EUSIPCO 2013 1569744031

BAND SELECTION FOR HYPERSPECTRAL IMAGES BASED ON SELF-TUNING SPECTRAL CLUSTERING

Vineet Kumar *Jürgen Hahn* *Abdelhak M. Zoubir*

Signal Processing Group
Technische Universität Darmstadt
Merckstraße 25, 64283 Darmstadt, Germany
E-mail: {vkumar, jhahn, zoubir}@spg.tu-darmstadt.de

ABSTRACT

Hyperspectral imaging (HSI) is an emerging technique, which allows to consistently capture images in the visible as well as infrared light range. Many materials can be easily discriminated by means of their spectra, rendering HSI an interesting method for the reliable classification of contents in a scene. As the number of features for each pixel in hyperspectral images is considerably high, further processing and classification is time consuming and stresses resources. Thus, efficient methods to select useful bands are required. We present a novel two-step scheme based on a clustering approach followed by representatives selection from each cluster. The classification results of real hyperspectral images demonstrate that the proposed method easily outperforms common as well as state-of-the-art methods.

Index Terms— Hyperspectral imaging, band selection, classification

1. INTRODUCTION

Hyperspectral imaging (HSI), frequently described as imaging spectroscopy, is the fusion of spectroscopy and image processing. Thus, hyperspectral images contain the reflectance of the visible and infrared light decomposed in tens or hundreds of bands.

Objects of different materials absorb and reflect light at different frequencies due to their molecular structure [1], resulting in discriminative spectra or signatures. Therefore, identification of a great diversity of materials becomes fairly easy, rendering HSI interesting for a wide area of engineering tasks and research. Today, HSI is often used in mineralogy, agriculture and surveillance [2].

As stated by the well-known curse of dimensionality, the high number of features or bands of hyperspectral images can be problematic for classification. The problem consists in the fact that the feature space becomes highly sparse as the number of features increases. This makes finding suitable class representations difficult and thus, often leads to low classification accuracies. Additionally, due to the high amount of data, processing and evaluating hyperspectral images is time consuming and stresses computational resources. Therefore, a key challenge is to reduce the number of features leading to feature or band selection.

For this purpose, several methods have been proposed to date, such as principal component analysis (PCA), information entropy (IE), or contrast measure (CM) [3]. Chang et al. proposed modifications of PCA, namely maximum-variance PCA (MVPCA) and maximum-SNR PCA (MSNRPCA) [4]. These methods have basically two disadvantages: first, they heavily depend on suitable parameter selection and second, data transform may lead to information loss and thus low classification accuracy [5].

In [5], a different approach, constrained band selection (CBS), is explained. Additionally, linearly constrained minimum variance (LCMV) based on constrained energy minimization (CEM) and virtual dimensionality (VD) [6] is proposed. The key idea is to linearly constrain each band of a signal of interest, while minimizing the band correlation or band dependence from other bands. In [7] and [8], band selection is performed by clustering, resulting in many cases in higher classification accuracies. Martinez-Uso et al. [7] cluster the bands by similarity measures such as the Kullback-Leibler divergence, reducing the computational complexity, while Qian et al. [8] utilize affinity propagation (AP). Additionally, Qian et al. use wavelet shrinkage to remove noisy bands.

In this contribution we propose a novel clustering based band-selection algorithm that groups similar bands of the hyperspectral image into the same cluster. Representatives are chosen from each cluster that are finally considered for classification. This method provides high classification accuracy while the parameters are either automatically estimated or can be set to fixed values that show good performances with the tested datasets.

This paper is structured as follows. In Section 2, we briefly review the spectral clustering approach of [9], as it forms an important step of our approach. In Section 3, we explain the framework of the proposed band selection algorithm. Results

of three dataset are shown in Section 4 and are compared with common and state-of-the-art band selection algorithms. Finally, a short conclusion is drawn in Section 5.

2. SELF TUNING SPECTRAL CLUSTERING

Many clustering algorithms such as k-means or EM clustering pose strong assumptions on the distribution of the data or the features, which in practice do not hold and therefore lead to poor results. A promising alternative is the graph cut algorithm that has shown accurate segmentation results in image processing [10]. It is based on finding the optimal flow in a graph representation of the data, where the links between the nodes represent their affinities. However, this method suffers from the problem that the sizes of the clusters can be heavily unbalanced. An attempt to solve this problem consists in normalized cuts [11] that results however in an NP-hard problem. A relaxation of this problem is given by spectral clustering [12].

In the sequel, we briefly describe spectral clustering [9, 12]. Let $\mathbf{f}_m \in \mathbb{R}^N$ with $1 \leq m \leq M$ denote an N-dimensional feature or data vector that is to be clustered. The normalized graph Laplacian $\mathbf{L}$ [12] is then defined as

$$\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2} \tag{1}$$

where $\mathbf{A}$ is a $M \times M$ matrix that describes the affinities between all features. The affinity between the i-th and j-th feature is defined as $a_{ij} = \exp(-\|\mathbf{f}_i - \mathbf{f}_j\|/2\sigma^2)$, with σ denoting a scaling factor. Note that the self-affinity a_{ii} is set to zero. The diagonal matrix $\mathbf{D}$ is computed by the sums over the rows of $\mathbf{A}$. The k eigenvectors corresponding to the k largest eigenvalues of $\mathbf{L}$ are chosen and stacked into $\tilde{\mathbf{L}}$. Finally, the transformed data points $\tilde{\mathbf{L}}$ are clustered by a conventional clustering method such as k-means.

The values of the kernel width or the scaling σ and the number of final clusters K have a strong impact on the result. Perona and Zelnik-Manor [9] present an efficient method to estimate local scales and a suitable number of clusters. A local scale $\sigma_m, m = 1, \ldots, M$ can be considered as the distance between f_m and its k-th nearest neighbor where k is determined by the dimension of the feature vector [9]. Thus, the affinity can be reformulated as $\tilde{a}_{ij} = \exp(-\|\mathbf{f}_i - \mathbf{f}_j\|)/(\sigma_i\sigma_j)$. In order to estimate the number of clusters K, Perona and Zelnik-Manor propose minimizing the distortion caused by clustering. For details, we refer the reader to [13, 14, 9].

3. BAND SELECTION BASED ON SPECTRAL CLUSTERING

We propose a band selection scheme which can be divided into two parts: (i) cluster formation and (ii) representative selection. In the first step, the algorithm groups similar bands into a cluster. In the second step, most informative features are extracted from each cluster. Afterwards, the image can be further processed or classified using the extracted features. An overview of this procedure is depicted in Fig. 1.

3.1. Cluster formation

Let $\mathbf{X}$ denote an $N_1 \times N_2 \times M$ hyperspectral image where N_1 and N_2 are the spatial dimensions and M the number of bands. We consider each vectorized band $\mathbf{b}_m \in \mathbb{R}^N$ with $1 \leq m \leq M$ and $N = N_1 \cdot N_2$ as a feature vector. The bands are clustered by spectral clustering [12] as explained in Section 2. We suggest spectral clustering by Perona et al. [9] as this method is able to automatically estimate its parameters. Note that basically any clustering method can be used here.

After clustering, the bands are grouped into K clusters $S^{(k)}, k = 1, \ldots, K$. In the subsequent step, representations of each cluster are selected.

3.2. Representative selection

For each of the K clusters, lower dimensional representations have to be found, i.e. a mapping $d : \mathbb{R}^{M^{(k)}} \rightarrow \mathbb{R}^{\tilde{M}^{(k)}}$ where $M^{(k)}$ is the number of bands and $\tilde{M}^{(k)} < M^{(k)}$ is the number of reduced bands in the respective cluster. We have tested many different mappings, e.g. contrast measure, pooling, virtual dimensionality, and information entropy [3]. In general, PCA has shown the best performance.

Thus, PCA is applied to each cluster to obtain the representations, i.e. we estimate the covariance matrix $\mathbf{C}^{(k)} \in \mathbb{R}^{M^{(k)} \times M^{(k)}}$ of the bands contained in each cluster $S^{(k)}$, compute its eigenvalues $\lambda_1^{(k)}, \lambda_2^{(k)}, \ldots \lambda_{M^{(k)}}^{(k)}$, and the corresponding eigenvectors $\mathbf{v}_1^{(k)}, \mathbf{v}_2^{(k)}, \ldots \mathbf{v}_{M^{(k)}}^{(k)}$. The eigenbasis of each cluster is then given by

$$\mathbf{V}^{(k)} = \left[\mathbf{v}_1^{(k)}, \ldots \mathbf{v}_{M^{(k)}}^{(k)}\right]. \tag{2}$$

Ideally, each cluster would be represented by the eigenvector exhibiting the largest eigenvalue. In some cases, the estimated number of clusters $\hat{K}$ is too low meaning that the dimension of the feature space is dramatically reduced. Thus, the discrimination of the classes becomes difficult and the classification accuracy is decreased.

For estimating how many eigenvectors should be kept in each cluster, we define the content of a cluster $C^{(k)}$ as the sum over all eigenvalues,

$$C^{(k)} = \sum_{i=1}^{M^{(k)}} \lambda_i^{(k)} \tag{3}$$

which reflects the variation of data in this cluster [3]. In order to retain content, we suggest to determine the number of kept eigenvalues $\tilde{M}^{(K)}$ for each cluster by finding a minimal $\tilde{M}^{(K)}$ that fulfills

$$\sum_{i=1}^{\tilde{M}^{(K)}} \lambda_i^{(k)} \geq \alpha C^{(k)} \tag{4}$$

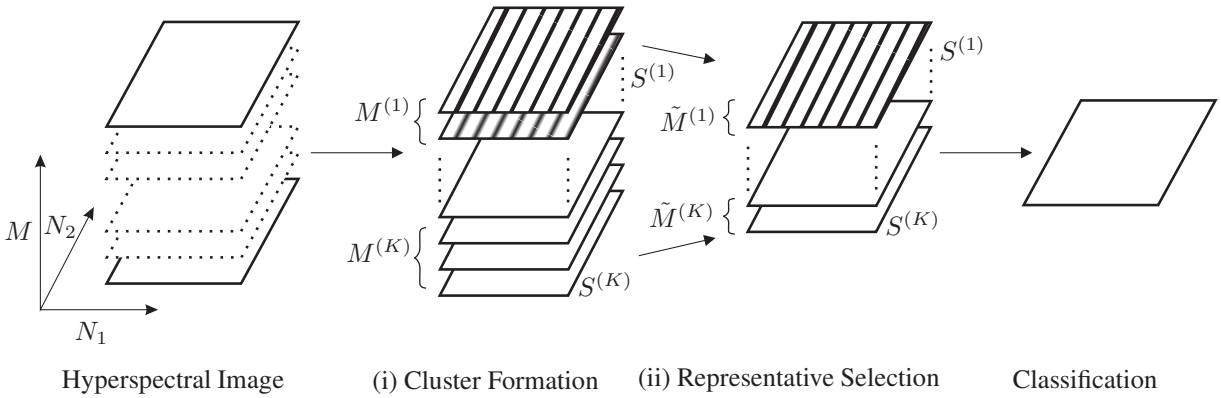

Fig. 1: Overview of the proposed band selection approach

with $0 \leq \alpha \leq 1$. By stacking the eigenvectors, we finally obtain the truncated eigenbasis

$$\tilde{\mathbf{V}}^{(k)} = \left[\mathbf{v}_1^{(k)}, \dots \mathbf{v}_{\tilde{M}^{(k)}}^{(k)}\right]. \tag{5}$$

Note that the presented scheme allows to find the reduced number of bands $\tilde{M} = \sum_1^K \tilde{M}^{(k)}$ automatically based on α.

3.3. Classification

During training, the clusters $S^{(k)}, k = 1, \dots, K$ and the basis transforms $\tilde{\mathbf{V}}^{(k)}$ are learned. Using $\tilde{\mathbf{V}}^{(k)}$, the bands are projected into a lower dimensional subspace,

$$\tilde{\mathbf{x}}_{i,j} = \left[\tilde{\mathbf{V}}^{(1)} \quad \dots \quad \tilde{\mathbf{V}}^{(K)}\right]^T \mathbf{x}_{i,j} \tag{6}$$

where $\mathbf{x}_{i,j} \in \mathbb{R}^M$ denotes the (i,j)-th signature of the image with $1 \leq i \leq N_1, 1 \leq j \leq N_2$ and $\tilde{\mathbf{x}}_{i,j}^{(k)}$ its projection. Finally, we obtain the band-reduced image $\tilde{\mathbf{X}}$ by stacking the coefficient vectors $\tilde{\mathbf{x}}_{i,j}$. As in pixel-wise classification, the coefficient vectors can be classified directly. Alternatively, spatial filtering or spatial feature extraction can easily be applied to improve the classification performance.

4. RESULTS

The proposed method (SCP) has been evaluated in terms of classification accuracy. For this purpose, we evaluate the overall accuracy (OA), which is the fraction of all correctly classified samples versus all samples and the class accuracy (CA) reflecting the number of correctly classified samples of each class. In the following, we present results for the Indian Pines (IP) [15], Center of Pavia (CP) [16] as well as University of Pavia (UP) [16] datasets. The Indian Pines image has a size of 145×145 pixels showing 16 classes of mainly vegetation. After removing water absorption bands [15], the image contains 200 bands. In contrast, the University of Pavia and Center of Pavia images were captured in urban area, comprise 9 classes, and have a size of 610×340 and 1096×490 pixels, respectively. After removing the noisy bands, the images contain 103 and 102 bands, respectively.
For classification, we consider the k-nearest neighbor (KNN) [17] classifier and support vector machines (SVM) [18]. The experiments have been repeated ten times and the training samples have been randomly chosen. For IP, we consider in total 660 training samples, for CP 5536 and UP 3921. During model selection, the parameters of the classifiers have been estimated, i.e. the number of neighbors K_N for KNN and the regularization C and the kernel bandwidth γ for SVM.

4.1. Influence of the number of bands

First, we investigate how the proposed method (SCP) is affected if only few bands are considered. In order to compare this method with recent methods, we also provide results for affinity propagation (AP) [8], linearly constrained minimum variance with band dependence minimization (LCMV) [5], MVPCA [4], and all bands (All) using the same setup. As explained in Section 3, the proposed approach is designed to find an appropriate number of bands automatically. However, this value can still be set manually which allows a fair comparison with other methods. Thus, we manually set the number of reduced bands between 5 and 50 with a step size of 5.
For the Indian Pines image, using KNN, the proposed method performs similar as AP leading to significantly higher accuracies than LCMV or MVPCA (Fig. 2(a)). However, considering all bands yields the best result. This holds also for the SVM (Fig. 2(b)). The major difference consists in the fact that SCP outperforms even AP and is very close to the results of considering all bands.
Since we have a lot more training samples for the University of Pavia dataset, we generally obtain higher classification accuracies. For KNN, LCMV and MVPCA show the best results as depicted in Fig. 2(c). Still, SCP gives more accurate predictions than AP. Using SVMs, SCP yields the best performance in all cases (Fig. 2(d)). However, especially LCMV and MVPCA are prone to small numbers of features.

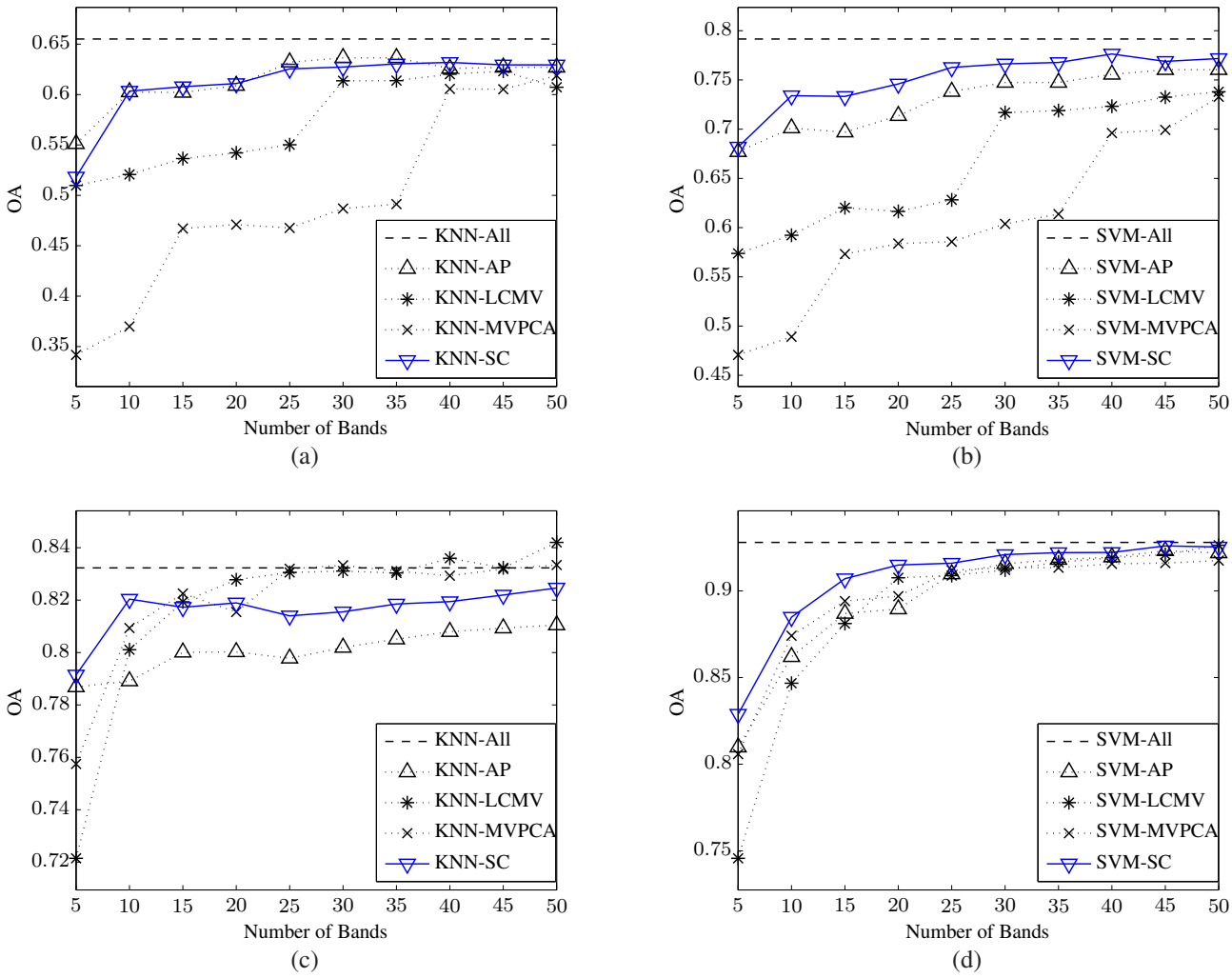

Fig. 2: Classification results for Indian Pines, (a) KNN and (b) SVM, and University of Pavia, (c) KNN and (d) SVM

We obtain similar results for the Center of Pavia image except that even for low numbers of features high accuracies are achieved. Due to space limitations the figures are not shown.

4.2. Automatic number-of-bands selection

For estimating a suitable number of bands, the content parameter α has to be set. By cross-validation, we found $\alpha = 99.99\%$ yielding high accuracies (see Tab. 1). Note that despite the high value of α, the number of bands is often significantly reduced. For IP, we obtain 182 bands, for CP 54 and UP 50, i.e. a reduction of roughly 20-50% of the original data is achieved. In Tab. 1, we also provide results for the case that morphological operations (opening and closing), denoted by SCP+M and All+M, have been applied after the band reduction. The high accuracies show that despite reducing the number of bands, high classification accuracies can be maintained.

5. CONCLUSION

In this contribution, we presented a novel, effective band selection scheme for hyperspectral image evaluation. Our approach is based on self tuning spectral clustering to group similar bands into clusters. Then, from each cluster representatives are chosen. For this purpose, we suggest the use of PCA. Finally, the band-reduced image can further be processed or classified. We have shown based on three real datasets that our scheme outperforms common as well as state-of-the-art methods like affinity propagation and performs especially well for very low number of bands. In many cases, even considering all bands does not result in higher accuracies. Further processing of the band-reduced image, such as morphological operations, can help to improve classification accuracy. Future directions will include exploitation of spatial information within the bands.

Table 1: Classification results of the datasets with automatically estimated number of bands

	Indian Pines				University of Pavia				Center of Pavia			
	SVM		KNN		SVM		KNN		SVM		KNN	
	OA	CA	OA	CA	OA	CA	OA	CA	OA	CA	OA	CA
All	79.16	85.85	**65.51**	**75.13**	**92.80**	**93.20**	83.22	**86.47**	**99.09**	**97.95**	**98.05**	**95.78**
SCP	**79.35**	**85.95**	65.50	**75.13**	92.71	93.13	**83.23**	86.35	99.08	97.92	**98.05**	95.77
All+M	91.31	95.57	88.34	**93.91**	99.20	**99.39**	**98.01**	**98.88**	**99.93**	**99.85**	**99.87**	99.69
SCP+M	**92.49**	**96.04**	**88.95**	93.85	**99.63**	99.68	97.70	98.60	99.92	99.84	99.84	**99.70**

6. ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe from Purdue University, U.S., for kindly providing the Indiana's Indian Pines data and Prof. P. Gamba from University of Pavia, Italy, for providing the University of Pavia and Center of Pavia images.

7. REFERENCES

[1] G.A. Shaw and H. K. Burke, "Spectral imaging for remote sensing," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 3–28, 2003.

[2] R.A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, Elsevier, 3rd edition, 2007.

[3] P. Bajcsy and P. Groves, "Methodology for hyperspectral band selectionâ, photogrammetric engineering and remote," *Sensing Journal*, vol. 70, pp. 793–802, 2004.

[4] C.-I Chang, Q. Du, T.-L. Sun, and M.L.G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2631 –2641, 1999.

[5] C.-I Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1575–1585, 2006.

[6] C.-I Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608–619, 2004.

[7] A. Martinez-Uso, F. Pla, J.M. Sotoca, and P. Garcia-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4158 –4171, 2007.

[8] Y. Qian, F. Yao, and S. Jia, "Band selection for hyperspectral imagery using affinity propagation," *IET Computer Vision*, vol. 3, no. 4, pp. 213 –222, 2009.

[9] P. Perona and L. Zelnik-Manor, "Self-tuning spectral clustering," *Advances in neural information processing systems*, vol. 17, pp. 1601–1608, 2004.

[10] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222 –1239, 2001.

[11] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.

[12] U. Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[13] G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 1996.

[14] V. K. Goyal and M. Vetterli, "Block transform adaptation by stochastic gradient descent," in *IEEE Dig. Sig. Proc. Workshop*, 1998.

[15] S. Tadjudin and D.A. Landgrebe, "Covariance estimation with limited training samples," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 4, pp. 2113–2118, 1999.

[16] P. Gamba, "A collection of data for urban area characterization," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp. IGARSS '04*, 2004, vol. 1.

[17] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 13, no. 1, pp. 21 –27, 1967.

[18] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, Software available at *http://www.csie.ntu.edu.tw/∼cjlin/libsvm*.