HYPERSPECTRAL CLASSIFICATION USING A COMPOSITE KERNEL DRIVEN BY NEAREST-NEIGHBOR SPATIAL FEATURES

Vineetha Menon†, Saurabh Prasad‡, James E. Fowler‡

†Department of Electrical and Computer Engineering
Distributed Analytics and Security Institute
Geosystems Research Institute
Mississippi State University, USA

‡Department of Electrical and Computer Engineering
University of Houston, USA

ABSTRACT

There is increasing interest in driving supervised classification of hyperspectral imagery by a support vector machine using a composite kernel employing both spectral and spatial features. While the spectral signature of the current hyperspectral pixel is often used directly to supply the spectral feature, a statistic—such as the mean—calculated across a spatial window surrounding the pixel is typically employed as a spatial feature. In contrast, a nearest-neighbor spatial feature is proposed in which the nearest neighbors in Euclidean distance to the current pixel are used to calculate the spatial feature. It is argued that the proposed nearest-neighbor spatial feature is more likely to incorporate relevant, same-class neighbor pixels than window-based features for which borders between coherent single-class regions may give rise to misclassification. Experimental results illustrate the performance advantage of the proposed nearest-neighbor framework at supervised hyperspectral classification in comparison to several competing benchmark algorithms that also employ kernel-based support vector machines.

Index Terms—nearest neighbor, composite kernel, hyperspectral classification.

1. INTRODUCTION

While the spectral signature of a hyperspectral pixel has been the focus of traditional efforts toward the classification of hyperspectral imagery, it has been recognized (e.g., [1–4]) that the incorporation of spatial context can significantly improve classification performance. Since kernel-based support vector machines (SVMs) have become ubiquitous for the supervised classification of hyperspectral data (e.g., [5, 6]), a recently popular approach to providing spatial-spectral classification is to drive an SVM with a composite kernel consisting of both spatial and spectral components [1].

In constructing such spatial-spectral composite kernels, it is necessary to extract both spectral as well as spatial features to drive the individual spectral and spatial kernels that make up the composite kernel. Usually, the spectral signature of the current hyperspectral pixel itself is simply used as the spectral feature for that pixel. On the other hand, there are a variety of spatial features that have been proposed to drive the spatial kernel (see [1] for an overview). Typical spatial features consist of some statistic (such as the mean) that is calculated over a spatially square window of $W \times W$ pixels surrounding the current pixel. The motivation for using a window about the current pixel to define a spatial feature for composite kernels rests on the assumption that spectral signatures of spatially adjacent hyperspectral pixels are highly correlated to the current pixel and thus belong to the same class. While this is likely to be true for a majority of the pixels in a hyperspectral dataset, this assumption will fail in the case of isolated pixels coming from other classes (which may arise in the case of anomalies), or, in particular, along the boundaries of otherwise coherent single-class regions.

To address this deficiency associated with window-based spatial features, we propose in this paper an alternative spatial feature based on nearest neighbors, in terms of spectral-domain Euclidean distance, to the current hyperspectral pixel. We argue that such a nearest-neighbor spatial feature is more likely to incorporate relevant (i.e., same-class) neighbor pixels than a window-based spatial feature that considers only spatial locality, particularly along region boundaries.

The remainder of our discussion is organized as follows. Next, in Sec. 2, we describe our proposed approach to a composite kernel incorporating a spatial feature driven by nearest neighbors. We follow with Sec. 3, which couples spatial preprocessing in the form of a median filter to the proposed SVM-based classifier using the nearest-neighbor composite kernel, the filtering intended to further enhance classification performance. We empirically evaluate the proposed approach in Sec. 4, comparing to several prominent SVM-based benchmarks. Finally, we make several concluding remarks in Sec. 5. We note that, although classification using a variety of nearest-neighbor paradigms has been previously considered (e.g., [7–9]), the present work constitutes the first use, to our knowledge, of a composite kernel driven by nearest-neighbor spatial features.
2. A COMPOSITE KERNEL WITH SPATIAL NEAREST NEIGHBORS

Define the spatial feature vector associated with hyperspectral pixel \( x_i \) as \( x_i^s \), while the corresponding spectral feature vector is \( x_i^ω \). The weighted-summation composite kernel [1] is

\[
K(x_i, x_j) = \mu K_s(x_i^s, x_j^s) + (1 - \mu) K_ω(x_i^ω, x_j^ω),
\]

(1)

where \( K_s(\cdot, \cdot) \) and \( K_ω(\cdot, \cdot) \) are the spatial and spectral kernels, respectively, and \( 0 < \mu < 1 \). Commonly, the hyperspectral pixel is itself used as the spectral feature vector (i.e., \( x_i^ω = x_i \)), while typically a spatial statistic (such as the mean) is calculated over a \( W \times W \) window of pixels surrounding \( x_i \) and is used as the spatial feature [1]. For example, in [10], the mean (calculated in each spectral band) over a \( 5 \times 5 \) window surrounding \( x_i \) was used as \( x_i^s \); i.e.,

\[
x_i^s = \frac{1}{|N_i|} \sum_{x_i' \in N_i} x_i',
\]

(2)

where \( N_i \) denotes the \( 5 \times 5 \) spatial neighborhood surrounding \( x_i \), and \(|N_i| = 24\).

As previously observed, the use of a square window may cause pixels from other classes to be incorporated in the calculation of the spatial feature given by (2), particularly in the occurrence of isolated pixels from other classes or at the edges of coherent single-class regions. Both effects may give rise to misclassification. As a consequence, we propose to replace the commonly used window-based formulation of the spatial feature vector with a neighborhood driven by nearest-neighbors. Specifically, (2) becomes

\[
x_i^s = \frac{1}{k} \sum_{x_i' \in N_i^k} x_i',
\]

(3)

where now the neighborhood \( N_i^k \) consists of the \( k \) nearest neighbors to \( x_i \), i.e., the \( k \) vectors \( x_i' \) from the hyperspectral image with the closest spectral-domain Euclidean distance to \( x_i \), where the spectral-domain Euclidean distance between two hyperspectral pixel vectors \( x_i \) and \( x_i' \) is \( d(x_i, x_i') = \|x_i - x_i'\|_2 \). For a hyperspectral image with \( M \) total vectors, we collect the distances \( d(\cdot, \cdot) \) between all possible pairs of pixel vectors into an \( M \times M \) adjacency matrix.\(^1\) For each pixel \( x_i \), we select the neighborhood \( N_i^k \) as the \( k \) pixels corresponding to the \( k \) columns in the adjacency matrix with the smallest entries in row \( i \).

A crucial parameter for our proposed technique is the number of nearest neighbors \( k \). If \( k \) is too large, then the spatial feature of (3) can incorporate spatially distant pixels that may not be as relevant as spatially nearer pixels, which may give rise to misclassification. On the other hand, a value of \( k \) that is too small can ignore pixels that might be highly relevant. In our experiments, we have chosen \( k = 8 \) to empirically balance the tradeoff between these two effects.

For the spatial and spectral kernels that make up the composite kernel in (1), we use a linear spatial kernel,

\[
K_s(x_i^s, x_j^s) = (x_i^s, x_j^s),
\]

(4)

and a radial-basis-function (RBF) spectral kernel,

\[
K_ω(x_i^ω, x_j^ω) = \exp \left( -\frac{\|x_i^ω - x_j^ω\|^2}{2\sigma^2} \right),
\]

(5)

where the \( \sigma \) parameter in the RBF spectral kernel is chosen using a grid search. Finally, we have found empirically that the optimal weight parameter is \( \mu \approx 0.5 \) for the kernels and features we use here (indeed, performance changes little over the range \( \mu \in [0.2, 0.7] \)); consequently, we set \( \mu = 0.5 \), meaning that (1) becomes a direct-summation composite kernel [1].

3. SPATIAL NOISE REMOVAL

Following the methodology outlined in [11, 12], we further enhance classification performance by implementing a spatial noise-removal preprocessing prior to classification with the composite-kernel driven SVM described in Sec. 2. Specifically, noise reduction is accomplished via a median filter applied with a \( 3 \times 3 \) window. A median filter is chosen due to excellent noise-reduction capability for “salt-and-pepper” and impulse noise with an added advantage of less smoothing of edges as compared to other lowpass filters. The window size is chosen empirically considering the trade off between noise reduction and misclassification due to smoothing of border pixels.

4. EXPERIMENTAL RESULTS

We now test the performance of our proposed classification approach in comparison to other SVM benchmarks. Specifically, we validate four techniques: traditional SVM with a single spectral RBF kernel (denoted RBF); composite-kernel SVM with a window-based spatial kernel (i.e., spatial features in the form of (2) with a window \( N_i \) of size \( 3 \times 3 \)) coupled with an RBF spectral kernel (denoted CK); composite-kernel SVM with a nearest-neighbor spatial kernel (i.e., spatial features in the form of (3) with \( k = 8 \)) coupled with an RBF spectral kernel (denoted kNN); and, finally, the kNN approach preceded by the spatial noise removal described in Sec. 3 (kNN+NR).

All the experiments are implemented using libSVM\(^2\). We note that our choices of spatial features for the CK and kNN approaches is fair in the sense that \(|N_i| = |N_i^k| = 8 \) for the neighborhoods \( N_i \) and \( N_i^k \) in (2) and (3), respectively.

We use the Indian Pines\(^3\) and Pavia Centre [13] datasets in our experiments. The Indian Pines dataset was acquired by

\(^1\)Although we do not consider it here, it would be possible to limit the spatial extent over which the neighborhood \( N_i^k \) was drawn to a spatial area immediately surrounding \( x_i \). Here, we let \( N_i^k \) draw from the entire image.

\(^2\)http://www.csie.ntu.edu.tw/~cjlin/libsvm

\(^3\)https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html
the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over the Indian Pines test site in northwestern Indiana. The original data consists of 224 spectral bands which were reduced to 200 bands after removal of 20 water-absorption bands. The dataset has a spatial dimension of $145 \times 145$ with a spatial resolution of 20 m; there are 16 classes of land cover. The Pavia Centre dataset was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) over an urban area of Pavia in north Italy. This dataset has 102 spectral bands each having a spatial dimension of $1096 \times 1096$ with a spatial resolution of 1.3 m; there are 9 classes of land cover. Each testing and training dataset was generated by random selection of pixels from the original hyperspectral data. Ground-truth maps for the two datasets are provided in Figs. 1 and 2.

The advantage to the nearest-neighbor formulation of spatial features as proposed in (3) is illustrated in Figs. 3 and 4 which depict the connected components of the Indian Pines and Pavia Centre datasets, respectively. Specifically, if we form a graph between hyperspectral pixels such that $N^k_i$ describes an edge between $x_i$ and each of its $k$ nearest neighbors, then we see from Figs. 3 and 4 that the connected components in such a graph exhibit sharp boundaries at the edges of coherent single-class regions. This effect implies increased classification accuracy with respect to the simple spatial windows used in, e.g., [1, 10], wherein pixels of other classes that lie within the spatial region can contribute to misclassification of the current pixel.

The efficacy of the resulting classification driven by SVM using the proposed nearest-neighbor-based composite kernel is validated in Figs. 5 and 6, wherein we note that the

---

**Fig. 1.** Ground-truth map for the Indian Pines dataset.

**Fig. 2.** Ground-truth map for the Pavia Centre dataset.

**Fig. 3.** Connected components for the Indian Pines dataset for $k = 8$.

**Fig. 4.** Connected components for the Pavia Centre dataset for $k = 8$. 
kNN+NR formulation outperforms the competing methods. This observation is particularly pronounced for the Indian Pines dataset. We note that, although the other methods perform more competitively for Pavia Centre, we still observe superior classification accuracy using the kNN+NR technique. Even with training samples as low as 5%, we are able to achieve substantially higher classification accuracy with kNN+NR.

To further validate the efficacy of our approach, we test its classification performance in the presence of additive white Gaussian noise (AWGN) with zero mean and variance $0.01$. We choose specifically AWGN in an effort to confound the median filter that is used for spatial noise removal in our approach, the median filter being most effective, rather, for impulsive “salt-and-pepper” noise. In Figs. 7 and 8, we note that our method outperforms the competing methods even in the presence of noise for both datasets considered.

5. CONCLUSIONS

In this paper, we proposed the incorporation of a spatial feature vector consisting of the mean of the $k$ nearest neighbors to a hyperspectral pixel as an alternative to the more common approach of extracting such features from a surrounding spatial window. The resulting nearest-neighbor spatial feature was used to drive a composite kernel for supervised SVM-based classification of hyperspectral imagery. The proposed nearest-neighbor spatial feature was argued to be more likely to incorporate relevant, same-class neighbor pixels than window-based features particularly along borders of coherent single-class regions, resulting in less chance of misclassification. Experimental results confirmed that the proposed nearest-neighbor-based classification system, coupled with a median-filter spatial preprocessing, empirically offered the best performance among several competing SVM-based benchmark classification algorithms.
6. REFERENCES


