ABSTRACT
A decision-fusion approach is introduced for hyperspectral data classification based on minimum-distance classifiers in the wavelet domain. In the proposed approach, multi-scale features of each hyperspectral pixel are extracted by implementing a redundant discrete wavelet transform on the spectral signature. Following this, a pair of minimum distance classifiers—a local mean-based nonparametric classifier and a nearest regularization subspace—are applied on wavelet coefficients at each scale. Classification results are finally merged in a multi-classifier decision-fusion system. Experimental results using real hyperspectral data demonstrate the benefits of the proposed approach—in addition to improved classification performance compared to a traditional single classifier, the resulting classifier framework is effective even for low signal-to-noise-ratio images.

Index Terms—decision fusion, nearest neighbors, hyperspectral data, pattern classification.

1. INTRODUCTION
Hyperspectral imagery (HSI) provides hundreds of narrow contiguous bands which include abundant spectral information about the materials. It is common to use the simplified single-Gaussian or mixture-Gaussian distribution to model HSI [1]. However, such a statistical distribution may suffer due to insufficient training samples (as is common with analysis of HSI data), leading to a suboptimal predictive model.

A simple non-parametric classifier, such as a minimum-distance classifier (MDC) [2], may be preferred, since it does not depend on the underlying class distribution. MDC classifies a testing sample into the class to which the nearest prototype to the sample belongs. The definition of MDC is to compute the global mean of each labeled class and measure the distance to the unlabeled sample according to some metric, such as Euclidean distance. The class with the mean closest to the vector (i.e., with the smallest distance) is the label for the testing sample. Several extensions have been studied for MDC, such as local mean-based nonparametric classifier (LMNC) [3] and nearest regularized subspace (NRS) [4].

Recently, the redundant discrete wavelet transform (RDWT) [5] was introduced to decompose a spectral signature into several levels of undecimated wavelet coefficients in HSI. RDWT is essentially an undecimated version of the DWT [6] which provides robust (e.g., shift-invariant) features by removing the downsampling procedure. RDWT results in a high-dimensional coefficient space since each feature subspace at any scale has the same dimensionality as the original data (i.e., the number of spectral bands in the case of hyperspectral data); the dimensionality of the RDWT-feature space is hence $m$ times larger than the original data, with $m$ being the number of scales employed in RDWT analysis.

In this work, we propose an approach of decision fusion for HSI classification based on LMNC or NRS in the RDWT-feature domain. The RDWT-based approach was first proposed in [5] using statistical classifiers—here, we extend this work to LMNC and NRS. Firstly, multi-scale features of each hyperspectral pixel are extracted by implementing RDWT on spectral signatures. Secondly, LMNC or NRS is considered for wavelet coefficients at each scale where the dimensionality equals the original space. Finally, classification results are merged in a multi-classifier decision-fusion system. It is expected that, at each level of decomposition, the wavelet features of the input data, with or without additional noise, provide different classification performance for the chosen classifier. Two different decision-fusion rules are employed—majority voting (MV) and logarithmic opinion pool (LOGP) [7]. We validate the proposed algorithm using real hyperspectral data to demonstrate the benefits and compare with existing methods.

2. DECISION FUSION IN THE WAVELET DOMAIN
2.1. LMNC
The LMNC [3] classifier is an extension of the original MDC. The essence behind the method is that the local mean vector
of the $k$-nearest neighbors in each class is used for classifying the testing sample. Consider a dataset with training samples $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$ in $\mathbb{R}^d$ ($d$ is the dimensionality, i.e., the number of spectral bands) and class labels $\omega_i \in \{1, 2, \ldots, C\}$, where $C$ is the number of classes, and $n$ is the total number of training samples. Let $n_l$ be the number of available training samples for the $l$th class, $\sum_{l=1}^C n_l = n$. First, the $k$-nearest training samples per class from $\mathbf{X}$ are selected. Following this, the local mean vector, represented as $\tilde{\mathbf{y}}_l$, is calculated using the $k$-nearest neighbor training samples, $\{\mathbf{x}_l^{(1)}, \mathbf{x}_l^{(2)}, \ldots, \mathbf{x}_l^{(k)}\}$:

$$
\tilde{\mathbf{y}}_l = \frac{1}{k} \sum_{j=1}^k \mathbf{x}_l^{(j)}.
$$

Finally, after obtaining the local mean vector per class, the class label of $\mathbf{y}$ is then determined according to the class which minimizes the residual. That is,

$$
\text{class}(\mathbf{y}) = \arg \min_{l=1, \ldots, C} r_l(\mathbf{y}),
$$

where $r_l(\mathbf{y}) = \|\tilde{\mathbf{y}}_l - \mathbf{y}\|_2^2$ is the residual between the mean vector and the corresponding testing sample.

### 2.2. NRS

We categorize the NRS [4] classifier as an extension of LMNC. In NRS, we replace the local mean vector $\tilde{\mathbf{y}}_l$ of (1) using an adaptive weighted representation. An approximation of the testing sample $\mathbf{y}$ is represented via a linear combination of available training samples per class, $\mathbf{X}_l$. That is, for each class, only from the training samples particular to class $l$, the class-specific approximation, $\mathbf{y}_l$, is calculated as $\mathbf{y}_l = \mathbf{X}_l\mathbf{\alpha}_l$, where $\mathbf{X}_l$ is of size $d \times n_l$, and $\mathbf{\alpha}_l$ is an $n_l \times 1$ vector of weighting coefficients. The weight vector $\mathbf{\alpha}_l$ for the linear combination is solved by an $\ell_2$-norm regularization,

$$
\mathbf{\alpha}_l = \arg \min_{\mathbf{\alpha}_l} \|\mathbf{y} - \mathbf{X}_l\mathbf{\alpha}_l\|_2^2 + \lambda \|\Gamma_{l,y}\mathbf{\alpha}_l\|_2^2,
$$

where $\Gamma_{l,y}$ is a biasing Tikhonov matrix specific to each class $l$ as well as the current testing sample $\mathbf{y}$, and $\lambda$ is a global regularization parameter which balances the minimization between the residual and the regularization term. Note that $\mathbf{\alpha}_l^*$ is a representation of $\mathbf{\alpha}_l$ with size of $n_l \times 1$. Specifically, the regularization term is designed in the form

$$
\Gamma_{l,y} = \begin{bmatrix}
\|\mathbf{y} - \mathbf{x}_{l,1}\|_2^2 & 0 & \cdots & 0 \\
0 & \|\mathbf{y} - \mathbf{x}_{l,2}\|_2^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \|\mathbf{y} - \mathbf{x}_{l,n_l}\|_2^2
\end{bmatrix},
$$

where $\mathbf{x}_{l,1}, \mathbf{x}_{l,2}, \ldots, \mathbf{x}_{l,n_l}$ are the columns of matrix $\mathbf{X}_l$ for the $l$th class. Following this, the weight vector $\mathbf{\alpha}_l$ can be recovered in a closed-form solution,

$$
\mathbf{\alpha}_l = (\mathbf{X}_l^T \mathbf{X}_l + \lambda^2 \Gamma_{l,y}^T \Gamma_{l,y})^{-1} \mathbf{X}_l^T \mathbf{y}.
$$

Once we obtain the weight vector, the adaptive weighted representation of $\mathbf{y}$ is $\mathbf{y}_l = \mathbf{X}_l \mathbf{\alpha}_l$. The class label of $\mathbf{y}$ is determined according to (2).

### 2.3. Proposed Classification Framework

Fig. 1 illustrates the flowchart of the proposed classification framework. In Fig. 1, the RDWT results in a new feature space; for the Indian Pine dataset (which will be introduced in Sec. 3), we use specifically six RDWT scales to form the feature space, i.e., $\{c_0, d_0, d_1, d_2, d_3, d_4, d_5\}$. The fine-scale RDWT coefficients, $d_0$, represent a higher-frequency component, providing more image detail; comparatively, $c_0$ provides coarser contour information. The feature subspace of wavelet coefficients at any scale has the same dimensionality as that of the original data. The previously-introduced NRS or LMNC is considered for each feature subspace. Before assigning the final label, classification results are merged in a multi-classifier decision-fusion system.

Two different decision-fusion rules are employed in this work—MV and LOGP. For MV, we count the number of times (represented as $N_l(\mathbf{y})$) class $l$ appears in the outputs of all the feature spaces. The final label of the testing sample is the one with the largest count, i.e.,

$$
\text{class}(\mathbf{y}) = \arg \max_{l=1, \ldots, C} N_l(\mathbf{y}).
$$

In contrast to MV at the class-label level, LOGP provides “soft” fusion at the posterior-probability level. LOGP uses the individual posterior probability $p_l(\omega | \mathbf{y})$ of each classifier to estimate a global membership function,

$$
P(\omega | \mathbf{y}) = \prod_{l=1}^m p_l(\omega | \mathbf{y})^{\alpha_l},
$$

where $m$ is the number of scales for the RDWT, and $\alpha_l$ is uniformly distributed. According to the residual output $r_l(\mathbf{y})$.
of NRS or LMNC, we employ a Gaussian mass function,

\[ p_t(\omega | y) = \exp(-r_t(y)), \quad (8) \]

which indicates that a smaller residual \( r_t(y) \) yields a higher the probability \( p_t(\omega | y) \).

3. EXPERIMENTS

The experimental hyperspectral dataset employed was acquired using NASA's AVIRIS sensor and was collected over northwest Indiana's Indian Pine test site in June 1992. The image represents a vegetation-dominated scenario with 145 × 145 pixels and 220 bands in the 0.4- to 2.45-μm region. The spatial resolution is 20 m. In our experiment, there are 8 classes to be classified, and, in total, 400 training samples (50 per class) and 8624 testing samples (1460, 834, 497, 489, 968, 2468, 614, and 1294 per class, respectively).

For this experimental data, we use 7 wavelet decomposition levels of RDWT transformation (the number of levels depends on the number of spectral bands, i.e., \( \text{floor}(\log_2 220) = 7 \)). We chose the db4 wavelet basis. First, we investigate the classification performance of NRS and LMNC in different subspaces (e.g., \{c0, d0, d1, d2, d3, d4, d5\}) of wavelet coefficients as illustrated in Fig. 2. For NRS, the regularization parameter \( \lambda \) is an important value, and, for LMNC, the number (e.g., \( k \)) of neighbors significantly affects the performance. Here, we empirically set \( \lambda = 1 \) for NRS and \( k = 3 \) for LMNC. From the results in Fig. 2, we observe that neither the low-frequency feature subspace (e.g., \( c0 \)) nor the high-frequency feature subspace (e.g., \( d5 \)) provides the best classification accuracy, while the mid-frequency feature subspace (e.g., \( d4 \)) does.

For the remainder of this paper, we use the following notations: NRS and LMNC combined with the MV and LOGP decision-fusion paradigms are denoted as NRS-MV, NRS-LOGP, LMNC-MV, and LMNC-LOGP, respectively. Additionally, the maximum outputs of NRS/LMNC implemented on individual wavelet coefficients subspace are denoted as NRS-MAX/LMNC-MAX. In other words, the best performance in Fig. 2 is considered as reference. We compare the classification performance of the proposed strategies with the original NRS and LMNC based on spectral signatures.

Fig. 3 illustrates classification accuracy versus varying parameters for the aforementioned methods using the experimental data. From the results, we have the following observations: (1) NRS-based methods are superior to LMNC-based methods, which verifies the advantages of the adaptive weighted representation against the local mean vector; (2) NRS-MAX and LMNC-MAX outperform NRS and LMNC, respectively, which indicates that the wavelet feature space provides higher discriminant power than does the original data space; (3) the classification accuracy of NRS-LOGP using the optimal \( \lambda \) (e.g., 0.3) is 83.55% which is obviously higher than that of NRS (e.g., 80.18% as \( \lambda = 0.9 \)), and the accuracy of LMNC-MV using the optimal \( k \) (e.g., 4) is 73.04% which is also higher than that of LMNC, resulting in an improvement of approximately 4%.

Then, we study the proposed method in highly noise-corrupted environments (measured by studying the signal-to-noise ratio (SNR)) [8]. Additional artificial noise is added with the assumption that the noise is independent of the signal information and in the form of white Gaussian noise of varying power. Fig. 3 illustrates classification accuracy versus varying parameters for the aforementioned methods with SNR = 30 dB. Basically, the performance is worse than those in Fig. 3. However, the proposed methods still outperform the existing methods.

4. CONCLUSIONS

In this paper, we proposed an extension to a recently developed RDWT-based multi-classifier, decision-fusion algorithm by combining the framework with NRS/LMNC. Preliminary studies have provided evidence on the discriminant power of the wavelet domain since the proposed classification framework achieves higher classification accuracy compared to the original NRS and LMNC. In future work, the proposed methods will be validated on other datasets under severely noisy conditions and with fewer training samples.

5. REFERENCES

Fig. 3. Classification accuracy versus varying parameters for the proposed methods as well as the original NRS and LMNC using the experimental data.

Fig. 4. Classification accuracy versus varying parameters for the proposed methods as well as the original NRS and LMNC under noisy condition (30 dB) using the experimental data.


