Spatial-Spectral Hyperspectral Classification Using Local Binary Patterns and Markov Random Fields

Zhen Ye, a,* James E. Fowler, b Lin Bai a

a Chang'an University, School of Electronics and Control Engineering, Xi'an 710064, China
b Mississippi State University, Distributed Analytics and Security Institute, Geosystems Research Institute, Department of Electrical and Computer Engineering, Mississippi State, MS 39762, USA

Abstract: Local binary patterns have been used extensively to yield spatial features for the classification of general imagery, and a few recent works have applied them to the classification of hyperspectral imagery. While the conventional local-binary-pattern formulation employs only the signs of differences between a central pixel and its surrounding neighbors, it has been recently demonstrated that the difference magnitudes also possess discriminative information. Consequently, a sign-and-magnitude local binary pattern is proposed to provide a spatial-spectral class-conditional probability for a Bayesian maximum a posteriori formulation of hyperspectral classification wherein the prior probability is provided by a Markov random field. Experimental results demonstrate that performance of the proposed approach is superior to that of other state-of-the-art algorithms, tending to result in smoother classification maps with fewer erroneous outliers even in the presence of noise.

Keywords: hyperspectral classification, local binary pattern, support vector machine, Markov random field.

1 Introduction

Traditionally, classification of hyperspectral imagery (HSI) has focused on the pixelwise application of classifiers operating on spectral information exclusively. However, there has been increasing interest in recent years in classification that exploits both spatial and spectral information. While there are several strategies that have been proposed for such spatial-spectral classification, one popular approach adopts a Bayesian maximum a posteriori (MAP) formulation coupled with a Markov random field (MRF) (e.g., Refs. 1-4). Specifically, for an n-sample dataset \( X = \{x_i\}_{i=1}^n \), we wish to find the image of class labels, \( y = \{y_i\}_{i=1}^n \), where \( y_i \in \{1, 2, \ldots, k\} \). By Bayes' rule, the posterior distribution of the class labels is

\[
p(y \mid X) = \frac{p(y)p(X \mid y)}{p(X)}.
\]

(1)

Since \( X \) is given, \( p(X) \) can be considered to be constant and can be disregarded, while \( p(y) \) is the prior probability of the classes, and \( p(X \mid y) \) is the class-conditional probability of \( X \). Applying a MAP formulation to Eq. 1, we arrive at the estimated class image being

\[
y = \arg \max_{y \in \{1, \ldots, k\}} \left( \log p(X \mid y) + \log p(y) \right).
\]

(2)

In the typical application of Eq. 2 to spatial-spectral HSI classification, \( p(y) \) is considered a spatial class prior and is often provided by an MRF in order to promote the assigning of neighboring pixels to the same class. Naturally then, the class-conditional probability, \( p(X \mid y) \), is typically designed to exploit spectral information, usually in the form of a pixelwise probability

* Address all correspondence to, Zhen Ye, E-mail: zhenye@chd.edu.cn
estimate driven by the spectral signature of the pixel being classified. To this end, support vector
machines (SVMs)\textsuperscript{1,2}, multinomial logistic regression (MLR)\textsuperscript{3}, and Gaussian mixture models
(GMMs)\textsuperscript{4} have been used to provide $p(X \mid y)$ for Eq. 2. In each of these formulations, however,
this class-conditional probability is devoid of spatial information and is, instead, derived exclusively from spectral signatures. While the partitioning of Eq. 2 into a spectral-only term and a spatial-only term makes some sense conceptually, it is not necessarily optimal. Indeed, the MRF prior tends to promote a piecewise smooth segmentation of the image—to increase the homogeneity of a region, the MRF may override otherwise correct results from the spectrally-driven class-conditional probability if the latter provides only relatively weak support (i.e., $p(X \mid y)$ is maximum for the correct class labeling but not by a wide margin over other labelings).

For HSI classification, extracting both spatial and spectral features can improve accuracy. For example, two-dimensional Gabor features extracted in principal-component subspaces have been proven efficient due to their ability to represent useful spatial and spectral information\textsuperscript{5,6}, while, in our previous work (Refs. 7 and 8), a three-dimensional discrete wavelet transform was proposed for spatial-spectral feature extraction, providing robustness in the presence of additive white Gaussian noise. Another technique, the extended morphological profile (EMP)\textsuperscript{9-11}, has been created from principal components in order to extract informative spatial-spectral features from remote-sensing data. Additionally, a number of techniques have used a local binary pattern (LBP)\textsuperscript{12} and its variants\textsuperscript{13,14} to successfully encapsulate the spatial structure of local image texture information for rotation-invariant texture classification in general pattern-recognition and computer-vision problems; subsequently, Refs. 15 and 16 adopted LBP for an effective spatial-feature extraction in the specific context of spatial-spectral HSI classification.

In this paper, we couple an LBP-driven class-conditional probability with an MRF-driven prior
probability in the MAP formulation of Eq. 2 to yield spatial-spectral HSI classification. As the first
contribution, we devise an LBP operator yielding a spatial-spectral feature that feeds into an SVM
to provide the class-conditional probability $p(X \mid y)$, in contrast to the more traditional approach
which employs a spectral-only $p(X \mid y)$. In addition, while the usual formulation of LBP involves
a binary pattern of signs of differences around a center point, later work (e.g., Ref. 14) has
demonstrated improvement by incorporating difference-magnitude information into the LBP as well, although empirical evaluation was restricted to grayscale visible-band imagery only. Here, our proposed LBP operator also incorporates both signs and magnitudes, demonstrating the effective discriminative ability of magnitude-based LBP information, which we believe is a first in the application of HSI classification. This stands in contrast to Refs. 15 and 16 which employ the traditional signs-only LBP for HSI. Furthermore, since principal component analysis (PCA) has long been a favorite methodology for HSI dimensionality reduction, we apply our proposed LBP operator in a reduced-dimensionality space resulting from a spectral principal-component projection. This also departs from prior work wherein Ref. 15 applies LBP in only a single prechosen spectral band, while Ref. 16 uses a handful of optimally determined bands.

The second contribution of this paper is the coupling of our proposed spatial-spectral LBP-
driven class-conditional probability with an MRF-driven spatial prior probability in the MAP
framework represented by Eq. 2. In empirical results, we demonstrate that the spatial refining of
the classification results that the MRF provides enhanced performance beyond that obtained from
the LBP-based spatial-spectral feature alone.

The remainder of this paper is organized as follows. In Sec. 2, we introduce our proposed
scheme, including the spatial-spectral LBP operator based on signs and magnitudes as well as its
coupling with the MRF prior. Then, in Sec. 3, we present an extensive body of experimental evaluations, following with Sec. 4 wherein we make several concluding remarks.

2 Proposed Approach

In this paper, we propose LBP-based feature extraction applied in PCA-projected subspaces for spatial-spectral classification. The LBP features consist of both sign and magnitude features, and we train an SVM on a suitable training set of the LBP feature vectors with known class labels, resulting in class-conditional probabilities $p(X|y)$. More specifically, to classify $n$-sample testing set $X=\{x_i\}_{i=1}^n$, we assume conditional independence of the features given the class labels so that the class-conditional probability $p(X|y)$ in Eq. 1 becomes $p(X|y) = \prod_{i=1}^n p(x_i|y_i)$, where the $p(x_i|y_i)$ probabilities are provided by the SVM (see Sec. 2.2 SVM below). With the prior probability $p(y)$ given by the MRF (see Sec. 2.3 MRF below), the MAP estimation of (2) becomes

$$\hat{y} = \max_{y \in \{1, \ldots, k\}^n} \left\{ \sum_{i=1}^n \log p(x_i|y_i) + \mu \sum_{(i,j) \in c} \delta(y_i - y_j) \right\}$$

(3)

where $\mu$ is a smoothing parameter, and $\delta(y)$ is the Kronecker delta function ($\delta(y) = 1$ if $y=0$, and $\delta(y) = 0$ if $y \neq 0$). As in Ref. 3, we compute (3) via a graph-cut algorithm; refer to Ref. 3 for more detail. A flowchart of the proposed approach is depicted in Fig. 1. Below, we discuss the various components of the proposed approach in more detail.

---

Fig. 1 Flowchart of the proposed classification scheme.
2.1 Magnitude-Enhanced LBP

As a form of spatial feature extraction, LBP is an efficient operator that characterizes local image structure by encoding the local differences, \( d_i = g_i - g_c \), between a grayscale center pixel \( g_c \) and its grayscale neighbors \( \{ g_i \}_{i=0}^{N-1} \), where \( N \) is the total number of neighbors lying on the neighborhood circle of radius \( R \) from the center pixel. A complete description of symmetric neighbor sets for different \((N,R)\) pairs is discussed in Ref. 12. We note that, as originally developed, LBP was developed for a single-band, grayscale image composed of grayscale pixels \( g_i \). In our use here, we will be deploying LBP on a \( B \)-band image composed of \( B \)-dimensional pixel vectors \( \mathbf{x}_i \)—more details on how this is done can be found below. First, however, we describe the fundamental grayscale LBP formulation in which the set of local differences is used to generate an LBP code which is in turn used as a feature for classification.

In the original formulation of LBP (i.e., Ref. 12), signs of the local differences are used exclusively to generate the feature. More specifically, the LBP feature for center pixel \( g_c \) is

\[
L_{\text{LBP}}(g_c) = \sum_{i=0}^{N-1} s(g_i - g_c)2^i, \tag{4}
\]

where \( s(\cdot) \) is the sign function. Eq. 4 indicates that the local neighborhood is effectively thresholded with the gray value of the center pixel, producing \( 2^N \) distinct LBP-code values.

While the sign-based formulation of the original LBP provides an effective spatially driven feature, Ref. 14 argues that additional discriminative information is available in the difference magnitudes as well. That is, we can express the local differences \( d_i \) for center pixel \( g_c \) in sign-magnitude form,

\[
d_i = s_i \cdot m_i, \quad \text{where} \quad \begin{cases} s_i = \text{sign}(d_i) \\ m_i = |d_i| \end{cases} \tag{5}
\]

where \( s_i \) denotes the sign component (\( s_i = 1 \) if \( d_i \geq 0 \), and \( s_i = 0 \) if \( d_i < 0 \)), and \( m_i \) denotes the magnitude component, and then define a magnitude-based LBP (LBP_M) operator as

\[
L_{\text{LBP}_M}(g_c) = \sum_{i=0}^{N-1} t(m_i, \bar{m})2^i \tag{6}
\]

where \( t(m_i, \bar{m}) = 1 \) if \( m_i \geq \bar{m} \), and \( t(m_i, \bar{m}) = 0 \) if \( m_i < \bar{m} \). As in Ref. 14, \( \bar{m} \) is a threshold which is chosen to be the mean value of the \( m_i \) values from the whole image. While Ref. 14 observes that the sign component of the differences typically preserves more local structure information, the magnitude component can contribute some additional discriminative information. Fig. 2 shows an example of a \( 3 \times 3 \) local pixel area containing an LBP neighborhood circle of radius \( R = 1 \); here, the neighborhood comprises \( N = 8 \) neighbors. By comparing the value of center pixel \( g_c \) with the 8 neighbors, the local differences are calculated [Fig. 1(c)], and the sign component [Fig. 1(d)] is coded by the LBP operator as binary labels [Fig. 1(e)], while the magnitude component [Fig. 1(f)] is coded by the LBP_M operator (also resulting in a binary feature vector).
Fig. 2 (a) Center pixel and its 8 circular neighbors with radius $R = 1$; (b) the $3 \times 3$ sample block; (c) local differences; (d) sign component; (e) binary sign feature; and (f) magnitude component.

Fig. 3 illustrates a comparison between the LBP sign and magnitude features for band 63 of the University of Pavia dataset (described in detail in Sec. 3). It can be observed that the sign-features image is much more detailed, comprising a significant degree of local spatial information (such as edges, corners, and spots). The magnitude-features image, on the other hand, provides additional, complementary information.

In order to combine the sign and magnitude features into a single feature for current pixel $g_c$, we follow Ref. 14 and calculate histograms for the sign and magnitude features over a $W \times W$ patch centered at $g_c$. The histograms are calculated over the patch for the sign and magnitude features separately, and then the histograms are concatenated to yield the final feature vector.

We also observe that, in addition to the magnitude features considered here, Ref. 14 proposes incorporating a thresholded version of the center pixel of the neighbor $g_c$ into what Ref. 14 calls a “completed LBP.” However, in our empirical observations, we have found that this center pixel
offers little discriminative information for HSI. Consequently, we consider only the sign and magnitude features as proposed here, which we call magnitude-enhanced LBP (MELBP).

Finally, we note that the preceding discussion formulates LBP and MELBP for a single-band, grayscale image. For a $B$-band HSI dataset, we first apply dimensionality reduction in the form of PCA prior to MELBP feature extraction, retaining $P$ principle components as the reduced dimensionality ($P < B$). MELBP is applied within each principle-component image independently (i.e., each principal-component image is treated as a grayscale image composed of grayscale pixels $g_i$), and the resulting MELBP features are concatenated to form a single feature vector to drive the subsequent SVM.

Parameters for the proposed MELBP-based scheme—i.e., the number of principle components ($P$), the patch size ($W$), the number of neighbors ($N$), and the neighborhood-circle radius ($R$)—are empirically tuned over wide ranges to obtain optimal classification results. The optimal parameters are chosen to maximize the overall classification accuracy (OA) over the training set†.

For the University of Pavia dataset, we find that $P = 6$ and $W = 17$ are optimal; for the Indian Pines dataset, $P = 10$ and $W = 33$ are optimal; and for the Salinas Valley dataset, $P = 6$ and $W = 27$ are optimal, as shown in Table 1. As for the LBP-neighborhood parameter set, $(N, R)$, Table 2 indicates that $N = 8$ is found to be optimal for the University of Pavia and the Salinas Valley datasets, and, for Indian Pines, it is 6, while the optimal $R$ for the three experimental datasets is 1. Of the four parameters, $P$ and $N$ impact the computational complexity of the MELBP approach. To quantify this effect, execution times (in seconds) are provided in Tables 1 and 2 (all experiments were carried out using MATLAB on a 2.8GHz machine with 8 GB of RAM).

† We note that, in order to expedite computation, the training sets used in Tables 1 and 2 to determine optimal parameter settings are half the size of those used in later results; i.e., the number of training samples per class are half those shown in Tables 3–5.
### Table 1 OA (%) versus varying $W$ and $P$ for the three datasets.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Patch size</th>
<th>Number of principle components</th>
<th>$P=4$</th>
<th>$P=6$</th>
<th>$P=8$</th>
<th>$P=10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Pavia</td>
<td>$W=11$</td>
<td></td>
<td>89.30</td>
<td>89.86</td>
<td>90.11</td>
<td>89.98</td>
</tr>
<tr>
<td></td>
<td>$W=13$</td>
<td></td>
<td>91.20</td>
<td>91.73</td>
<td>92.21</td>
<td>92.00</td>
</tr>
<tr>
<td></td>
<td>$W=15$</td>
<td></td>
<td>90.75</td>
<td>92.64</td>
<td>92.69</td>
<td>92.83</td>
</tr>
<tr>
<td></td>
<td>$W=17$</td>
<td></td>
<td>91.17</td>
<td>93.39</td>
<td>92.76</td>
<td>92.62</td>
</tr>
<tr>
<td></td>
<td>$W=19$</td>
<td></td>
<td>90.80</td>
<td>92.66</td>
<td>92.61</td>
<td>92.65</td>
</tr>
<tr>
<td></td>
<td>$W=21$</td>
<td></td>
<td>90.85</td>
<td>91.97</td>
<td>92.59</td>
<td>92.18</td>
</tr>
<tr>
<td></td>
<td>Time(s)</td>
<td></td>
<td>105.8</td>
<td>146.7</td>
<td>201.5</td>
<td>275.4</td>
</tr>
<tr>
<td>Indian Pines</td>
<td>$W=25$</td>
<td></td>
<td>91.10</td>
<td>91.05</td>
<td>91.23</td>
<td>91.10</td>
</tr>
<tr>
<td></td>
<td>$W=27$</td>
<td></td>
<td>91.18</td>
<td>91.00</td>
<td>91.29</td>
<td>91.49</td>
</tr>
<tr>
<td></td>
<td>$W=29$</td>
<td></td>
<td>91.46</td>
<td>91.23</td>
<td>91.50</td>
<td>91.65</td>
</tr>
<tr>
<td></td>
<td>$W=31$</td>
<td></td>
<td>92.34</td>
<td>92.17</td>
<td>92.12</td>
<td>92.54</td>
</tr>
<tr>
<td></td>
<td>$W=33$</td>
<td></td>
<td>91.43</td>
<td>91.90</td>
<td>92.43</td>
<td>92.73</td>
</tr>
<tr>
<td></td>
<td>$W=35$</td>
<td></td>
<td>90.33</td>
<td>90.97</td>
<td>91.62</td>
<td>91.90</td>
</tr>
<tr>
<td></td>
<td>Time(s)</td>
<td></td>
<td>23.1</td>
<td>35.0</td>
<td>47.8</td>
<td>56.6</td>
</tr>
<tr>
<td>Salinas Valley</td>
<td>$W=19$</td>
<td></td>
<td>88.33</td>
<td>91.84</td>
<td>92.18</td>
<td>91.45</td>
</tr>
<tr>
<td></td>
<td>$W=21$</td>
<td></td>
<td>90.25</td>
<td>91.93</td>
<td>92.22</td>
<td>92.96</td>
</tr>
<tr>
<td></td>
<td>$W=23$</td>
<td></td>
<td>91.46</td>
<td>92.11</td>
<td>93.14</td>
<td>93.16</td>
</tr>
<tr>
<td></td>
<td>$W=25$</td>
<td></td>
<td>91.68</td>
<td>92.48</td>
<td>93.71</td>
<td>93.43</td>
</tr>
<tr>
<td></td>
<td>$W=27$</td>
<td></td>
<td>91.82</td>
<td>93.73</td>
<td>93.54</td>
<td>93.50</td>
</tr>
<tr>
<td></td>
<td>$W=29$</td>
<td></td>
<td>92.42</td>
<td>93.38</td>
<td>93.50</td>
<td>93.69</td>
</tr>
<tr>
<td></td>
<td>Time(s)</td>
<td></td>
<td>83.1</td>
<td>124.0</td>
<td>162.8</td>
<td>202.6</td>
</tr>
</tbody>
</table>

### Table 2 OA (%) versus varying $R$ and $N$ for the three datasets.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Radius</th>
<th>Neighbors</th>
<th>$N=4$</th>
<th>$N=6$</th>
<th>$N=8$</th>
<th>$N=10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Pavia</td>
<td>$R=1$</td>
<td></td>
<td>92.50</td>
<td>92.97</td>
<td>93.39</td>
<td>92.56</td>
</tr>
<tr>
<td></td>
<td>$R=2$</td>
<td></td>
<td>91.52</td>
<td>92.47</td>
<td>92.69</td>
<td>92.05</td>
</tr>
<tr>
<td></td>
<td>$R=3$</td>
<td></td>
<td>90.59</td>
<td>92.18</td>
<td>91.79</td>
<td>91.66</td>
</tr>
<tr>
<td></td>
<td>$R=4$</td>
<td></td>
<td>90.10</td>
<td>91.16</td>
<td>91.41</td>
<td>91.34</td>
</tr>
<tr>
<td></td>
<td>Time(s)</td>
<td></td>
<td>79.9</td>
<td>110.8</td>
<td>167.8</td>
<td>287.2</td>
</tr>
<tr>
<td>Indian Pines</td>
<td>$R=1$</td>
<td></td>
<td>92.03</td>
<td>92.80</td>
<td>92.73</td>
<td>92.63</td>
</tr>
<tr>
<td></td>
<td>$R=2$</td>
<td></td>
<td>92.28</td>
<td>92.62</td>
<td>92.70</td>
<td>92.27</td>
</tr>
<tr>
<td></td>
<td>$R=3$</td>
<td></td>
<td>92.14</td>
<td>91.97</td>
<td>91.74</td>
<td>91.74</td>
</tr>
<tr>
<td></td>
<td>$R=4$</td>
<td></td>
<td>91.52</td>
<td>91.85</td>
<td>91.58</td>
<td>91.61</td>
</tr>
<tr>
<td></td>
<td>Time(s)</td>
<td></td>
<td>41.2</td>
<td>44.0</td>
<td>52.8</td>
<td>64.6</td>
</tr>
<tr>
<td>Salinas Valley</td>
<td>$R=1$</td>
<td></td>
<td>92.90</td>
<td>93.29</td>
<td>93.73</td>
<td>93.20</td>
</tr>
<tr>
<td></td>
<td>$R=2$</td>
<td></td>
<td>90.84</td>
<td>92.12</td>
<td>91.19</td>
<td>90.07</td>
</tr>
<tr>
<td></td>
<td>$R=3$</td>
<td></td>
<td>90.12</td>
<td>90.76</td>
<td>90.61</td>
<td>85.64</td>
</tr>
<tr>
<td></td>
<td>$R=4$</td>
<td></td>
<td>91.03</td>
<td>91.89</td>
<td>89.60</td>
<td>90.74</td>
</tr>
<tr>
<td></td>
<td>Time(s)</td>
<td></td>
<td>105.9</td>
<td>114.8</td>
<td>130.8</td>
<td>165.2</td>
</tr>
</tbody>
</table>
2.2 SVM

Once spatial features are extracted following the MELBP procedure outlined above, an SVM is employed to yield the class-conditional distribution $p(X \mid y)$. Specifically, given a training dataset $X' = \{x_i\}_{i=1}^{n'}$ of $d$-dimensional feature vectors $x_i \in \mathbb{R}^d$, class labels $y_i \in \{-1, +1\}$, and nonlinear kernel mapping $\phi(\cdot)$, an SVM is designed to solve

$$\min_{w, \xi, b} \left\{ \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \right\}$$

constrained by $y_i' (\phi(x_i') \cdot w + b) \geq 1 - \xi_i$, $\forall i = 1, \ldots, n'$, where $W$ is the normal to the optimal decision hyperplane, $n'$ is the number of training samples, and $b$ is an offset. The regularization parameter $C$ controls the generalization capabilities of the classifier, and $\xi = [\xi_1 \cdots \xi_n]$ is a vector of positive slack variables. Typically, Eq. 6 is solved via its Lagrangian dual problem.

Being inherently a binary classifier, an SVM alone does not produce probability estimates for the individual classes. However, logistic regression is commonly fit to SVM classifier scores which transforms the SVM output into a probability distribution. Here, we use such logistic regression in conjunction with SVM as implemented by libSVM to obtain the class-conditional probabilities $p(X \mid y)$ in Eq. 2. We use a radial basis function (RBF) for kernel $\phi(\cdot)$.

2.3 MRF

For an $n'$-sample testing dataset $X' = \{x_i\}_{i=1}^{n'}$, we use an MRF to describe the probability of the field of classes, $y = \{y_i\}_{i=1}^{n'}$. As described in Refs. 3 and 4, the joint probability density for an MRF is a Gibbs distribution in the form of

$$p(y) = \frac{1}{Z} \exp \left( U(y) \right)$$

where $Z$ is a normalization constant, energy function $U(y) = \mu \sum_{(i,j) \in C} \delta(y_i - y_j)$ sums clique potentials $\delta(y_i - y_j)$ over the set of cliques $C$, and $\mu$ is a smoothing parameter. We define a clique as a set of neighboring pixels as is done in Ref. 3. Consequently, Eq. 7 effectuates a spatial prior that reflects that it is highly probable that adjacent pixels belong to the same class; consequently, Eq. 7 encourages a piecewise-smooth classification map.

3 Experimental results

3.1 Experimental data

We now evaluate the performance of the proposed scheme on several HSI datasets. The first experimental dataset used is the University of Pavia dataset collected by a Reflective Optics System Imaging Spectrometer (ROSIS) sensor. The dataset has a spatial size of $610 \times 340$ pixels with 103 spectral bands after removing noisy bands; the geometric resolution is 1.3 m/pixel. We employ nine classes from the dataset (see the ground-truth map in Fig. 4), using 60 labeled pixels per class for training and a total of 42,776 pixels for testing (see the information in Table 3).
The second experimental dataset is the Indian Pines dataset collected by an Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor; this dataset has a spatial size of 145 × 145 pixels over 220 spectral bands with a geometric resolution of 20 m/pixel. We employ 16 classes from the dataset (see the ground-truth map in Fig. 5), using a total of 734 labeled pixels for training and a total of 10,249 pixels for testing (see the information in Table 4).

The final experimental dataset is the Salinas Valley dataset collected by the AVIRIS sensor; this dataset has 512 × 217 pixels over 224 spectral bands, with a geometric resolution of 20 m/pixel. We employ 16 classes from the dataset (see the ground-truth map in Fig. 8), using 20 labeled pixels per class for training and a total of 53,649 pixels for testing (see the information in Table 5).

<table>
<thead>
<tr>
<th>Table 3</th>
<th>The training and testing classes for the University of Pavia dataset.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>Classes</td>
</tr>
<tr>
<td>1</td>
<td>Asphalt</td>
</tr>
<tr>
<td>2</td>
<td>Meadows</td>
</tr>
<tr>
<td>3</td>
<td>Gravel</td>
</tr>
<tr>
<td>4</td>
<td>Trees</td>
</tr>
<tr>
<td>5</td>
<td>Painted Metal Sheets</td>
</tr>
<tr>
<td>6</td>
<td>Bare Soil</td>
</tr>
<tr>
<td>7</td>
<td>Bitumen</td>
</tr>
<tr>
<td>8</td>
<td>Self-Blocking Bricks</td>
</tr>
<tr>
<td>9</td>
<td>Shadows</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>The training and testing classes for the Indian Pines dataset.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>Classes</td>
</tr>
<tr>
<td>1</td>
<td>Alfalfa</td>
</tr>
<tr>
<td>2</td>
<td>Corn-no till</td>
</tr>
<tr>
<td>3</td>
<td>Corn-min till</td>
</tr>
<tr>
<td>4</td>
<td>Corn</td>
</tr>
<tr>
<td>5</td>
<td>Grass/Pasture</td>
</tr>
<tr>
<td>6</td>
<td>Grass/Trees</td>
</tr>
<tr>
<td>7</td>
<td>Grass/Pasture-moved</td>
</tr>
<tr>
<td>8</td>
<td>Hay-windowed</td>
</tr>
<tr>
<td>9</td>
<td>Oats</td>
</tr>
<tr>
<td>10</td>
<td>Soybean-no till</td>
</tr>
<tr>
<td>11</td>
<td>Soybean-min till</td>
</tr>
<tr>
<td>12</td>
<td>Soybean-clean till</td>
</tr>
<tr>
<td>13</td>
<td>Wheats</td>
</tr>
<tr>
<td>14</td>
<td>Woods</td>
</tr>
<tr>
<td>15</td>
<td>Building-Grass-Trees-Drives</td>
</tr>
<tr>
<td>16</td>
<td>Stone-steel Towers</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>
Table 5  The training and testing classes for the Salinas Valley dataset.

<table>
<thead>
<tr>
<th>No.</th>
<th>Classes</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Broccoli-green-weeds-1</td>
<td>20</td>
<td>2,009</td>
</tr>
<tr>
<td>2</td>
<td>Broccoli-green-weeds-2</td>
<td>20</td>
<td>3,726</td>
</tr>
<tr>
<td>3</td>
<td>Fallow</td>
<td>20</td>
<td>1,976</td>
</tr>
<tr>
<td>4</td>
<td>Fallow-rough-plow</td>
<td>20</td>
<td>1,394</td>
</tr>
<tr>
<td>5</td>
<td>Fallow-smooth</td>
<td>20</td>
<td>2,678</td>
</tr>
<tr>
<td>6</td>
<td>Stubble</td>
<td>20</td>
<td>3,959</td>
</tr>
<tr>
<td>7</td>
<td>Celery</td>
<td>20</td>
<td>3,579</td>
</tr>
<tr>
<td>8</td>
<td>Grapes-untrained</td>
<td>20</td>
<td>11,271</td>
</tr>
<tr>
<td>9</td>
<td>Soil-vineyard-develop</td>
<td>20</td>
<td>6,203</td>
</tr>
<tr>
<td>10</td>
<td>Corn-senesced-green-weeds</td>
<td>20</td>
<td>3,278</td>
</tr>
<tr>
<td>11</td>
<td>Lettuce-romaine-4wk</td>
<td>20</td>
<td>1,068</td>
</tr>
<tr>
<td>12</td>
<td>Lettuce-romaine-5wk</td>
<td>20</td>
<td>1,927</td>
</tr>
<tr>
<td>13</td>
<td>Lettuce-romaine-6wk</td>
<td>20</td>
<td>916</td>
</tr>
<tr>
<td>14</td>
<td>Lettuce-romaine-7wk</td>
<td>20</td>
<td>1,070</td>
</tr>
<tr>
<td>15</td>
<td>Vinyard-untrained</td>
<td>20</td>
<td>7,268</td>
</tr>
<tr>
<td>16</td>
<td>Vinyard-vertical-trellis</td>
<td>20</td>
<td>1,807</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>320</td>
<td>5,4129</td>
</tr>
</tbody>
</table>

3.2 Results and Discussion

We now empirically evaluate the classification performance of the proposed approach and compare to other state-of-the-art algorithms. Specifically, we apply a conventional SVM classifier with the RBF kernel directly to spectral signatures (using a one-against-all strategy for multiclass classification), which we denote simply as “SVM”. Additionally, we employ a similar SVM pixelwise classifier followed by a separate stage of MRF-based spatial-regularization postprocessing as described in Ref. 2, which we call “SVM-MRF.” We also evaluate several strategies using common approaches to spatial-spectral feature extractions: namely, a technique which uses two-dimensional Gabor features in a PCA-projected subspace with SVM (“Gabor-SVM”), a strategy using extended-morphological-profile features with SVM (“EMP-SVM”), as well as our previous work consisting of a 3D-DWT feature extraction with SVM (“3DDWT-SVM”).

We also compare to two LBP-based approaches to spatial-spectral classification. Specifically, the original signs-only LBP operator applied in a PCA-projected subspace is used to produce spatial-spectral features to drive SVM in an approach we denote as “LBP-SVM.” Finally, we also consider an identical approach that replaces LBP with the MELBP proposed in Sec. 2.1 Magnitude-Enhanced LBP, which we denote as “MELBP-SVM”.

Tables 6 and 7 report the classification performance in terms of OA and \( \kappa \) coefficient for the University of Pavia and Indian Pines datasets, respectively. The parameters for all algorithms are tuned to be optimal. To avoid any bias induced by random sampling, the classification results of all methods are averaged over 20 trials with standard deviation reported. For the University of Pavia dataset, it can be seen that the OA of the proposed approach is 11% higher than traditional SVM, about 5% higher than SVM-MRF, and 4% higher than LBP-SVM. For the Indian Pines dataset, the OA of the proposed approach is 19% higher than SVM, 7% higher than SVM-MRF and 7% higher than LBP-SVM.

Fig. 4 depicts the classification maps for the University of Pavia dataset. We see that the proposed approach tends to provide more coherent classification regions with fewer isolated pixel
misclassifications that tend to plague the SVM, SVM-MRF, and 3DDWT-SVM strategies but with fewer misclassified regions such as seen in the EMP-SVM, Gabor-SVM, and LBP-SVM maps. Fig. 5 depicts the classification maps for the Indian Pines dataset, which also shows that the proposed approach results in smoother classification maps with fewer erroneous outliers as compared to the aforementioned methods.

We also see in Tables 6 and 7 that, by comparing MELBP-SVM to LBP-SVM, the incorporation of magnitude information into the LBP operator as done in MELBP-SVM yields 1% to 2% increased classification performance as compared to using signs alone (as is done in LBP-SVM). Additionally, the proposed technique, which couples MELBP for the class-conditional probability with an MRF prior, demonstrates that the combination of the MELBP and MRF results in an additional gain of 1% to 1.5%.

| Table 6 Classification performance for the University of Pavia dataset. Results are averaged over 20 trials and reported in the form of $\eta \pm \sigma$, where $\eta$ is the mean and $\sigma$ is the standard deviation. |
|---|---|---|---|---|---|---|---|---|
| SVM | SVM-MRF | 3DDWT-SVM | EMP-SVM | Gabor-SVM | LBP-SVM | MELBP-SVM | Proposed |
| OA | 85.5±1.3 | 91.6±1.5 | 91.4±1.5 | 94.9±1.0 | 89.4±1.7 | 92.3±1.2 | 95.1±0.8 | 96.6±0.7 |
| $\kappa$ | 81.3±1.6 | 89.0±1.7 | 88.9±1.7 | 93.3±1.3 | 86.2±2.0 | 90.0±1.5 | 93.5±0.1 | 95.5±0.9 |

| Table 7 Classification performance for the Indian Pines dataset. Results are averaged over 20 trials and reported in the form of $\eta \pm \sigma$, where $\eta$ is the mean and $\sigma$ is the standard deviation. |
|---|---|---|---|---|---|---|---|
| SVM | SVM-MRF | 3DDWT-SVM | EMP-SVM | Gabor-SVM | LBP-SVM | MELBP-SVM | Proposed |
| OA | 77.8±1.1 | 89.8±1.9 | 83.6±0.7 | 83.9±0.9 | 94.5±0.8 | 95.1±0.7 | 96.3±0.8 | 97.4±0.8 |
| $\kappa$ | 75.0±1.2 | 88.5±2.0 | 81.5±0.7 | 81.6±1.0 | 93.8±0.9 | 94.4±0.8 | 95.8±1.0 | 96.9±0.9 |

Fig. 6 illustrates that the classification performance of the various algorithms decreases with decreasing number of training samples, which is as expected. However, the proposed approach always produces higher classification accuracies as compared to the other algorithms, even with an exceedingly small training-sample size (e.g., 20 training samples).

We also consider the performance of the algorithms in question under additive white Gaussian noise, which is intended to model the distortions typically occurring in the multiple stages within the HSI acquisition process. Fig. 7 presents performance results in the presence of noise, with error bars indicating the 95% confidence interval. We observe that the proposed approach outperforms all other methods over a wide range of signal-to-noise ratio (SNR).

Finally, Fig. 8 depicts classification maps and overall classification accuracies resulting for the Salinas Valley dataset for two different SNRs. It can be seen that the LBP-based approaches offer superior noise robustness in contrast to conventional SVM as well as SVM-MRF. Again, we see in Fig. 8 that MELBP-SVM outperforms LBP-SVM, while the proposed approach that couples MELBP with MRF provides further improvement of performance as well as smoother classification maps with fewer erroneous outliers.
Fig. 4 Classification maps for the University of Pavia dataset: (a) False-color map, (b) Ground truth, (c) SVM, (d) SVM-MRF, (e) 3DDWT-SVM, (f) EMP-SVM, (g) Gabor-SVM, (h) LBP-SVM, (i) MELBP-SVM, (j) Proposed.
Fig. 5 Classification maps for the Indian Pines dataset: (a) False-color map, (b) Ground truth, (c) SVM, (d) SVM-MRF, (e) 3DDWT-SVM, (f) EMP-SVM, (g) Gabor-SVM, (h) LBP-SVM, (i) MELBP-SVM, (j) Proposed.
Fig. 6 Classification performance for varying number of training samples: (a) University of Pavia, (b) Indian Pines.

Fig. 7 Classification performance for varying SNR: (a) University of Pavia, (b) Indian Pines.
Fig. 8 Classification maps for the Salinas Valley dataset.
4 Conclusion

In this paper, we proposed a MAP formulation for spatial-spectral HSI classification that couples LBP-based feature extraction for the class-conditional probability with an MRF-based prior. Instead of relegating spatial information to solely the class prior, a LBP feature extraction employing both sign and magnitude components provided a class-conditional probability that comprised both spatial and spectral information. Three datasets were used to evaluate the proposed methodology and compare it to several advanced classification strategies. Experimental results demonstrated that the proposed methodology yielded outstanding classification accuracy while being robust under small-training-sample as well as Gaussian-noise conditions.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China under Grants No. 41601344 and 61601059, and in part by the Fundamental Research Funds for the Central Universities under Grant No. 310832171006. The authors would like to thank the editor and the anonymous reviewers for their suggestions in improving the quality of this paper.

References


**Zhen Ye** received her PhD degree in Information and Communication Engineering from Northwestern Polytechnical University, Xi'an, China, in 2015. She spent one year as an exchange student of Mississippi State University. Currently, she is an instructor in the School of Electronics and Control Engineering at Chang'an University. Her research interests include hyperspectral image analysis, data compression, and machine learning.

**James E. Fowler** received the B.S. degree in computer and information science engineering and the M.S. and Ph.D. degrees in electrical engineering from The Ohio State University. He is currently a Billie J. Ball Professor and the Graduate Program Director of the Department of Electrical and Computer Engineering, Mississippi State University. He is the Editor-in-Chief of *IEEE Signal Processing Letters* and a Fellow of the IEEE.

**Lin Bai** received his PhD degree in Information and Communication Engineering from Northwestern Polytechnical University, Xi'an, China, in 2011. Currently, he is an associate professor in the School of Electronics and Control Engineering at Chang'an University. His research interests include hyperspectral image analysis, pattern recognition, and machine learning.