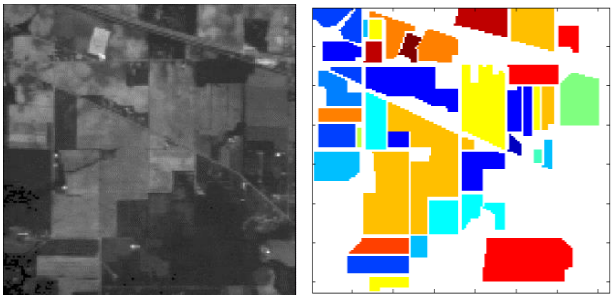
We performed all the experiments on the public data sets Indian Pines and Salinas [11]. The class distribution of Indian Pines and Salinas, both sets having 16 classes are represented in Figures 3 and 4. By analyzing Table I especially we observe that the Indian Pines class has some classes with a very small number of samples in comparison with other classes. This distribution of data samples can create problems in the classification portion of the framework due to the unbalanced number of labels for each class.

B. *Parameters*

In the case of unbalanced classes for a data base, it is desirable to split the data into train and test sets in such a way that we preserve the same proportions of samples in each class. This type of separation is defined as a stratified train-test split and can be achieved be setting a "stratify" argument for the label set that represents the data. This arguments is used by a function called *train_test_split()* which ensure that both the train and test sets will have the same sample distribution in each class that based on the distribution of the labels. We performed each experiment 10 times and report both the overall accuracy (OA) and the standard deviation.

In order to assess the performance of the proposed approach in similar conditions, we conceived the experimental plan in similar manner to the approaches taken in [12]: the number of the PCA selected bands fixed to 7, the patch size fixed to 17×17 for Indian Pines and, respectively, fixed to 10 bands and a patch size of 25 x 25 pixels for Salinas dataset.

As known by classical results, the classification accuracy strongly depends on the choice of the training and test data are chosen. Moreover, for patch-based representation, due potential overlapping of the patches, independence between the training set and the testing set should be considered, as illustrated in [13]. In order to evaluate the classification performance of the proposed model, we chose the support vector machine (SVM) as the classifier. Depending on the type of data we want to classify, we performed a grid search in order to determine

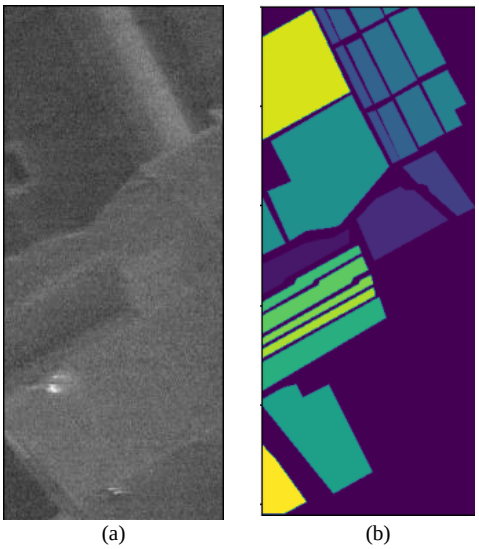


(a) (b)

Figure 3. Image and ground truth classes for the Indian Pines dataset (a) Sample band; (b) Ground truth classes represented using different colors

TABLE I. CLASSES FOR THE INDIAN PINES DATASET AND ASSOCIATED SAMPLE NUMBER

	Class	Samples
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-mintill	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pasture-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-mintill	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings-Grass-Trees-Drives	386
16	Stone-Steel-Towers	93



(a) (b)

Figure 4. Image and ground truth classes for the Salinas dataset (a) Sample band; (b) Ground truth classes using different colors

TABLE II. CLASSES FOR THE INDIAN PINES DATASET AND ASSOCIATED SAMPLE NUMBER

#	Class	Samples
1	Brocoli_green_weeds_1	2009
2	Brocoli_green_weeds_2	3726
3	Fallow	1976
4	Fallow_rough_plow	1394
5	Fallow_smooth	2678
6	Stubble	3959
7	Celery	3579
8	Grapes_untrained	11271
9	Soil_vinyard_develop	6203
10	Corn_senesced_green_weeds	3278
11	Lettuce_romaine_4wk	1068
12	Lettuce_romaine_5wk	1927
13	Lettuce_romaine_6wk	916
14	Lettuce_romaine_7wk	1070
15	Vinyard_untrained	7268
16	Vinyard_vertical_trellis	1807

the optimal parameters for the radial basis function kernel of the SVM i.e *gamma* and *c*.

For the spatial feature extraction LBP method we studied three different pairs for the $LBP_{P,R}$: (8, 1), (16, 2) and (24, 3) as illustrated in [10].

For all the tests performed in this paper we use $LBP_{16,2}$ to obtain the intensity texture features. For the wavelet filters, an important parameter represents the level of decomposition, known as *N*. The selected values for the decomposition level are correlated with the complexity of the hyperspectral vector. Therefore, considering different decomposition levels, the wavelet filter can extract the desired features from the input for a better representation for the signal. In this paper we have tested the performance of our model for three decomposition levels, *N=1, N=2* and *N=3*.

C. *Classification performance*

We investigated both the role of the wavelet type and of the number of the associated coefficients to represent the spectral signature.

Table III resumes the results for the two considered data bases with different ratios defining the training samples per class and the testing data sets, using only the spectral signature. The ratios were set as 10%, respectively 30% for the data train.

In Table IV we show the mean accuracy and associated standard deviation for the Indian Pines data set based on the two wavelet filters when using the proposed approach

TABLE III. SPECTRAL CLASSIFICATION ACCURACY (%) , USING A RATIO OF 10% AND 30% OF SAMPLES PER CLASS FOR TRAINING FOR INDIAN PINES AND SALINAS SCENES

	10% ratio/class	30% ratio/class
Indian Pines	74.03 ± 0.39	79.34 ± 0.58
Salinas	79.34 ± 0.18	89.34 ± 0.14

the classification of the spectral dimension, after the wavelet transforms were applied. using Haar and Daubechies filters, and a variable number of decomposition levels. The overall accuracies represent the classification of the spectral dimension, after the wavelet transforms were applied.

Based on the concept of spectral similarity between pixels as presented in [10], we considered mainly the low coefficients, labeled in this paper as L, LL or LLL, depending on the level of decomposition for the wavelet. Beside the low frequencies, we take into consideration the influence of the high frequencies for the same hyperspectral vector. In this paper we consider the H, LH and LLH bands to observe a different percentage of the high frequencies can influence the classification results.

If we take the first level of decomposition (N=1) for the hyperspectral vector, the overall accuracy, when considering only the high frequencies, has a smaller value, of ~ 89.08%, in comparison with the classification of the low frequency spectrum when we obtain a value of ~92.14%. By choosing the low band and the high band together we have a decrease of ~1% in performance when we have a larger training data set. On the contrary, if we have a smaller representation of samples per class, the 10% ratio case, we have a decrease in overall accuracy. Based on this observation we will no longer consider only the high frequency band, H, for the rest of the decomposition levels. As a comparison, between the two wavelet filters, the Haar and Daubechies, the performances are similar in terms of overall accuracy.

In Table V we show the performance results obtained for the Salinas data set, taking into consideration all 16 classes. Our proposed parallel architecture has good generalization capacities, when we take into consideration only the low frequencies. In terms of overall classification values, the Figure 5 and Figure 6 represent the classification maps for Indian Pines and Salinas. In order to illustrate the time consumption of the proposed model, we chose the same experimental results:

TABLE IV. CLASSIFICATION ACCURACY (%) FOR SVM (KERNEL:RBF, GAMMA: 0.01, C:100), USING 10% AND 90% RATIO FOR THE TRAINING DATA SET PER CLASS FOR INDIAN PINES WITH DIFFERENT DECOMPOSITION LEVELS (N) FOR HAAR AND DAUBECHIES FILTERS

		Haar		db20	
		10% samples/class	30% samples/class	10% samples/class	30% samples/class
N=1	**L**	92.14 ± 0.17	92.87 ± 0.24	92.30 ± 0.14	92.02 ± 0.45
	H	89.08 ± 0.32	92.88 ± 0.26	87.41 ± 0.10	91.85 ± 0.22
	L+H	90.76 ± 0.37	**93.92 ± 0.20**	90.57 ± 0.35	**93.21 ± 0.51**
N=2	**LL**	92.03 ± 0.22	**94.94 ± 0.18**	92.83 ± 0.12	**94.25 ± 0.63**
	LH	90.40 ± 0.26	93.66 ± 0.11	90.55 ± 0.32	92.99 ± 0.22
	LL+LH	91.63 ± 0.31	94.53 ± 0.08	90.63 ± 0.17	93.84 ± 0.35
N=3	**LLL**	91.83 ± 0.26	**94.83 ± 0.16**	91.13 ± 0.46	**94.14 ± 0.26**
	LLH	91.07 ± 0.40	93.99 ± 0.28	90.15 ± 0.32	93.37 ± 0.48
	LLL+LLH	91.82 ± 0.19	94.58 ± 0.22	91.43 ± 0.27	93.51 ± 0.25

TABLE V. CLASSIFICATION ACCURACY (%) FOR SVM (KERNEL:RBF, GAMMA: 0.01, C:100), USING 10% AND 90% RATIO FOR THE TRAINING DATA SET PER CLASS FOR SALINAS WITH DIFFERENT DECOMPOSITION LEVELS (N) FOR HAAR AND DAUBECHIES FILTERS

		Haar		db20	
		10% samples/class	30% samples/class	10% samples/class	30% samples/class
N=1	**L**	95.49 ± 0.03	**96.67 ± 0.03**	95.13 ± 0.01	**96.56 ± 0.02**
	H	95.22 ± 0.02	95.78 ± 0.07	94.02 ± 0.06	95.82 ± 0.09
	L+H	95.43 ± 0.08	96.57 ± 0.06	95.23 ± 0.04	95.96 ± 0.07
N=2	**LL**	95.43 ± 0.04	96.50 ± 0.02	95.55 ± 0.05	96.84 ± 0.06
	LH	95.47 ± 0.03	96.51 ± 0.04	95.62 ± 0.03	96.77 ± 0.03
	LL+LH	95.58 ± 0.03	**96.61 ± 0.08**	95.97 ± 0.05	**96.86 ± 0.04**
N=3	**LLL**	95.31 ± 0.05	96.41 ± 0.04	95.28 ± 0.02	96.54 ± 0.02
	LLH	95.02 ± 0.07	95.98 ± 0.03	95.66 ± 0.05	96.13 ± 0.05
	LLL+LLH	96.59 ± 0.04	**96.79 ± 0.06**	95.44 ± 0.02	**96.58 ± 0.04**

TABLE VI. TRAINING AND TESTING TIMES FOR INDIAN PINES DATASET WITH 30% RATIO FOR TRAINING SET - DATA CONCATENATION WITH HAAR WAVELET AND $LBP_{16,2}$

		training time(s)	**testing time(s)**	**OA(%)**
$LBP^{riu}_{16,2}$+$N_{fft:512}$		29	16	91.94 ± 0.78
N=2	**LL**	9	7	94.94 ± 0.18
	LL+LH	11	8	94.53 ± 0.08
N=3	**LLL**	7	6	94.83 ± 0.16
	LLL + LLH	8	6	94.58 ± 0.22

TABLE VII. TRAINING AND TESTING TIMES FOR FOR SALINAS WITH 30% RATIO FOR TRAINING SET - DATA CONCATENATION WITH HAAR WAVELET AND $LBP_{16,2}$

		training time(s)	**testing time(s)**	**OA(%)**
$LBP^{riu}_{16,2}$+$N_{fft:512}$		229	315	94.56 ± 0.07
N=2	**LL**	108	192	96.62 ± 0.03
	LL+LH	120	210	96.61 ± 0.08
N=3	**LLL**	67	136	96.41 ± 0.04
	LLL+LLH	76	150	96.59 ± 0.06

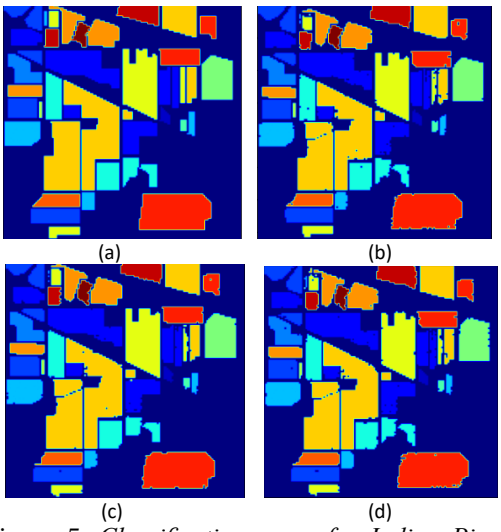


Figure 5. Classification maps for Indian Pines with 30% sample ratio/class for the training set per class for training and different decomposition levels: (a) Ground truth, (b) N=1: $LBP^{riu}_{16,2}$+$Haar_L$,(c) N=2: $LBP^{riu}_{16,2}$+ $Haar_{LL}$, (d) N=3: $LBP^{riu}_{16,2}$+ $Haar_{LLL}$

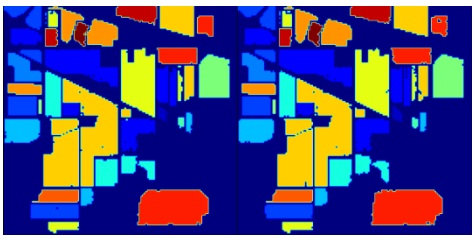


Figure 5 (cont.) Classification maps for Indian Pines with 30% sample ratio/class for the training set per class for training and different decomposition levels: (e) N=2: $LBP^{riu}_{16,2}$+ $Haar_{LL+LH}$, (f) N=3: $LBP^{riu}_{16,2}$+ $Haar_{LLL+LLH}$

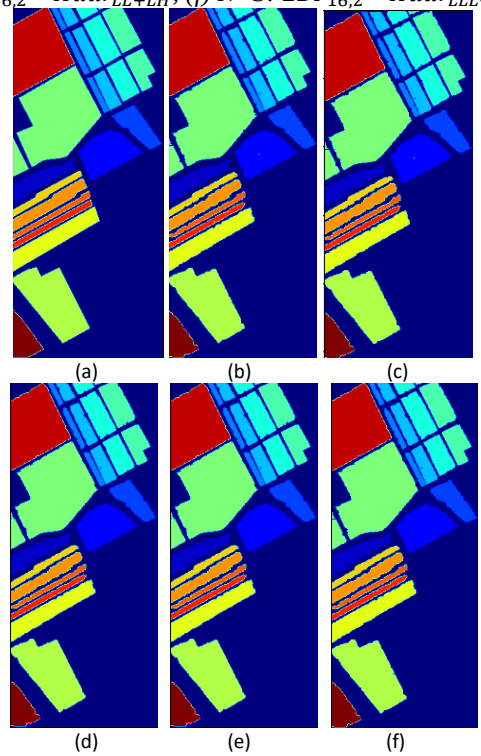


Figure 6. Classification maps obtained for Salinas with 30% sample ratio/class for the training set per class different decomposition levels:(a) Ground truth, (b) N=1: $LBP^{riu}_{16,2}$+$Haar_L$,(c) N=2: $LBP^{riu}_{16,2}$+ $Haar_{LL}$, (d) N=3: $LBP^{riu}_{16,2}$+ $Haar_{LLL}$,(e) N=2: $LBP^{riu}_{16,2}$+ $Haar_{LL+LH}$, (f) N=3: $LBP^{riu}_{16,2}$+ $Haar_{LLL+LLH}$,

for Indian Pines an average of 18 samples per class, so all the classes can be represented, respectively 200 samples per for Salinas generated using the two ratios for separating the training and testing data. The training and testing times were evaluated for the parallel architecture proposed in this paper. Also we have tested the computation time for the case in which on the spectral dimension we applied a DFT transform to obtain the frequency spectrum representation. By analyzing Table V and Table VI we show how efficient the parallel architecture is, being capable to label the hyperspectral data, for Indian Pines for example from 9 second up to 11 seconds, depending on the dimension of the concatenated vector. Comparing the wavelet coefficients with the DFT coefficients based on the same table, we observe a different of ~10s for training and testing times. Another important aspect is represented by the ~3% in accuracy brought by the wavelet coefficients in comparison with result obtained using the $LBP_{16,2}^{riu}$+$N_{\text{fft:512}}$ model, from ~91% to ~94%. The same behavior is present for Salinas.

IV. CONCLUSIONS

A novel approach for the classification of hyperspectral data in proposed. Our approach takes into consideration the dependence between the spatial and spectral dimensions. The feature vector to be classified is represented by the concatenation between patch-based LBP histograms for the spatial domain and wavelet coefficients for the spectral one. The proposed model has good capabilities of extracting the features from the hyperspectral image and provides better accuracy figures in classification that similar approaches that employs Fourier coefficients for the spectral domain. The testing and training time are also smaller compared to the Fourier case.

Despite the increase in performance, this parallel design can still be improved by considering different types of methods applied on the spatial and spectral dimensions.

Future work will be devoted to the design of train/test separation that take into account pre imposed distances in order to insure true independence between the train and test data samples in such a patch-based method.

REFERENCES

[1] R. N. Handcock, B. Evans, N. Liu, R. J. Harper. “Seasonal timing for estimating carbon mitigation in revegetation of abandoned agricultural land with high spatial resolution remote sensing”, Remote Sensing, 9(6), 2017, pp. 545.

[2] Q. Li, L. Zhuang, B. Zhang, J. M. Bioucas-Dias et al., “A new low-rank representation based hyperspectral image denoising method for mineral mapping”, Remote Sensing, 9(11), 2017, pp. 1145.

[3] W. Wang, C. Lin, X. Zhang, J. Cui, “A study for texture feature extraction of high-resolution satellite images based on a direction measure and gray level co-occurrence matrix fusion algorithm”, Remote Sensing, 17(7), 2017, pp. 1474.

[4] H. Zhang, L. Zhang, H. Shen, Q. Yuan, C. Jiang, ”Hyperspectral image denoising with a combined spatial and spectral weighted hyperspectral total variation model”, Canadian Journal of Remote Sensing, 42, 2016.

[5] A. Plaza, J. Plaza, G. Martín, “Incorporation of spatial constraints into spectral mixture analysis of remotely sensed hyperspectral data”, IEEE International Workshop on Machine Learning for Signal Processing. MLSP, 2009, doi: 10.1109/MLSP.2009.5306202.

[6] T. Ojala, M. Pietikäinen, T. Maenpaa, “Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24, 2002, pp. 971-987.

[7] Li, W. Li, C. Chen, H. Su et al., “Local Binary Patterns and Extreme Learning Machine for Hyperspectral Imagery Classification”, IEEE Transactions on Geoscience and Remote Sensing. **53**, 2015, pp. 1-13.

[8] S. Feng, P. Yuki, M Duarte, “Wavelet-Based Semantic Features for Hyperspectral Signature Discrimination”, 2016, arXiv:1602.03903

[9] S. Dev Vishnu, S. Rajan, V. Sowmya and K. P. Soman, "Hyperspectral image denoising: A least square approach using Computing”, Communications and Informatics (ICACCI), Udupi, 2017, pp. 805-811, doi: 10.1109/ICACCI.2017.8125941.

[10] A. V. Miclea, R. Terebes and S. Meza, "Local binary patterns and Fourier transform based hyperspectral image classification," 2020 International Symposium on Electronics and Telecommunications (ISETC), Timisoara, Romania, 2020, pp. 1-4, doi: 10.1109/ISETC50328.2020.9301080.

[11] [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title= Hyperspectral _Remote_Sensing_Scenes [15.02.2021].

[12] S. Jia, J. Hu, “Three-Dimensional Local Binary Patterns for Hyperspectral Imagery Classification”, IEEE Transactions on Geoscience and Remote Sensing. 2017, pp. 1-15.

[13] S. Barburiceanu, R. Terebes and S. Meza, "Hyperspectral Image Classification using the MRELBP Texture Descriptor," 2019 E-Health and Bioengineering Conference (EHB), Iasi, Romania, 2019, pp. 1-6, doi: 10.1109/EHB47216.2019.8969874.