Three-Dimensional Wavelet Texture Feature Extraction and Classification for Multi/Hyperspectral Imagery

Xian Guo, Xin Huang, *Member, IEEE*, and Liangpei Zhang, *Senior Member, IEEE*

***Abstract*—A 3-D wavelet-transform-based texture feature extraction algorithm for the classification of urban multi/hyperspectral imagery is investigated in this study. It is widely agreed that it is necessary to simultaneously exploit the spectral and spatial information for image classification. In this context, the 3-D discrete wavelet transform (3-D DWT) is studied since it considers the local imagery patch as a cube and, hence, is capable of representing the imagery information in both spectral and spatial domains. The notable characteristic of the 3-D DWT is the ability to decompose an image into a set of spectral–spatial components. Specifically, we propose three approaches for 3-D DWT texture extraction, namely, pixelwise, non-overlapping, and overlapping cube. Experiments conducted on AVIRIS hyperspectral and WorldView-2 multispectral images revealed that the 3-D DWT textures achieved much better results than the widely used spectral–spatial classification methods.**

***Index Terms*—Classification, feature extraction, high resolution, hyperspectral, texture, wavelet.**

I. Introduction

MULTI/HYPERSPECTRAL imagery has the potential to precisely discriminate different land-cover types using the relatively abundant spectral information. However, with the rapidly increasing spatial resolution of the available remote sensing images, it has been argued that the spectral information alone is not adequate for classification [1]. Consequently, various spatial feature extraction methods have been proposed for providing discriminative information in pattern classification recently, such as pixel shape index [2], object-based procedure [3], Markov random field [4], extended morphological profiles (EMPs) [5], and attribute profiles (APs) [6]. However, the aforementioned conventional 2-D spatial feature extraction methods are mainly based on a single spectral band or the first few principal components of the multispectral bands, which are not efficient for spectral–spatial representation of the hyperspectral remotely sensed imagery.

Three-dimensional feature extraction is capable of simultaneously extracting image information from the spectral and spatial directions, which provides more effective features for interpretation of multi/hyperspectral imagery. Bau *et al.* [7] developed a series of spatial–spectral features based on the 3-D Gabor filters to integrate the orientation, scale, and wavelength-dependent properties of hyperspectral imagery. Tsai and Lai [8] proposed a 3-D gray-level cooccurrence matrix (GLCM), extending the conventional GLCM to its 3-D version. In recent years, 3-D discrete wavelet transform (3-D DWT) has been employed for data mining of remotely sensed data. Qian *et al.* [9] combined the 3-D DWT texture extraction with the sparse logistic regression for hyperspectral image classification.

In this context, we propose a new 3-D wavelet feature extraction and classification framework for multi/hyperspectral imagery. Specifically, three feature extraction algorithms are proposed in order to simultaneously exploit the spectral and spatial information via per-pixel, non-overlapping, and overlapping per-cube sliding. Afterward, the energy of the subband wavelet coefficients is employed to characterize the texture.

The remainder of this letter is organized as follows. A brief review of 3-D DWT is introduced in Section II. The proposed 3-D spectral–spatial feature extraction approaches are described in Section III, followed by the experimental data sets and their results in Section IV. Section V concludes this letter.

II. Three-Dimensional Discrete Wavelet Decomposition

DWT is derived from the multiresolution analysis theory using wavelet filter banks. The low-pass (L) and high-pass (H) filtering branches of the filter banks retrieve the approximations and details of the input signal [10], respectively. In the 1-D DWT, the lower branch consists of low-pass filters and an upsampling procedure, along with the impulse response converging to the low-pass sequences, i.e., $l(n)$

$$\phi(t) = \sum_{n=0}^{N} l(n)\phi(2t-n) \tag{1}$$

where $\phi(t)$ serves as the scaling function, and t and n are the scaling and translation parameters, respectively. For the high-pass sequences $h(n)$, the wavelet function $\psi(t)$ can be similarly defined as

$$\psi(t) = \sum_{n=0}^{N} h(n)\psi(2t-n). \tag{2}$$

For the multi/hyperspectral imagery, 3-D DWT is carried out by extending the conventional 1-D wavelet to the spectral and

Manuscript received February 21, 2014; revised May 9, 2014; accepted May 9, 2014. This work was supported in part by the National Natural Science Foundation of China under Grant 41101336 and 91338111, by the Program for New Century Excellent Talents in University of China under Grant NCET-11-0396, and by the Foundation for the Author of National Excellent Doctoral Dissertation of China under Grant 201348.

The authors are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: huang_whu@163.com).

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/LGRS.2014.2323963

spatial domains of an image, i.e., applying the 1-D DWT filter banks in turn to each of the three dimensions (two for the spatial domain and one for the spectral direction). The 3-D DWT is constructed by a tensor product [11]

$$\begin{aligned} I^{(x,y,z)} &= (\mathrm{L}^x \oplus \mathrm{H}^x) \otimes (\mathrm{L}^y \oplus \mathrm{H}^y) \otimes (\mathrm{L}^z \oplus \mathrm{H}^z) \\ &= \mathrm{L}^x\mathrm{L}^y\mathrm{L}^z \oplus \mathrm{L}^x\mathrm{L}^y\mathrm{H}^z \oplus \mathrm{L}^x\mathrm{H}^y\mathrm{L}^z \oplus \mathrm{L}^x\mathrm{H}^y\mathrm{H}^z \\ &\quad \oplus \mathrm{H}^x\mathrm{L}^y\mathrm{L}^z \oplus \mathrm{H}^x\mathrm{L}^y\mathrm{H}^z \oplus \mathrm{H}^x\mathrm{H}^y\mathrm{L}^z \oplus \mathrm{H}^x\mathrm{H}^y\mathrm{H}^z \end{aligned} \quad (3)$$

where $\oplus$ and $\otimes$ denote the space direct sum and tensor product, respectively. L and H represent the low- and high-pass filters along the x, y, and z axes, respectively. In practice, the x and y directions denote the spatial coordinates of an image, and z is the spectral axis. After a single-level processing, the volume data are decomposed into eight subbands, i.e., LLL, LLH, LHL, LHH, HLL, HLH, HHL, and HHH, which can be separated into three categories: subbands of approximation (LLL), spectral variation (LLH, LHH, HLH), and spatial variation (LHL, HLL, HHL), respectively.

III. Three-Dimensional DWT Texture Feature Extraction

Three-dimensional wavelet analysis has been successfully applied to video compression [11] and medical imaging [12]. In this letter, we propose a novel 3-D spectral–spatial feature extraction and classification framework for multi/hyperspectral remotely sensed imagery by defining a series of local cubes around the central pixel. The 3-D DWT is implemented in each local cube along the spectral and spatial directions, resulting in eight subbands after one-level decomposition. The generated 3-D DWT coefficients represent the amount of spectral and spatial variation and, therefore, can help to distinguish different classes. The 3-D DWT textures are calculated by measuring the energy functions of the 3-D wavelet coefficients with three levels: pixelwise, non-overlapping, and overlapping cube, respectively.

A. Pixel-Based Shift

The pixel-based strategy for the 3-D DWT textural extraction is based on a local cube around each pixel. In this pixel-based DWT (PDWT) algorithm, the energy statistic, i.e., the quadratic sum of the wavelet coefficients for each subband obtained by wavelet decomposition, is used to characterize the texture property

$$E(W) = \sum_{i=1}^{B/2^L} \sum_{j=1}^{B/2^L} \sum_{k=1}^{N/2^L} P(i,j,k)^2 \quad (4)$$

where W is a $B \times B \times N$ local cube with B and N as the dimensions in the spatial and spectral domains, respectively. $P(i,j,k)$ is the wavelet coefficient in the cube centered by the pixel (i,j,k).

It should be pointed out that the pixel-based cube shift is subject to a large computational burden since each pixel in the image is used as the center of a moving cube. Moreover, high correlation exists between features extracted from the highly overlapped areas, resulting in a large amount of redundant information. Therefore, we propose non-overlapping and overlapping approaches for the 3-D texture representation.

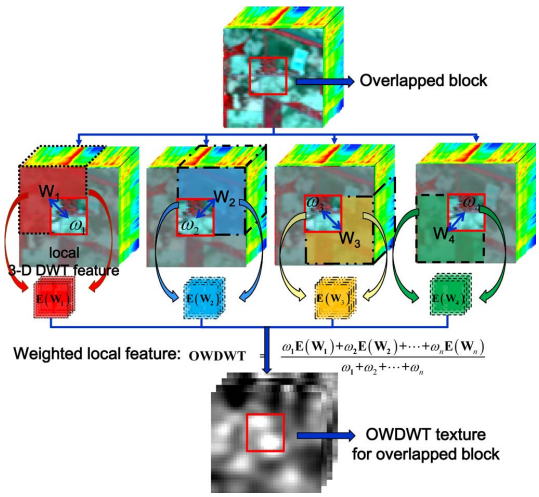

Fig. 1. Overlapping cube strategy for the 3-D DWT feature extraction.

B. Non-Overlapping Cube

Given a window size B, we can equally divide an input image into non-overlapping blocks with $B \times B$ pixels for each one. Different from the pixelwise approach, in the non-overlapping mode, 3-D DWT is performed in each block, and the feature value of the cube is assigned to all the pixels in the block. The computational cost of this non-overlapping sliding window DWT (WDWT) is much lower than the pixel-based approach. In this way, however, the edge and details cannot be well preserved in the classification.

C. Overlapping Cube

As stated earlier, the non-overlapping approach does not take the spatial relationship of adjacent cubes into consideration, which leads to spatial information loss and edge-blurring effects. Therefore, the overlapping DWT (OWDWT) strategy is proposed as a tradeoff between computational cost (pixelwise mode) and information loss (non-overlapping mode). To deal with the pixels in the overlapping area, a weighted feature representation method is used. As demonstrated in Fig. 1, there are four adjacent local cubes, i.e., W_1, W_2, W_3, and W_4 (red, blue, yellow, and green dashed blocks), overlapping at the central block (red cell). Since the overlapping area ($S \times S \times N$ pixels, $S < B$) is a common region of the four cubes (i.e., $B \times B \times N$ cube), 3-D DWT feature extraction is first computed at each cube. To incorporate the spectral–spatial information contained in these overlapping cubes, the overlapping 3-D texture feature is subsequently computed as a weighted linear combination of the neighboring texture measurement values, i.e.,

$$\mathrm{OWDWT} = \frac{\omega_1 E(W_1) + \omega_2 E(W_2) + \cdots + \omega_n E(W_n)}{\omega_1 + \omega_2 + \cdots + \omega_n} \quad (5)$$

where ω_N and $E(W_N)$ denote the weight and texture measurement of the cube W_N, respectively. The weight is defined as the Euclidean distance between a specific cube and the overlapping area. The optimal value of S is related to spatial characteristics

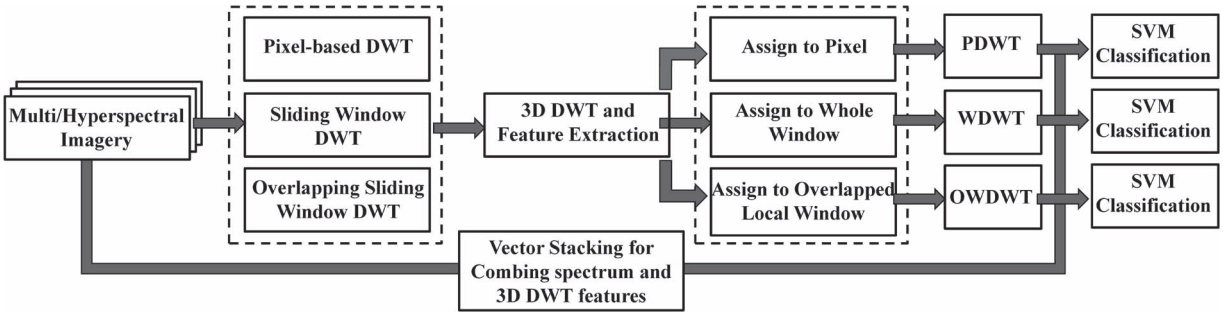

Fig. 2. Proposed 3-D DWT feature extraction and classification framework for multi/hyperspectral imagery.

of classes in different image scenes and, hence, is manually tuned in this study.

The advantages of the overlapping cube strategy include the following: 1) It is able to achieve comparable accuracy but with much less computational cost than the pixelwise approach. 2) The spatial relationship of adjacent pixels, as well as the spectral information, is exploited more efficiently due to the overlapping processing.

Fig. 2 presents the overview of the proposed 3-D DWT feature extraction and classification framework.

IV. Experiments and Comparisons

The proposed algorithms are validated using the AVIRIS hyperspectral image and the WorldView-2 multispectral data set. Furthermore, a comparison is made between the proposed 3-D DWT algorithms and the state-of-the-art 2-D methods, e.g., GLCM [13], EMPs [5], and extended APs (EAPs) [6], together with the newly developed 3-D GLCM [8]. The spectral and textural features are concatenated and then fed into a support vector machine (with a radial basis function kernel) for classification. A threefold cross validation is used to select the kernel parameter γ and penalty coefficient C. The classification accuracy is assessed by the overall accuracy (OA) with an available range of training samples randomly chosen from the ground-truth reference. Each classification scenario is repeated ten times with different starting training sets, and the mean and standard deviation of the classification results are reported.

A. Data Sets

The first data set was collected by the AVIRIS sensor in 1992, referring to an agricultural area of Indian Pines in northern Indiana, USA. This hyperspectral image contains 16 classes with 145×145 pixels for 220 spectral bands. The second data set is a multispectral image captured by the WorldView-2 in 2011 in Wuhan, central China. This image refers to an urban landscape with 512×512 pixels of eight bands, where six information classes were manually labeled for model validation. The test images and ground-truth references for the two data sets are shown in Fig. 3.

B. AVIRIS Experiment

In the experiments, the spectral method denotes the classification using spectral information only. The overlapping area S is set to 4, and window size B varies from 8, 16, 32, and 64 for the OWDWT. B is fixed to 32 for PDWT and WDWT. The size

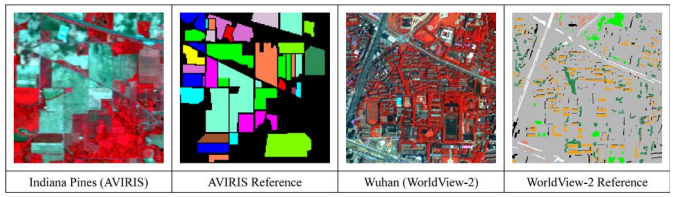

Fig. 3. AVIRIS and WorldView-2 data sets.

of the window of the GLCM is set to 32. The kernel size of the 3-D GLCM is $32 \times 32 \times 5$. The first four principal components are used as the base images of the EMPs and EAPs, which are generated using a diamond-shaped structural element with a series of sizes (4, 8, and 12). Afterward, 1%, 5%, and 10% of the labeled samples were randomly selected from each class of the reference as training samples while keeping the remaining samples for testing.

By analyzing the accuracy levels of the algorithms with different sizes of training samples, as shown in Table I, the proposed OWDWT produces the best classification accuracy when compared with other spectral–spatial classification methods, even with a small training sample size. In the case of 5% sampling, the OWDWT algorithm improves the OA from 68.41% (spectral classification) to 95.31% and significantly outperforms GLCM (85.21%), 3-D GLCM (76.16%), EMPs (88.95%), and EAPs (85.97%). Note that the pixelwise and non-overlapping approaches also give much better results than the spectral classification; however, their accuracy levels are slightly lower than the 2-D feature extraction algorithms. It can be stated that the proposed overlapping algorithm has the potential for providing efficient 3-D texture representation for image classification.

Fig. 4 shows the classification maps obtained by different features with 5% training samples. A general observation shows that the addition of spatial features can considerably improve the classification performance compared with the spectral classification. GLCM and 3-D GLCM provide neighboring information, which helps relieve the speckle noise for classification. Due to the invariance to local contrast change, EMPs also significantly improve the results. EAPs can further reduce the misclassifications by considering the attribute filters. As for the proposed 3-D DWT features, PDWT provides textural information for each pixel, which is sensitive to local change and leads to speckle noise. WDWT is less sensitive to local change; however, it leads to edge-blurring effects due to the window-based processing. OWDWT achieves a balance between texture representation and computational cost, and it yields the best classification performance.

TABLE I
CLASSIFICATION ACCURACY LEVELS OF VARIOUS FEATURES (AVIRIS DATA SET)

OA(%)	Spectrum	GLCM	3D GLCM	EMPs	EAPs	PDWT	WDWT	OWDWT
1%	54.64 ± 0.87	63.13 ± 2.15	58.97 ± 1.46	72.05 ± 2.94	71.26 ± 2.27	60.17 ± 1.66	74.25 ± 1.89	81.67 ± 3.00
5%	68.41 ± 0.79	85.21 ± 1.21	76.16 ± 0.88	88.95 ± 0.61	85.97 ± 1.03	79.55 ± 1.35	88.00 ± 0.80	95.31 ± 0.73
10%	76.50 ± 1.14	93.24 ± 0.28	83.11 ± 0.82	93.11 ± 0.55	91.51 ± 0.55	86.83 ± 0.55	90.70 ± 0.54	97.21 ± 0.47

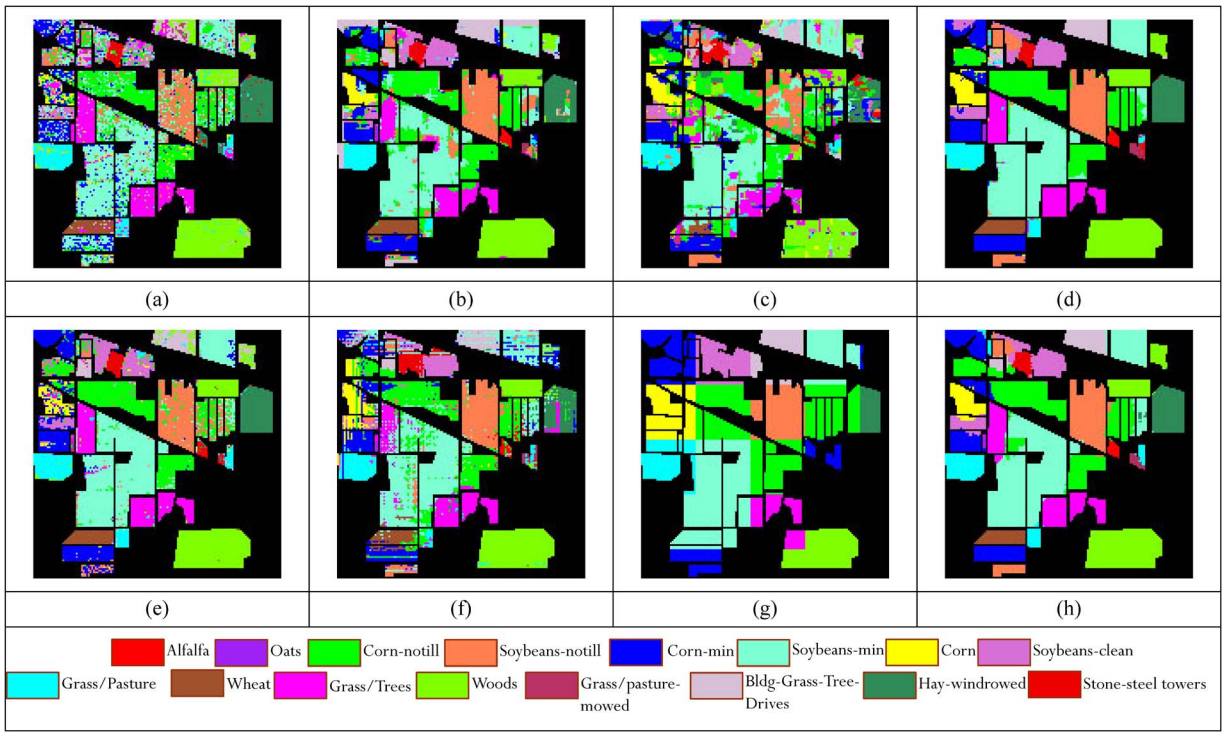

Fig. 4. Classification maps of the AVIRIS data (15 pixels for training per class). (a) Spectrum. (b) GLCM. (c) Three-dimensional GLCM. (d) EMPs. (e) EAPs. (f) PDWT. (g) WDWT. (h) OWDWT.

TABLE II
CLASSIFICATION ACCURACY LEVELS OF VARIOUS FEATURES FOR THE WORLDVIEW-2 DATA SET
("PTS" INDICATES THE NUMBER OF TRAINING SAMPLES IN PIXELS FOR EACH CLASS)

OA	Spectrum	GLCM	3D GLCM	EMPs	EAPs	PDWT	WDWT	OWDWT
10 pts	84.18 ± 2.82	84.58 ± 3.37	86.07 ± 5.49	86.69 ± 4.10	88.36 ± 2.30	87.70 ± 2.22	82.59 ± 2.20	90.05 ± 1.38
20 pts	88.57 ± 1.35	90.23 ± 0.99	90.63 ± 2.36	92.02 ± 1.43	92.23 ± 1.46	89.79 ± 0.98	88.55 ± 2.40	92.63 ± 0.91
30 pts	90.45 ± 0.66	90.70 ± 1.11	90.49 ± 1.09	93.87 ± 1.16	94.71 ± 1.02	90.09 ± 0.85	90.68 ± 0.86	95.25 ± 0.90

C. Experiment With the WorldView-2 Imagery

The challenge for classifying the WorldView-2 Wuhan image is to distinguish the spectrally similar classes such as soil–roads–buildings and trees–grass in a complicated urban environment. A total of 42 931 pixels are manually labeled to evaluate the effectiveness of various algorithms. In the experiments, 10, 20, and 30 pixels per class are randomly selected as training samples. For the OWDWT, the overlapping area S is set to 4, and the window size B varies from 6, 8, 10, to 12, and B is set to 7 for PDWT and WDWT. The statistical window for GLCM is chosen as 7, whereas the kernel size of 3-D GLCM is $5 \times 5 \times 5$. EMP and EAP features are generated using a diamond-shaped structural element with a series of sizes (3, 6, and 9) based on the first four principal components (see Table II).

The classification maps in Fig. 5 show that when using only spectral information, there are many misclassifications between buildings, roads, and soil due to their similar spectral properties. This phenomenon is alleviated by taking spatial features into account. We zoom in the central bottom part [highlighted in red rectangle in Fig. 6(a)] of the image in Fig. 6 to compare the classification performance of different methods with ten training samples per class. GLCM and 3-D GLCM slightly improve the OA obtained by the spectral classification (84.18%) to 84.58% and 86.07%, respectively. Note that, although EMPs and EAPs achieve high scores in OA, i.e., 86.69% and 88.36%, respectively, a large amount of misclassifications between trees–building and building–roads can be observed. Benefiting from the consideration of the spatial–spectral relationship with the overlapping region, the OWDWT method achieves the highest OA among all the classification algorithms

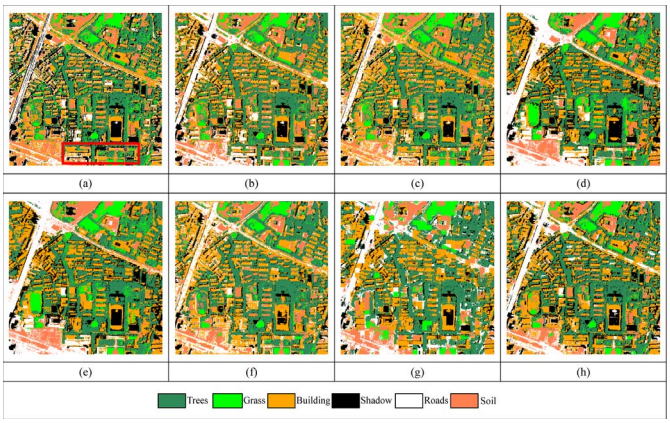

Fig. 5. Classification maps of the WorldView-2 image obtained by (a) spectral, (b) GLCM, (c) 3-D GLCM, (d) EMPs, (e) EAPs, (f) PDWT, (g) WDWT, and (h) OWDWT.

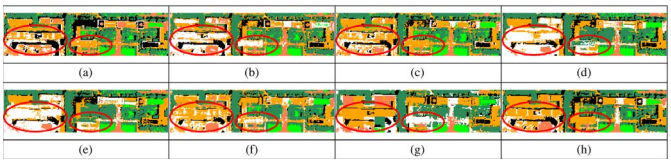

Fig. 6. Zoomed classification maps of the WorldView-2 image obtained by (a) spectral, (b) GLCM, (c) 3-D GLCM, (d) EMPs, (e) EAPs, (f) PDWT, (g) WDWT, and (h) OWDWT. Detailed results are underlined in red regions.

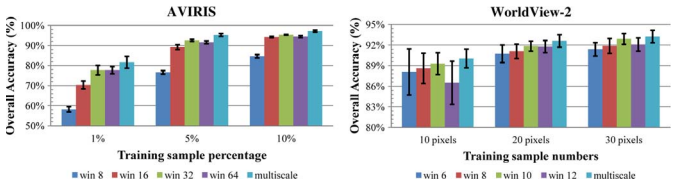

Fig. 7. OA levels (%) for different window sizes.

considered in this study, regardless of the number of training samples.

D. Discussion

The classification accuracy of the OWDWT algorithm is influenced by the window size B. Small windows do not contain sufficient spatial information, whereas large windows may complicate the problem and blur the boundary when they cover multiple objects or classes. An effective strategy for addressing the issue of window size is to concatenate the textural features in multiple windows (denoted "multiscale" in Fig. 7). It is interesting to see that the multiwindow approach outperforms the single-window methods in all the cases.

It should be underlined that this study is quite different from Qian *et al.* [9], which is also a 3-D wavelet texture classification algorithm. The differences are summarized as follows: 1) Qian *et al.* [9] focus on the global texture extraction, i.e., a hyperspectral image is decomposed into a set of sub-images that have the same size as the original one by removing the downsampling step. However, in this study, the windowed or local 3-D wavelet transformation is considered, which is more relevant to the local texture information. 2) A series of specific strategies for 3-D DWT feature extraction and representation is proposed. 3) A comparison is conducted between the proposed algorithms and the state-of-the-art spectral–spatial classification methods. This comparative study is necessary for providing references for the research on the spectral–spatial classification and is important for showing the prospects of the 3-D feature extraction. 4) The proposed OWDWT method outperforms the method of Qian *et al.* [9] in terms of the classification accuracy levels for the AVIRIS data set, which can be attributed to the consideration of local 3-D texture information.

V. Conclusion

In this letter, we have presented a new framework using 3-D wavelet transform for the classification of hyper/multispectral imagery, considering that 3-D textures are able to simultaneously exploit the rich spectral and spatial information in the images. Specifically, three feature extraction methods are investigated, including pixelwise, non-overlapping, and overlapping cube. Moreover, the proposed methods are compared in detail with the state-of-the-art 2-D and 3-D spectral–spatial classification methods (e.g., GLCM, EMPs, and EAPs), and it is shown that the OWDWT method produces the optimal results in terms of both visual inspection and quantitative accuracy. It can be stated that the proposed 3-D wavelet textures exhibit a promising prospect of the 3-D feature extraction for multi/hyperspectral image analysis in future research.

References

[1] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. 110–122, Sep. 2009.

[2] L. Zhang, X. Huang, B. Huang, and P. Li, "A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2950–2961, Oct. 2006.

[3] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.

[4] L. Jun, 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.

[5] 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.

[6] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.

[7] T. C. Bau, S. Sarkar, and G. Healey, "Hyperspectral region classification using a three-dimensional Gabor filterbank," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3457–3464, Sep. 2010.

[8] F. Tsai and J.-S. Lai, "Feature extraction of hyperspectral image cubes using three-dimensional gray-level cooccurrence," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3504–3513, Jun. 2013.

[9] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2013.

[10] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.

[11] S.-T. Hsiang and J. W. Woods, "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank," *Signal Process. Image Commun.*, vol. 16, no. 8, pp. 705–724, May 2001.

[12] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Resonance Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.

[13] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.