ON THE SAMPLING STRATEGIES FOR EVALUATION OF JOINT SPECTRAL-SPATIAL INFORMATION BASED CLASSIFIERS

Jun Zhou¹, Jie Liang², Yuntao Qian³ and Yongsheng Gao¹

¹Institute for Integrated and Intelligent Systems, Griffith University, Nathan, Australia.  
²Research School of Engineering, Australian National University, Canberra, Australia  
³College of Computer Science, Zhejiang University, Hangzhou, P.R. China.

ABSTRACT

Joint spectral-spatial information based classification is an active topic in hyperspectral remote sensing. Current classification approaches adopt a random sampling strategy to evaluate the performance of various classification systems. Due to the limitation of benchmark data, sampling of training and testing data is performed on the same image. In this paper, we point out that while training with random sampling is practical for hyperspectral image classification, it has intrinsic problems in evaluating spectral-spatial information based classifiers. This statement is supported by several experiments, and has lead to the proposal of a new sampling strategy for comparing spectral spatial information based classifiers.

Index Terms—Hyperspectral classification, spectral-spatial analysis, feature extraction, sampling

1. INTRODUCTION

Joint spectral-spatial information based classification has attracted increasing attention during the past several years [1, 2]. By bringing spatial information into the traditional single pixel based spectral analysis, research in this area leads to better modelling of the joint spectral-spatial feature distribution of local structures in the image, and facilitates more accurate land cover classification.

Spatial information can contribute to three stages in hyperspectral image classification. Firstly, in image preprocessing, spatial neighborhood can be used for image denoise, morphology, and segmentation. Image denoise enables the reduction of random noises introduced from sensor, photon effects, and calibration errors. Several approaches have been exploited for this purpose, for example, smoothing filters, anisotropic diffusion, multi-linear algebra, wavelet shrinkage, and sparse coding methods [3]. In most cases, denoising can be done by applying a local filter with designed or learned kernel across the whole image. In mathematical morphology, operations are performed to extract spatial structures of objects according to their spectral responses [4, 5]. Similar information is explored in image segmentation, which groups spatially neighboring pixels into clusters based on their spectral distribution [6, 7].

Secondly, more common usage of joint spectral-spatial information lies in the feature extraction stage. While traditional spectral features are extracted at single pixels in hyperspectral images, spectral-spatial feature extraction methods use spatial neighborhood to calculate features. Typical examples include texture features such as 3D discrete wavelet [8], 3D Gabor wavelet [9], 3D Scattering Wavelet Transform [10], and local binary patterns [11]. Morphological profiles, alternatively, use closing, opening, and geodesic operators to enhance spatial structures of objects [12]. Other spectral-spatial features include spectral saliency [13], spherical harmonics [14], and affine invariant descriptors [15]. Heterogeneous features can be further fused using feature selection or reduction approaches [16].

Thirdly, some image classification approaches rely on spatial relation between pixels for model building. A direct way of doing so is calculating the similarity between a pixel and its surrounding pixels [17]. Markov random fields, for example, use spectral information in the local neighborhood to help pixel class prediction [7, 18]. Similar spatial structures are explored in conditional random fields [19], hypergraph modelling [2], and multi-scale analysis [20]. While supervised learning approaches, such as K-nearest neighbors, linear discriminant analysis, and support vector machines, are widely used in these classification tasks [21], some approaches adopt semi-supervised or active learning strategies [1, 22].

An assumption of the supervised learning process is that pixels with the same class shall have similar spectral responses or spectral-spatial features. Therefore, a good classifier shall be able to model the feature distribution and be generalised to predict the label of unseen samples. In order to properly evaluate classification approaches, random sampling strategies are widely adopted to alleviate the influence of noise or outliers in the data. In most hyperspectral image classification experiments, training and testing are performed on the same image due to limited availability of benchmark dataset and high cost of ground truth data collection.

In this paper, through a series of experiments, we show that such random sampling strategy for classifier evaluation has serious problems when training and testing are performed on the same image. Firstly, the dependency between spatial locations of pixels and their labels makes random sampling inappropriate. Secondly, the overlap between neighboring training and testing regions gives significant advantages to the spectral-spatial analysis approaches over pure spectral based option due to the explicit or implicit using of the testing data in the training process. Therefore, it is necessary to revisit the sampling strategy and the existing data for supervised hyperspectral image classification.

2. SPATIAL LOCATION AND LABEL DEPENDENCY

In many benchmark datasets, there are strong dependencies between the spatial location of pixels and land cover classes in hyperspectral
Fig. 1. Ground truth labels of several hyperspectral datasets: (a) Indian Pines; (b) Botswana; (c) Pavia University; (d) Salinas.

images. Figure 1 shows the ground truth labels of four commonly used hyperspectral datasets, i.e., Indian Pines, Botswana, Pavia University and Salinas. In these datasets, same class pixels appear almost all in large segments, blobs, or other continuous local structure. This is natural as pixels within the same ground objects or of the same type of land covers shall be assigned with the same class label.

Fig. 2. An example of randomly sampled training set from Indian Pines dataset.

In most cases, when performing random sampling for supervised learning, such spatial dependency is overlooked. As the training samples are drawn from the whole image, they may be selected from every connected local structures and somehow be homogeneously distributed over the whole image. When the rest of the dataset are used as testing samples, a nearby training sample from the same local structure can always be found for each testing sample. This can be exemplified by Figure 2, in which 10% of training data is sampled from the Indian Pines dataset.

In order to show how classification accuracy is influenced by spatial information, we performed comparison experiments on six benchmark datasets, i.e., Indian Pines, Botswana, Pavia University, and Salinas [23]. We employed nonlinear support vector machine (SVM) classifiers for spatial locations, i.e., 2D coordinates, and both linear and nonlinear SVMs for spectral responses of randomly sampled data, respectively. Please note that linear SVM is not used for spatial locations as they are not linearly separable. The parameters of the SVMs are learned by cross validation. Three sampling rates were explored, i.e., 5%, 10%, and 25%, to generate the training data from all labelled samples, while the rest data were used as the testing samples. The overall classification accuracies are shown in Table 1 for the spatial feature, and in Table 2 for the spectral feature, respectively.

Table 1. Classification accuracy using spatial feature only.

<table>
<thead>
<tr>
<th>Nonlinear SVM</th>
<th>%5</th>
<th>%10</th>
<th>%25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indian Pines</td>
<td>91.2</td>
<td>97.8</td>
<td>99.8</td>
</tr>
<tr>
<td>Botswana</td>
<td>83.3</td>
<td>97.0</td>
<td>100.0</td>
</tr>
<tr>
<td>PaviaU</td>
<td>88.5</td>
<td>96.5</td>
<td>99.6</td>
</tr>
<tr>
<td>Salinas</td>
<td>96.0</td>
<td>99.6</td>
<td>99.8</td>
</tr>
</tbody>
</table>

Table 2. Classification accuracy using spectral feature only.

<table>
<thead>
<tr>
<th>Nonlinear SVM</th>
<th>Linear SVM</th>
<th>Nonlinear SVM</th>
<th>Linear SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indian Pines</td>
<td>77.0</td>
<td>68.9</td>
<td>73.7</td>
</tr>
<tr>
<td>Botswana</td>
<td>86.1</td>
<td>85.5</td>
<td>92.4</td>
</tr>
<tr>
<td>PaviaU</td>
<td>93.8</td>
<td>95.8</td>
<td>87.7</td>
</tr>
<tr>
<td>Salinas</td>
<td>92.8</td>
<td>96.3</td>
<td>91.4</td>
</tr>
</tbody>
</table>

3. OVERLAP OF TRAINING AND TESTING REGIONS IN SPECTRAL-SPATIAL FEATURE EXTRACTION

More severe problem of the random sampling strategy lies in the spectral-spatial analysis, especially for the feature extraction stage. When only spectral responses is used, feature extraction is performed at single pixel, without exploring its spatial neighborhood. Therefore, there is no overlap between training and testing samples. However, extraction of spectral-spatial requires additional spectral information from neighboring pixels. This is normally implemented by a sliding window with specific sizes, for example, $3 \times 3$, $5 \times 5$, $7 \times 7$ and so on. These neighboring pixels are mostly in the testing set. Figure 3 shows the extent of overlapping between training and testing data on the Indian Pines dataset. In the figure, the white dots show the locations of training samples, and the surrounding white squares cover the region used for spectral-spatial feature extraction. When 5% training data are sampled, 31.4% testing samples are covered by the training regions. When random sampling rate increases to 25%, the extent of overlap becomes 86.3%.
Such overlap leads to using of the testing data for training purpose, and gives significant advantages to the spectral-spatial feature extraction approaches. Depending on how feature is extracted, benefit of testing data may be explicit, for example when the spectral-spatial feature is extracted by concatenating the spectral responses of pixels in a neighborhood, or implicit, for example, by extracting texture features based on time frequency analysis such as discrete wavelet transform.

Figure 4 shows the overall classification accuracy on the Indian Pines dataset when different sizes of neighborhood are used for spectral-spatial feature extraction.

To see how the continuous sampling strategy influences classification with spectral-spatial features, another experiment was developed which used 3D discrete wavelet transform (3D-DWT) for feature extraction. 3D-DWT exploits the correlation along the wavelength axis, as well as along the spatial axes, so that both spatial and spectral structures of hyperspectral imagery can be more adequately mapped into the 3D-DWT based features. In the implementation, we followed the multiple scale setting described in [8], however, without the feature selection step. The experimental results are shown in Table 4. On all datasets, continuous sampling leads to lower accuracy with spectral-spatial features. It means that with the spatial information introduced on top of the spectral information. When the size of neighborhood increases, more testing data contribute to the training step, therefore the classification accuracy increases. However, when the neighborhood becomes too large, i.e., at $9 \times 9$, unlabelled data or samples from other classes may be involved, which starts to deteriorate the constructed feature.

When testing data are used implicitly during training, their influence to the training step is more difficult to measure. In many cases, it is not clear whether better classification performance is caused by more effective spectral-spatial features, or simply due to the incorporation of testing data for training purpose. Such impact shall be evaluated case by case, and is left for future work.

### 4. AN ALTERNATIVE SAMPLING STRATEGY

Given two problems of random sampling, alternative sampling strategies shall be explored. The new strategy shall avoid drawing samples homogeneously over the whole image and overlap between training and testing regions. A direct approach is to sample continuously from a local area for each class. The randomness can be guaranteed by choosing different local areas across the data. Though this approach can not completely eliminate overlap, the influence of testing data on the training step can be greatly reduced.

We performed an experiment with raw spectral features on Indian Pines dataset to compare these two sampling strategies. Again linear SVMs were used for classification. The results in Table 3 show that compared with the random sampling strategy in Table 2, there is significant reduction on the classification accuracy when continuous sampling is used. This is because the benefit of capturing spectral distribution of testing regions has been removed. Furthermore, variations on the same class data in different regions are less sufficiently captured as some of them may not be included in the training samples. Therefore, the new sampling strategy has made the hyperspectral classification a much more difficult problem, and forces more rigorous evaluation to the feature extraction and classification approaches. Nonetheless, this is desired, as improper using of testing data is avoided.

To see how the continuous sampling strategy influences classification with spectral-spatial features, another experiment was developed which used 3D discrete wavelet transform (3D-DWT) for feature extraction. 3D-DWT exploits the correlation along the wavelength axis, as well as along the spatial axes, so that both spatial and spectral structures of hyperspectral imagery can be more adequately mapped into the 3D-DWT based features. In the implementation, we followed the multiple scale setting described in [8], however, without the feature selection step. The experimental results are shown in Table 4. On all datasets, continuous sampling leads to lower accuracy with spectral-spatial features. It means that with the spatial information provided by the random sampling, it cannot achieve significant higher accuracy than spectral feature based methods. 3D-DWT has shown much better generalization capability than pure spectral feature to cope with unseen samples, by comparing the results in Table 3 and Table 4. This is especially the case when the sampling rate increases to 25%, because the amount of testing data in the same region as the training data reduces, the classification performance using pure spectral feature drops on Pavia University and SalinasA.

### Table 3. Overall classification accuracy using spectral features with new sampling strategy.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>%5</th>
<th>%10</th>
<th>%25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indian Pines</td>
<td>55.6</td>
<td>59.8</td>
<td>63.6</td>
</tr>
<tr>
<td>Botswana</td>
<td>77.2</td>
<td>81.5</td>
<td>85.8</td>
</tr>
<tr>
<td>PaviaU</td>
<td>71.2</td>
<td>68.3</td>
<td>66.1</td>
</tr>
<tr>
<td>Salinas</td>
<td>82.0</td>
<td>88.4</td>
<td>84.4</td>
</tr>
</tbody>
</table>
datasets. However, 3D DWT feature still shows increased classification accuracy than that with the smaller sampling rates.

5. DISCUSSIONS

We have presented several experiments, highlighting the problem of random sampling strategy that is widely used in hyperspectral image classification. Though experiments were only performed to showcase the influence of spectral-spatial feature extraction, similar problem can also be caused by image processing and classification approaches that involve spatial operations. However, due to limited space in this paper, we have not include experimental analysis on these two factors.

Though a new continuous strategy has introduced, it may still be necessary to keep the random sampling as an alternative setting in the evaluating the performance of classifiers, as most published works have followed this option. Retaining it allows convenient comparison with the state-of-the-art spectral-spatial classification approaches. Therefore, a suggestion is that both sampling strategies shall be explored in the experiment development.

The problem of such sampling strategy comes from the training and testing on the same image. This is ultimately due to the lack of labelled hyperspectral data that are available to the researchers. A more urgent task for the research community, thus, is to build more benchmark datasets to facilitate spectral-spatial image analysis research.

6. CONCLUSION

We have presented an experimental study on the influence of sampling strategies for classifier performance evaluation. The results show that random sampling has some problems as it has ignored the spatial label dependency and overlap between training and testing samples. An alternative continuous sampling strategy is proposed to solve these problems. This approach provides a better way to evaluate the effectiveness of spectral spatial features and corresponding classifiers. In the future, we will expand this work to analyse the influence of random sampling on other spectral spatial operations such as preprocessing and classification models.

7. REFERENCE


