SEMI-SUPERVISED LEARNING WITH GRAPHS: COVARIANCE BASED SUPERPIXELS FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Philip Sellars[1], Angelica I. Aviles-Rivero[2], Nicolas Papadakis[3],
David Coomes[4], Anita Faul[5] and Carola-Bibiane Schönlieb[1]

[1] DAMTP and [2]DPMMS, Faculty of Mathematics, University of Cambridge, UK.
[3] IMB, Université Bordeaux France, [4] Department of Plant Science, University of Cambridge, UK.
[5] Laboratory for Scientific Computing, University of Cambridge, UK.

ABSTRACT

In this paper, we present a graph-based semi-supervised framework for hyperspectral image classification. We first introduce a novel superpixel algorithm based on the spectral covariance matrix representation of pixels to provide a better representation of our data. We then construct a superpixel graph, based on carefully considered feature vectors, before performing classification. We demonstrate, through a set of experimental results using two benchmarking datasets, that our approach outperforms three state-of-the-art classification frameworks, especially when an extremely small amount of labelled data is used.

Index Terms— Hyperspectral Imaging, Superpixels, Covariance, Graphs, Semi-Supervised Learning, Classification

1. INTRODUCTION

Hyperspectral image (HSI) classification is an active area of research and poses unique challenges. The high dimensional nature of the data allows for a detailed description of an image, and class labels can be assigned on a pixel-by-pixel basis. The majority of classification frameworks for HSIs are supervised learning (SL) frameworks; during the training phase information is only gained from the initial labelled data. These include kernel methods [1], deep-learning [2] and sparse representation methods [3]. However, the problem with SL methods is that they rely upon the existence of a large and accurately labelled training set. In application such as HSI classification, label collection is time consuming and expensive.

Another set of algorithmic approaches are based on unsupervised learning frameworks. However, the intrinsic nature of the problem makes the classification task a strongly ill-posed problem, and therefore, specific assumptions are needed to mitigate the lack of correspondence between the produced clusters and the known classes.

In practice, the size of the labelled set is often very small compared to the amount of unlabelled data. In such applications, there are large advantages to using semi-supervised learning (SSL) approaches [4]. SSL methods use information present in the labelled and unlabelled data during the training process, which can lead to much higher classification accuracy. These approaches can be divided into three groups: generative, low-density separation and graph-based methods.

This paper follows the graph perspective. This is motivated by the advantages of using graphs: i) a natural representation for HSI data, ii) a way to gain scalability and therefore computational tractability and iii) a structure with mathematical desirable properties (e.g. sparseness). However, a fundamental problem in graph based learning is *how to construct the graph in order to produce an accurate classification?* This is of great interest as the performance of the classifier heavily depends on the feature selection and the structure design, and this question is addressed in this paper.

Contributions. We present a novel graph base framework for HSI classification, which we call hyperspectral superpixel graph classification (HSGC). Our approach achieves state-of-the-art classification results using a semi-supervised graph based approach alongside a carefully designed feature space. Our highlights are: 1) spectral covariance based superpixels for feature extraction which uses local covariance matrix representation to include spatial and spectral information, 2) a graphical representation where each node represents a superpixel and the edges represent the similarity being the superpixel feature vectors and 3) a unified SSL framework. We show that learning with minimal supervision is highly beneficial when one removes feature space redundancy whilst strengthening the synergy between feature selection and graph construction.

2. LEARNING WITH MINIMAL SUPERVISION

This section is divided into three key parts. Firstly, we introduce our proposed algorithmic approach to segment HSIs. Secondly, we describe our feature extraction process. Finally, we construct the graphical representation and perform the classification task.

2.1. Spectral Covariance Based Superpixels

Let $\mathbf{I} = \{I_b\}, b = 1, .., \mathcal{B}$ be a HSI with dimensions $\mathcal{W} \times \mathcal{H} \times \mathcal{B}$ representing the width, height and number of bands

respectively and $I_b : \mathcal{W} \times \mathcal{H} \rightarrow D$ where D is the image representation of a band. Our framework starts by performing dimensionality reduction, via PCA [5], on $\mathbf{I}$ for computational efficiency. We construct a dimensionally reduced HSI $\widehat{\mathbf{I}} = \{\widehat{I_a}\}, a = 1, .., \mathcal{A}$ where $\mathcal{A} \ll \mathcal{B}$.

We then aim to find a better representation of the HSI data to increase the performance of our classifier. This is an important step as classification accuracy is highly dependent on setting relevant local regions. Local regions are commonly set by using either a fixed size or dynamic window. However, this provides a limited representation. An alternative is to use superpixels, which has been explored in [1]. Unlike [1] which relied on an existing superpixel algorithm, we propose a new superpixel approach that is designed with HSIs in mind.

Denoting an individual pixel as $p \in \widehat{\mathbf{I}}$, superpixels split the HSI into a family of disjoint sets, $\widehat{\mathbf{I}} = \cup_{i=1}^K \mathcal{S}_i$, $\mathcal{S}_i \cap \mathcal{S}_j = \emptyset$, where $\mathcal{S}_i$ corresponds to an individual superpixel and K is the number of superpixels, which is initially set by the user. Each superpixel $\mathcal{S}_i$ is made up of a set of n_i connected pixels, $\mathcal{S}_i = \{p_{i,1}, ..., p_{i,n_i}\}$. Our superpixel segmentations are produced via minimisation of the following objective, $Q(\{\mathcal{S}_1, .., \mathcal{S}_K\}) = \sum_{i=1}^K \sum_{p \in \widehat{\mathbf{I}}} d((p, \widehat{\mathbf{I}}(\mathbf{p})), F(\mathcal{S}_i))$ where d is a distance function and $F(\mathcal{S}_i)$ is the average of $\mathcal{S}_i$.

We have built our superpixel algorithm on top of the commonly used SLIC algorithm [6]. SLIC has drawbacks [7] found in the fixed size localised search range. To improve upon this, we adopt the observation of [7], in which the search range is dynamically adjusted by the local content density in the image. This information is given by the function g which maps each pixel to a positive real number. Therefore, in our superpixel algorithm, we used the following search range

$$d((p, \widehat{I}(p)), F(\mathcal{S}_i)) \text{ if } |p - (F(\mathcal{S}_i))_1| \leq 2\sqrt{\tfrac{\mathcal{W}\mathcal{H}}{K}} g(F(S_i));$$
$$\text{Otherwise } d((p, \widehat{I}(p)), F(\mathcal{S}_i)) = \infty;$$

where $(F(\mathcal{S}_i))_1$ represents that spatial part of the feature function, and $| \cdot |$ is the euclidean distance on the image grid. Furthermore, instead of using the Euclidean spectral distance found in [6, 7], we instead use covariance matrix representation [8] and the Log-Euclidean distance (LED) [9] which is better suited for HSIs. For each pixel $p \in \widehat{\mathbf{I}}$ we construct a covariance matrix $\mathbf{C}_p$ describing the relationship between different hyperspectral bands, which extracts powerful spectral and spatial information. However, covariance matrices are symmetric positive definite matrices and they do not lie on a Euclidean space but instead on a Riemannian manifold. Therefore, the LED metric is used to construct the spectral distance between pixels,

$$d_{spectral}(p_x, p_y) = ||\mathrm{logm}(\mathbf{C}_{p_x}) - \mathrm{logm}(\mathbf{C}_{p_y})||_F, \quad (1)$$

In our superpixel construction, the reduced image $\widehat{\mathbf{I}}$ is passed into the covariance based superpixel algorithm and a 2-D superpixel label map is generated. This map is applied back to $\widehat{\mathbf{I}}$ to obtain our 3-D superpixel mapping.

2.2. Feature Extraction

From each superpixel $\mathcal{S}_i$ we extract three different features. By applying a mean filter to each superpixel we can extract localised spatial information. The mean feature vector is denoted as $\vec{\mathcal{S}}_i^m$ and it is defined as

$$\vec{\mathcal{S}}_i^m = \frac{\sum_{j=1}^{n_i} \widehat{\mathbf{I}}(p_{i,j})}{n_i}. \quad (2)$$

To obtain the spatial information surrounding a superpixel, we take a weighted combination of the information present in the adjacent superpixels. For each given superpixel $\mathcal{S}_i$, we define the set $\mathcal{Z}_i = \{z_1, z_2.., z_J\}$ which contains the J indexes of the adjacent superpixels. The weighted feature vector $\vec{\mathcal{S}}_i^w$ is given by

$$\vec{\mathcal{S}}_i^w = \sum_{j=1}^J w_{i,z_j} \vec{\mathcal{S}}_{z_j}^m, \quad (3)$$

where h is a predefined scalar parameter and the weight between adjacent superpixels w_{i,z_j} is defined as

$$w_{i,z_j} = \frac{\exp\left(-||\vec{\mathcal{S}}_{z_j}^m - \vec{\mathcal{S}}_i^m||_2^2/h\right)}{\sum_{j=1}^J \exp\left(-||\vec{\mathcal{S}}_{z_j}^m - \vec{\mathcal{S}}_i^m||_2^2/h\right)} \quad (4)$$

Finally, we extract the location of the centre of each superpixel $\vec{\mathcal{S}}_i^p$ which we calculate as

$$\vec{\mathcal{S}}_i^p = \frac{\sum_{j=1}^{n_i} p_{i,j}}{n_i}. \quad (5)$$

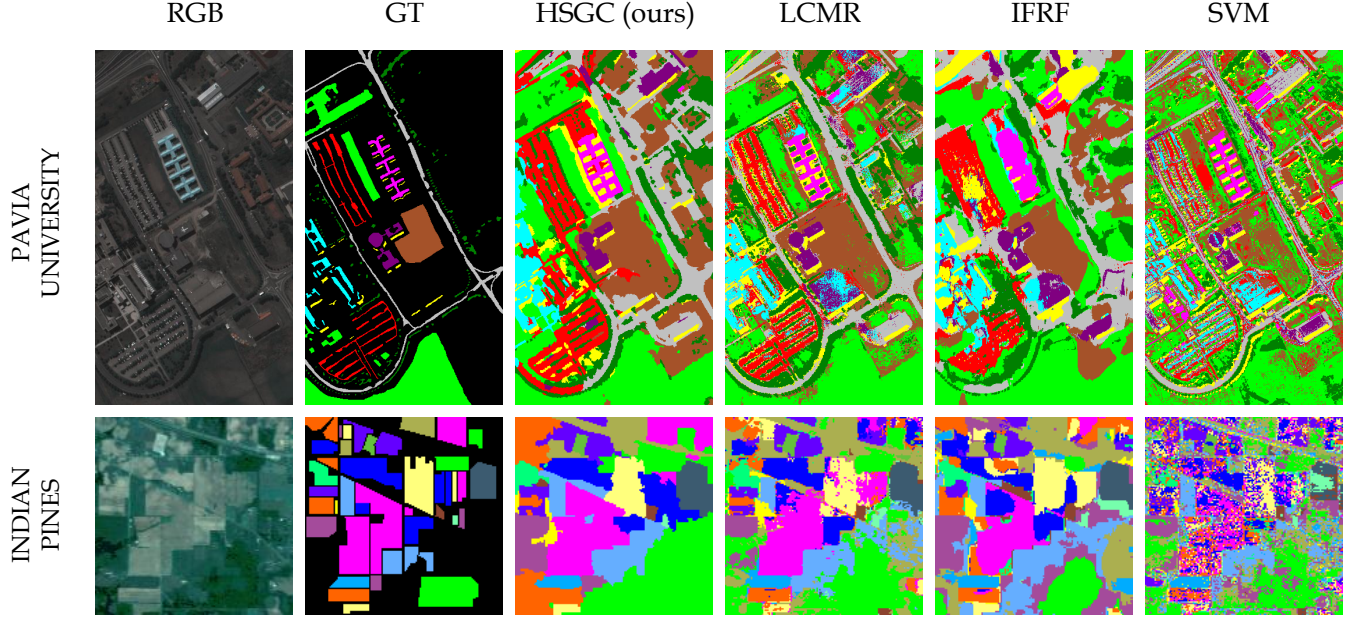## 2.3. Graph Construction and SSL Classification

Using these feature vectors we construct a weighted, undirected graphical representation $G = (V, E, W)$ where each node is a superpixel and the edge weights reflect the similarity between superpixels. Note that a similarity of 1 implies most similar and an decreasing number means less similar. The weight between superpixels is given by $w_{ij} = s_{ij}l_{ij}$,

$$s_{ij} = \exp\left(\frac{(\beta - 1)||\vec{\mathcal{S}}_i^w - \vec{\mathcal{S}}_j^w||_2^2 - \beta||\vec{\mathcal{S}}_i^m - \vec{\mathcal{S}}_j^m||_2^2}{\sigma_s^2}\right), \quad (6)$$

$$l_{ij} = \exp\left(\frac{-||\vec{\mathcal{S}}_i^p - \vec{\mathcal{S}}_j^p||_2^2}{\sigma_l^2}\right). \quad (7)$$

The parameter β weights the contribution of the mean and weighted feature vectors and σ_s, σ_l determine the width of the Gaussian kernels. The edge set is constructed using k-nearest neighbours.

The initial labelling of the superpixels is specified using the matrix $Y \in \mathbb{R}^{K \times c}$, where c is the number of classes present and K is the number of superpixels. Y_{vl} specifies the value of the seed label l for node v. The initial label information

Fig. 1. **The classification maps.** A comparison of the classification maps produced for the two data sets. From left to right: the RGB image, the ground truth (GT), and classification maps from: HSGC (our approach), LCMR [10], IFRF [11] and SVM [12] methods when ten labelled pixels for each class are used for training. Note that HSGC and LCMR are SSL based methods.

for each superpixel is taken as the average initial label of its set of pixels. If no pixel within a superpixel is labelled the superpixel has no initial label. The graph G and the labelling matrix Y are then fed into the *Learning with Local and Global Consistency algorithm* (LGC) by Zhou et al [13]. LGC propagates the labelling information across the graph and produces the final superpixel labelling matrix $F \in \mathbb{R}^{K \times c}$. The final label for each superpixel y_i is given by,

$$y_i = \underset{j \in \{1,..,c\}}{\mathrm{argmax}} F_{ij}. \tag{8}$$

Each superpixel label is then passed down to its corresponding set of pixels.

3. EXPERIMENTAL RESULTS

This section addresses the experimental methodology that we used to validate and assess our proposed approach.

Data Description. We use two benchmark datasets. *University of Pavia:* with dimensions of $610 \times 340 \times 103$, a spectral range from 0.43 to 0.86μm and a spatial resolution of 1.3m. *Indian Pines:* dimensions of $145 \times 145 \times 200$, a spectral range of 0.4 to 2.5μm and a spatial resolution of 20m.
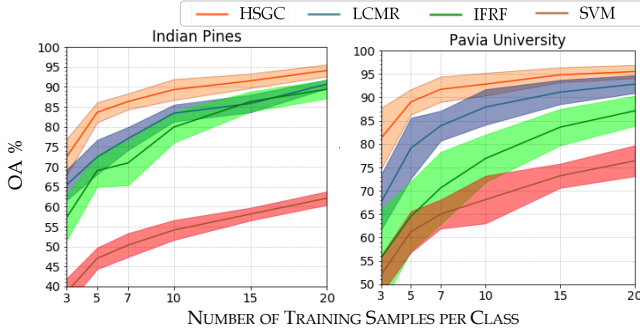
Experimental Setup. We compared our proposed approach against three state-of-the-art HSI classification methods: local covariance matrix representation (LCMR) [10], a SVM method [12] and image fusion and recursive filtering (IFRF) [11]. The first method is semi-supervised and the last two

Table 1. OA (%) AA (%) and Kappa (%) with ten training samples per class, and for the two HSI datasets.

	HSGC (ours)	LCMR [10]	IFRF [11]	SVM [12]
UNIVERSITY OF PAVIA DATASET				
OA	$91.5 \pm 2.6\%$	$87.9 \pm 3.8\%$	$79.1 \pm 3.8\%$	$69.7 \pm 3.1\%$
AA	$90.2 \pm 7.8\%$	$90.4 \pm 2.6\%$	$73.2 \pm 3.1\%$	$77.8 \pm 1.3\%$
Kappa	$88.9 \pm 3.3\%$	$84.3 \pm 4.7\%$	$73.0 \pm 4.6\%$	$61.9 \pm 3.3\%$
INDIAN PINES DATASET				
OA	$89.8 \pm 1.7\%$	$83.4 \pm 2.1\%$	$80.1 \pm 3.2\%$	$54.1 \pm 2.7\%$
AA	$93.8 \pm 7.7\%$	$90.6 \pm 1.5\%$	$74.8 \pm 2.8\%$	$67.2 \pm 1.6\%$
Kappa	$88.4 \pm 2.0\%$	$81.2 \pm 2.3\%$	$77.5 \pm 3.5\%$	$48.7 \pm 2.7\%$

are supervised methods. The parameters of the compared approaches were set to the default values referenced in the papers. For our method the parameters were determined by an empirical coarse to fine search method. The spectral dimension after PCA was set by demanding that the total explained variance ratio was ≥ 0.98. The number of superpixels chosen was 1000 for Indian Pines and 2000 for the University of Pavia. To demonstrate the performance of our method in *minimal supervision* settings, we only use very small training sets ranging from three to twenty labelled pixels per class. All experiments are repeated ten times and use three common metrics to evaluate our performance: overall accuracy (OA), average accuracy (AA) and the Kappa coefficient.

Comparison With Other Methods. we start by visually evaluating our approach against the three different classifiers. The results are visualised in Fig.1. We can observe that our

Fig. 2. **Overall Classification Accuracy (OA).** The lines represent the average data for ten runs whilst the shaded regions represent the standard deviation of: LCMR [10], IFRF [11], SVM [12] and the proposed HSGC.

Table 2. OA (%) on the two benchmarking datasets.

	UNIVERSITY OF PAVIA DATASET		
Labels\ Class	**HSGC** [ours]	**LCMR** [10]	**IFRF** [11]	**SVM** [12]
3	$81.2 \pm 6.5\%$	$67.5 \pm 5.9\%$	$55.6 \pm 10.3\%$	$52.0 \pm 4.0\%$
7	$91.7 \pm 2.7\%$	$83.9 \pm 3.2\%$	$70.6 \pm 7.7\%$	$65.0 \pm 3.1\%$
20	$95.5 \pm 1.4\%$	$92.8 \pm 1.9\%$	$87.1 \pm 3.2\%$	$76.4 \pm 3.3\%$
	INDIAN PINES DATASET			
3	$72.4 \pm 4.5\%$	$65.4 \pm 3.8\%$	$57.2 \pm 5.7\%$	$38.5 \pm 3.5\%$
7	$84.3 \pm 2.0\%$	$77.0 \pm 2.9\%$	$70.9 \pm 5.6\%$	$50.3 \pm 3.0\%$
20	$94.1 \pm 1.5\%$	$90.7 \pm 1.1\%$	$89.4 \pm 2.3\%$	$62.1 \pm 1.7\%$

method produces a smoother output, with significantly fewer outliers than the compared approaches. It deals well with the complex structures present in the Pavia data set and preserves the boundaries between the different regions in Indian Pines. This is reflected in the numerical results reported in Table 1, where the OA, the AA and Kappa coefficient were calculated using *ten* labelled pixels per class. We observe that our approach outperformed the other methods for both datasets. However, we can observe in Fig. 2 and Table 2 that our approach *significantly* outperforms the compared methods when the number of labelled pixels per class is reduced below 10. We can observe, that for all number of labels counts and for both datasets, our approach exhibits the highest overall classification accuracy (OA). Note that higher values of OA correspond to better classifier performance. Overall, our approach outperformed the compared state-of-the-art approaches. The main contribution of our approach is to produce very high classification accuracy even when the number of labelled pixels per class is incredibly small.

4. CONCLUSION

In this work, we present a novel framework, HSGC, for hyperspectral image classification. Our framework combines a novel, purpose built superpixel algorithm with a semi-supervised graph based approach. We demonstrate state-of-the-art results compared with a recent semi-supervised

and two supervised approaches. Our highlight is that HSGC produces the highest classification accuracy, even when the amount of labelled data is small. This shows the benefits and potential of learning with minimal supervision on graphs.

5. REFERENCES

[1] L Fang, S Li, W Duan, J Ren, and J A Benediktsson, "Classification of hyperspectral images by exploiting spectralspatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, 2015.

[2] D Ienco, R Gaetano, C Dupaquier, and P Maurel, "Land cover classification via multitemporal spatial data by deep recurrent neural networks," *IEEE Geoscience and Remote Sensing Letters*, pp. 1685–1689, 2017.

[3] Y Chen, N M Nasrabadi, and T D Tran, "Semi-supervised graphbased hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, 2011.

[4] O Chapelle, A Zien, and B Schölkopf, *Semisupervised learning*, MIT Press, 2006.

[5] I Jolliffe, "Principal component analysis," *New York, NY, USA: Wiley*, 2005.

[6] R Achanta and et al, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 2274–2282, 2012.

[7] G. Maierhofer, D. Heydecker, A. I. Aviles-Rivero, S. M. Alsaleh, and C. Schönlieb, "Peekaboo-where are the objects? structure adjusting superpixels," in *IEEE International Conference on Image Processing (ICIP)*, 2018.

[8] O Tuzel, F Porikli, and P Meer, "Region covariance: A fast descriptor for detection and classification," *Proc. Eur. Conf. Comput. Vis.*, pp. 589–600, 2006.

[9] V Arsigny, P Fillard, X Pennec, and N Ayache, "Geometric means in a novel vector space structure on symmetric positive-definite matrice," *SIAM J. Matrix Anal. Appl.*, vol. 29, no. 1, pp. 328–347, 2006.

[10] L Fang, N He, S Li, and J Plaza, "A new spatial-spectral feature extraction method for hyperspectral images using local covariance matrix representation," *IEEE Trans. Geosci. Remote Sens.*, pp. 3534–3546, 2018.

[11] X Kang, S Li, and J A Benediktsson, "Feature extraction of hyperspectral images with image fusion and recursive filtering," *IEEE Trans. Geosci. Remote Sens.*, 2014.

[12] F Melgani and L Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, 2004.

[13] D Zhou, O Bousquet, T Lal, J Weston, and B Schölkopf, "Learning with local and global consistency," *NIPS*, pp. 595–602, 2004.