Classification Based on 3-D DWT and Decision Fusion for Hyperspectral Image Analysis

Zhen Ye, *Student Member, IEEE*, Saurabh Prasad, *Member, IEEE*, Wei Li, *Student Member, IEEE*,
James E. Fowler, *Senior Member, IEEE*, and Mingyi He, *Member, IEEE*

Abstract—In this letter, a fusion-classification system is proposed to alleviate ill-conditioned distributions in hyperspectral image classification. A windowed 3-D discrete wavelet transform is first combined with a feature grouping—a wavelet-coefficient correlation matrix (WCM)—to extract and select spectral-spatial features from the hyperspectral image dataset. The adjacent wavelet-coefficient subspaces (from the WCM) are intelligently grouped such that correlated coefficients are assigned to the same group. Afterwards, a multiclassifier decision-fusion approach is employed for the final classification. The performance of the proposed classification system is assessed with various classifiers, including maximum-likelihood estimation, Gaussian mixture models, and support vector machines. Experimental results show that with the proposed fusion system, independent of the classifier adopted, the proposed classification system substantially outperforms the popular single-classifier classification paradigm under small-sample-size conditions and noisy environments.

Index Terms—Decision fusion, hyperspectral imagery, multiclassifiers, wavelets.

I. INTRODUCTION

HYPERSPECTRAL imagery (HSI) records hundreds of spectral bands for each pixel, each of which contains values that correspond to the detailed spectrum of reflected light [1]. However, the high dimensionality of HSI also reduces the generalization capability of statistical classifiers, such as the maximum-likelihood estimation (MLE) [2] and the Gaussian mixture model (GMM) [3] classifiers. Since HSI bands are often highly correlated, it is often necessary to project the data onto a lower-dimensional subspace

using dimensionality reduction. Traditional approaches include unsupervised methods, such as principal component analysis [4], and supervised methods, such as Fisher's linear discriminant analysis (LDA) [2] and its many variants [5], [6]. Compared with these classification approaches, support vector machines (SVMs) [7] have been found to be particularly promising for classification of hyperspectral data because of their lower sensitivity to the curse of dimensionality.

Despite high spectral resolution in HSI, the number of training samples is often limited and insufficient to reliably estimate classifier models for each class. To solve this problem, a variety of feature-extraction methods [8]–[10] and classification approaches have been proposed recently. In [11], the classification accuracy with SVM is improved by using stacked generalization and the complementary information of magnitude and shape features. In [12], a divide-and-conquer approach provides a noise-robust framework based on redundant discrete wavelet transforms. In [13], a 3-D discrete wavelet transform (3-D DWT) is combined with the local Fisher's discriminant analysis (LFDA)-GMM strategy [14] for extracting spectral-spatial information while preserving the multimodal structure of the statistical distributions, wherein a windowed 3-D DWT is first employed to extract wavelet features from the HSI.

In this letter, a fusion-classification system based upon a wavelet-coefficient correlation matrix (WCM) used as a feature grouping is proposed. Specifically, 3-D DWT coefficients are extracted as spectral-spatial features for classification. Partitioning based on the WCM breaks down the wavelet feature space into approximately independent contiguous subspaces. Afterwards, classification (LDA-MLE [6], LFDA-GMM [14], and SVM-radial basis function (RBF) [15]) along with decision fusion in the form of majority voting (MV) [16] is then invoked to combine these results into a single classification decision per pixel. The proposed algorithm solves the small-sample-size problem via WCM feature-grouping and also provides noise-robust classification.

The remainder of this letter is organized as follows. Section II presents the windowed 3-D DWT feature-extraction technique and WCM feature-grouping method. Section III describes the formulation of the proposed classification scheme. In Section IV, we describe the employed HSI dataset. A detailed discussion of the experimental results

Manuscript received September 22, 2012; revised October 12, 2012 and January 5, 2013; accepted January 16, 2013. Date of publication April 12, 2013; date of current version November 8, 2013. This work was supported in part by the National Aeronautics and Space Administration under Grant NNX12AL49G, the University of Houston, College of Engineering Startup Funds, and the National Natural Science Foundation of China under Grant 61171154.

Z. Ye and M. He are with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: yzh525@gmail.com; myhe@nwpu.edu.cn).

S. Prasad is with the University of Houston, Houston, TX 77004 USA (e-mail: saurabh.prasad@ieee.org).

W. Li is with the University of California, Davis, CA 95616 USA (e-mail: liwei089@ieee.org).

J. E. Fowler is with the Geosystems Research Institute and the Electrical and Computer Engineering Department, Mississippi State University, Starkville, MS 39762 USA (e-mail: fowler@ece.msstate.edu).

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.2013.2251316

Fig. 1. Windowed block for a 3-D DWT decomposition.

is presented in Section V, while Section VI concludes this letter.

II. FEATURE EXTRACTION AND SELECTION

A. Windowed 3-D DWT

In this letter, a windowed 3-D DWT is employed to extract spectral-spatial features of a hyperspectral image for classification. Specifically, we perform a sliding window analysis, wherein a spatial window of size $B \times B$ pixels is scanned across the hyperspectral imagery. Due to the 3-D DWT transformation, the spatial window size is restricted to be a power of 2. Fig. 1 illustrates the sliding window used for extracting spatial features from HSI. If the window size B is 4, there are 16 pixels in the window.

With each slide along the x - or y -axis, a windowed block of spatial size $B \times B$ and spectral size N will be produced, as depicted in Fig. 1 (N is the number of spectral bands), the 3-D DWT coefficients are then extracted from each $B \times B \times N$ block. Specifically, a single-level dyadic 3-D DWT using Haar filters is applied to each $B \times B \times N$ block. This 3-D DWT, which is implemented in the spatial-row, spatial-column, and spectral-slice directions, results in eight subbands for one block. The baseband, or LLL frequency subband, results from lowpass filtering in all three directions; in this letter, the set of LLL subbands from the blocks constitutes the extracted features used subsequently for classification. The dimensionality of each extracted feature is consequently $N_{\text{LLL}} = B^2 N / 8$.

B. WCM Feature Grouping

Subspace identification is the next step in the proposed multiclassifier and decision-fusion system. It involves intelligent partitioning of the hyperspectral feature space into contiguous subspaces such that each subspace provides increased class separation, while the statistical dependence between subspaces is minimized. In this letter, the WCM is defined as the correlation matrix of the LLL subband of the 3-D DWT extracted from the windowed blocks. We evaluated the 3-D DWT-based methods using all subbands individually and found that the LLL subband provided the best classification performance. For example, consider the results for 3-D DWT-LFDA-GMM as shown in Table I.

Fig. 2. Wavelet-coefficient correlation matrix (WCM) for Indian Pines.

The WCM technique breaks down the high-dimensional wavelet-coefficient feature space into contiguous subspaces. For example, for the Indian Pines dataset ($N = 220$), the feature vector extracted from each block has dimensionality $B^2 N / 8 = 440$. We note that, in this dataset, spectral bands $\{104-108\}$ and $\{150-163\}$ are water-absorption bands and are hence removed. The resulting WCM is depicted in Fig. 2; we see from this WCM that the 440-dimensional feature vector can be naturally partitioned into five groups such that, in each group, these features are highly correlated with one another. These five groups correspond to the five light-colored blocks residing on the diagonal of the WCM, as visible in Fig. 2.

The WCM, which drives the proposed feature partitioning, possesses the advantage of including the 3-D DWT into the feature-extraction process, yielding high classification accuracies at the subspace level. That is, adjacent 3-D DWT coefficients are highly correlated in the feature subspace. The adjacent 3-D DWT feature bands are intelligently grouped by exploiting the WCM structure. With such an approach, the problem of limited training sample size is overcome since the number of training samples required per subspace is substantially lower than that which would be required if all bands and 3-D DWT coefficients were used. On the other hand, noisy/redundant bands are discarded by subspace identification [17]. Furthermore, the low intergroup correlation is favorable for a noise-robust decision fusion of local decisions. This can be attributed to the fact that such a partitioning of the feature space for a multiclassifier system yields superior noise-robust performance for classification tasks, while ensuring robustness to the small-sample-size problem.

III. CLASSIFICATION AND METHODOLOGIES

A. MLE

MLE is a common solution that views system parameters as quantities whose values are fixed but unknown. The maximum-likelihood estimate of their values maximizes the probability of obtaining the samples actually observed. A popular parametric classification typically employed involves invoking an MLE classifier after dimensionality reduction based on Fisher's LDA.

Fig. 3. Flowchart of the proposed multiclassifier and decision-fusion system.

TABLE I
CLASSIFICATION ACCURACY VERSUS SUBBANDS
AND REDUCED DIMENSION

	LLL	LLH	LHL	LHH	HLL	HLH	HHL	HHH
R = 5	71.41	55.59	75.66	55.45	72.47	57.18	73.54	58.38
R = 7	79.65	52.53	81.12	56.38	80.32	57.71	80.98	60.51
R = 9	87.23	53.46	86.17	58.24	86.70	56.38	86.97	60.11
R = 11	87.23	53.86	84.31	56.65	86.57	55.85	84.71	58.64
R = 13	88.03	53.32	83.51	57.98	86.44	53.86	85.64	57.05
R = 15	85.64	53.19	86.17	56.78	86.57	52.26	86.04	56.52
R = 17	84.84	53.19	85.90	55.32	86.17	51.86	86.44	54.39
R = 19	84.57	53.06	85.51	53.46	85.64	51.86	85.64	55.19
R = 21	85.24	52.93	84.97	53.19	85.11	50.13	87.77	53.19
R = 23	84.84	52.53	84.57	52.66	84.44	50.80	87.37	53.32
R = 25	84.04	51.06	85.51	51.20	83.91	49.47	86.84	53.19

B. GMM

A GMM can be viewed as a combination of two or more normal Gaussian distributions. In a typical GMM representation, a probability density function for $X = \{x_i\}_{i=1}^n$ in $\mathbb{R}^d$ is written as the sum of K Gaussian components

$$p(X) = \sum_{k=1}^K \alpha_k N\left(X, \mu_k, \sum_k\right) \quad (1)$$

where $N(X, \mu_k, \sum_k)$ represents the k th Gaussian component of the mixture. K is the number of mixture components, and α_k , μ_k , and $\sum_k$ are the mixing weights, mean vector, and covariance matrix, respectively, of the k th component, which are expressed by the parameter vector $\Theta = \{\alpha_k, \mu_k, \sum_k\}$. Once the optimal number of components, K , per GMM has been determined, the parameters for the mixture model can be estimated by using expectation maximization [18], an iterative optimization strategy. It was recently demonstrated in [14] that LFDA, combined with GMM, resulted in high classification performance for HSI.

C. SVM

Given a training dataset $X = \{x_i\}_{i=1}^n$ in $\mathbb{R}^d$ with class labels $y_i \in \{-1 + 1\}$ and a nonlinear kernel mapping $\phi(\cdot)$, x_i is a pixel vector with a d -dimensional spectrum. The SVM method solves

$$\min_{w, \xi_i, b} \left\{ \frac{1}{2} \|w\|_2^2 + C \sum_i \xi_i \right\} \quad (2)$$

constrained to

$$y_i(\phi^T(x_i) \cdot w + b) \geq 1 - \xi_i \quad \forall i = 1, K, n \quad (3)$$

where w is the normal to the optimal decision hyperplane, n denotes the number of samples, and b represents the closest distance to the origin of the coordinate system. The regularization parameter C controls the generalization capabilities of the classifier, and ξ_i are positive slack variables [7]. This problem is solved by maximizing its Lagrangian dual problem, which consists of maximizing

$$\max_{\alpha_i} \left\{ \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \langle \phi(x_i), \phi(x_j) \rangle \right\} \quad (4)$$

constrained to $0 \leq \alpha_i \leq C$, and $\sum_i \alpha_i y_i = 0$, $i = 1, K, n$, where auxiliary variables α_i are Lagrange multipliers corresponding to constraints in (3). In this letter, the RBF is employed

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (5)$$

where σ is the parameter that we will need to tune experimentally.

D. Proposed System

In this letter, LDA-MLE, LFDA-GMM, and SVM-RBF are employed as the classification strategies. We propose a fusion system that is based on WCM feature-grouping and MV-based decision fusion [16]. As a hard decision-fusion rule, MV is

$$w = \arg \max_{j \in \{1, 2, \dots, C\}} N(j), \quad \text{where } N(j) = \sum_{i=1}^n \alpha_i I(w_i = j) \quad (6)$$

where α_i is the confidence score for the i th classifier, I is the indicator function, w is the class label from one of the C possible classes for the test pixel, i is the classifier index, n is the number of classifiers, and $N(j)$ is the number of times class j was detected in the bank of classifiers.

Fig. 3 shows the flowchart of the proposed multiclassifier and decision-fusion system. First, 3-D wavelet coefficients are extracted by a sliding window. Then, we break down the feature space into approximately independent contiguous subspaces by the WCM-based feature grouping. X^*Y represents the number of original pixels. Following this, multiclassifiers (LDA-MLE, LFDA-GMM, and SVM-RBF, respectively) are applied for obtaining local labels in each group. Finally, a decision-fusion system merges decisions from each classifier in the bank into a single decision by the MV rule.

IV. EXPERIMENTAL SETUP

In this letter, the experiments are conducted using two hyperspectral datasets. The first one is the Indian Pines dataset collected by the National Aeronautics and Space Administration's Airborne Visible/Infrared Imaging Spectrometer sensor. The dataset represents a vegetation-classification scenario with 145×145 pixels and 220 spectral bands. Approximately, 8600 labeled pixels are employed to train and validate/quantify the efficacy of the proposed system. Eight classes (Corn no till, Corn min till, Grass/Pasture, Hay windowed, Soybean no till, Soybean min till, Soybean clean till, Woods) are used in this letter, with an average of 188 training samples per class and 7120 testing samples (the ratio of testing and training samples is about 5:1). The second dataset is the Pavia University data, collected by the reflective optics system imaging spectrometer sensor. The dataset has 103 spectral bands with a spatial size of 610×340 pixels. Approximately 8100 labeled pixels are employed to train and validate/quantify the efficacy of the proposed system. This dataset is partitioned into approximately 1350 training samples and 6750 testing samples (the ratio of testing and training is 5:1).

V. RESULTS AND DISCUSSION

In this section, we applied the WCM feature-grouping and decision-fusion scheme with MLE-, GMM-, and SVM-based classifiers for HSI classification. For comparing the performances of the proposed methods with existing single classifier algorithms (wherein all the WCM coefficients are fed to a single classifier)—namely, LDA-MLE, LFDA-GMM, and SVM-RBF—we conduct experiments on two real-life situations—limited training samples and noisy datasets. We report the performance of these classification-fusion systems as measured by the overall classification accuracy with test data, along with 95% confidence intervals. To match challenging operating conditions, we provide results for a wide range of training samples. Noisy conditions are simulated by adding white Gaussian noise to hyperspectral image data, and the performance measures are reported over a wide range of signal-to-noise ratios (SNRs).

A. Optimizing WCM-Based Algorithms

For the basic single-classifier algorithms, LFDA-GMM and SVM-RBF need to optimize system parameters for any HSI classification task. For LFDA-GMM, we need to find an optimal parameter r that is the reduced dimensionality of the LFDA projection. For SVM-RBF, the kernel parameter σ is an important factor that impacts the generalization ability in the kernel-induced space [14]. We tune r over a wide range from 5 to 30, and from σ 0.1 to 1.0. Finally, the optimal parameters are chosen as those that maximize the training accuracy.

In the following experiment, we discuss the window size B as a parameter for the proposed fusion system. For the Indian Pines dataset, the available training data are further divided into 94 training samples and 94 testing samples per class. Fig. 4 depicts the classification accuracy of the proposed algorithms WCM-LDA-MLE, WCM-LFDA-GMM,

Fig. 4. Classification accuracy (%) of three fusion methods versus window size for Indian Pines.

and WCM-SVM-RBF as a function of window size B (varying from 2 to 10). We can conclude that the optimal value of B for both WCM-LDA-MLE and WCM-LFDA-GMM is 6. For WCM-SVM-RBF, the best classification accuracy is obtained at a value of 10.

B. Comparison Against Conventional Classifiers

In order to quantify the efficacy of the proposed fusion-based classification system, we compare three decision-fusion methods (WCM-LDA-MLE, WCM-LFDA-GMM, and WCM-SVM-RBF) with corresponding single-classifier methods (LDA-MLE [5], LFDA-GMM [14], and SVM-RBF [7]). The baseline of these methods listed above is measured in two experiments representing challenging real-world scenarios. Since labeling training samples for HSI classification is a complicated task, the number of training samples is often insufficient to reliably estimate the classifier models for each class. We study the classification accuracy as a function of training samples over a range from a small number to a reasonably high number. We also conducted an experiment to study the robustness of our system under noisy environments. We assume that the noise is independent of the signal information, and add different amounts of Gaussian noise to both training and testing samples. SNR is used to measure the quality of noisy hyperspectral dataset.

The classification accuracy of the proposed system is studied over an available range of total training samples (from 600 to 1600 for Indian Pines dataset and from 225 to 1350 for University of Pavia dataset). Fig. 5 (top) shows that the proposed fusion methods always produce higher overall classification accuracies, compared to traditional single-classifier methods. For the Indian Pines and the University of Pavia datasets, WCM-LFDA-GMM and WCM-SVM-RBF yield much higher performance, even under small training-sample sizes (e.g., 600 training samples for Indian Pines dataset and 225 training samples for the University of Pavia dataset).

We also conducted an experiment to estimate the noise robustness of our proposed system. These two HSI datasets are corrupted by white Gaussian noise added to each spectral band over a wide SNR range (from 9.3 dB to 36.36 dB

Fig. 5. Overall classification accuracy (%) as a function of training samples (top), and SNR (bottom) for Indian Pines (left) and University of Pavia (right).

for Indian Pines dataset and from 3.99 dB to 27.37 dB for University of Pavia dataset). We studied the sensitivity of the proposed fusion methods relative to conventional methods. As Fig. 5 (bottom) indicates, the proposed WCM-based multi-classifier and decision-fusion system yields classification that outperforms traditional single-classifier approaches. For the Indian Pine dataset, the proposed WCM-SVM-RBF significantly outperforms the other approaches at all SNRs. For the University of Pavia dataset, the proposed WCM-based system also outperforms traditional approaches.

VI. CONCLUSION

In this letter, a windowed 3-D DWT was employed for simultaneous spectral-spatial feature extraction, and a WCM feature-banding method was proposed to break down the wavelet feature space into approximately independent contiguous subspaces. Afterwards, an MV-based decision fusion, combined with a bank of multiclassifiers (LDA-MLE, LFDA-GMM, and SVM-RBF, respectively) was invoked to combine these results into a single classification decision. The experimental results demonstrated that the proposed system required only a small number of training samples for effective classification, while also providing robustness in the presence of additive white Gaussian noise.

REFERENCES

- [1] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Dordrecht, The Netherlands: Kluwer, 2003.
- [2] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: Wiley, 2001.
- [3] L. P. Heck and K. C. Chou, "Gaussian mixture model classifiers for machine monitoring," in *Proc. ICASSP*, 1994, pp. 133–136.
- [4] M. D. Farrell and R. M. Mersereau, "On the impact of PCA dimension reduction for hyperspectral detection of difficult targets," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 192–195, Apr. 2005.
- [5] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, no. 5, pp. 1027–1061, May 2007.
- [6] W. Li, S. Prasad, J. E. Fowler, and L. Mann Bruce, "Locality-preserving discriminant analysis in kernel-induced feature space for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 894–898, Sep. 2011.
- [7] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.
- [8] L. Shen and S. Jia, "Three-dimensional Gabor wavelets for pixel-based hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 5039–5046, Dec. 2011.
- [9] H. R. Kalluri, S. Prasad, and L. Mann Bruce, "Decision-level fusion of spectral reflectance and derivative information for robust hyperspectral land cover classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4047–4058, Nov. 2010.
- [10] G. Rellier, X. Descombes, F. Falzon, and J. Zerubia, "Texture feature analysis using a Gauss–Markov model in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1543–1551, Jul. 2004.
- [11] J. Chen, C. Wang, and R. Wang, "Using stacked generalization to combine SVMs in magnitude and shape feature spaces for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2193–2205, Jul. 2009.
- [12] S. Prasad, W. Li, J. E. Fowler, and L. Mann Bruce, "Information fusion in the redundant-wavelet-transform domain for noise-robust hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 9, pp. 3474–3486, Sep. 2012.
- [13] Z. Ye, S. Prasad, W. Li, J. E. Fowler, and M. He, "Locality-preserving discriminant analysis and Gaussian mixture models for spectral-spatial classification of hyperspectral imagery," in *Proc. WHISPERS*, Shanghai, China, Jun. 2012.
- [14] 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.
- [15] Y. Gu, C. Wang, D. You, Y. Zhang, S. Wang, and Y. Zhang, "Representative multiple kernel learning for classification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2852–2865, Jul. 2012.
- [16] S. Prasad and L. Mann Bruce, "Decision fusion with confidence-based weight assignment for hyperspectral target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1448–1456, May 2008.
- [17] Q. Du, W. Zhu, H. Yang, and J. E. Fowler, "Segmented principal component analysis for parallel compression of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 713–717, Oct. 2009.
- [18] A. Berge and A. H. S. Solberg, "Structured Gaussian components for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3386–3396, Nov. 2006.