HYPERSPECTRAL IMAGE INTERPRETATION BASED ON PARTIAL LEAST SQUARES

Andrey Bicalho Santos[1], *Arnaldo de Albuquerque Araújo*[1], *William Robson Schwartz*[1], *David Menotti*[2]

[1]Computer Science Department, Federal University of Minas Gerais, Belo Horizonte, Brazil
[2]Computing Department, Federal University of Ouro Preto, Ouro Preto, Brazil

ABSTRACT

Remote sensed hyperspectral images have been used for many purposes and have become one of the most important tools in remote sensing. Due to the large amount of available bands, e.g., a few hundreds, the feature extraction step plays an important role for hyperspectral images interpretation. In this paper, we extend a well-know feature extraction method called Extended Morphological Profile (EMP) which encodes spatial and spectral information by using Partial Least Squares (PLS) to emphasize the importance of the more discriminative features. PLS is employed twice in our proposal, i.e., to the EMP features and to the raw spatial information, which are then concatenated to be further interpreted by the SVM classifier. Our experiments in two well-known data sets, the Indian Pines and Pavia University, have shown that our proposal outperforms the accuracy of classification methods employing EMP and other baseline feature extraction methods with different classifiers.

Index Terms— Remote sensing, Hyperspectral image, Feature extraction, Extended morphological profile, Partial least squares.

1. INTRODUCTION

Remote sensed hyperspectral images have been used for many purposes and have become one of the most important tools in remote sensing [1]. First termed as *imaging spectroscopy* [2], and nowadays known as *hyperspectral imaging*, this technology has allowed many advances in analysis and interpretation of materials on surface of the earth [3]. In the 1980's, inside NASA *Jet Propulsion Laboratory*, researchers started developing new instruments, including the *Airborne Imaging Spectrometer* (AIS), later called *Airborne Visible Infra-Red Imaging Spectrometer* (AVIRIS). This sensor covers a range of the electromagnetic spectrum from $0.4\mu m$ to $2.5\mu m$, at spectral resolution of $10nm$ and more than two hundred spectral channels [4]. This technology can identify some materials that cannot be detected with traditional low-spectral resolution analysis [2], such as the *Multispectral* imaging systems. Hence, one can see *Hyperspectral* imaging as a natural evolution of *Multispectral* imaging [1], which acquires data only in four to seven spectral channels. Typically, images are acquired using an airborne or satellite sensor at short, medium or long distance [3] and the analysis and exploitation of such data can bring many advances to the field of land cover classification/interpretation such as urban planning, ecology, precision agriculture, forestry and military [1, 5].

Although the nature of hyperspectral images to produce rich information of an acquired scene, the scarce referenced available data make classification still a challenging task [1, 3, 5]. This challenge inspires new researches and has received attention in the past years for improving classification and/or interpretation. The conventional multispectral analysis techniques may not be suitable for hyperspectral imaging, emphasizing the need for more advanced approaches to deal with such high dimensionality [1, 5]. To surmount this problem, *kernel based* methods have been applied to this context [6]. In particular, the *Support Vector Machines* (SVM) learning algorithm with *Radial Basis Function* (RBF) kernel has demonstrated promising results among other traditional kernel methods and classifiers [7].

In terms of spatial information, the majority of hyperspectral systems have resolution between 1 and 30 meters per pixel [1], representing great capability of distinguishing structures and objects, which may be able to increase the classification accuracy. Therefore, many researches aim at integrating both spectral and spatial information in their classification systems have been performed [8, 9, 10, 11, 12]. Among them, the *Extended Morphological Profiles* (EMP) approach [8, 3, 13] has demonstrated to be useful to encode both spectral and spatial information in a vector of features.

In this work, we propose a novel classification approach based on the EMP, the *Partial Least Squares* (PLS) [14, 15], and the SVM learning algorithm, referred to as *SpecEmpPLS-SVM*. This approach performs the fusion of both spectral and spatial information to take advantage of the ability of the EMP to encode spatial information and the full spectral data given by the raw hyperspectral image, to form a new feature vector using the PLS transformation which emphasizes the importance of the more discriminative features. Then, the SVM is employed as the classifier.

The results obtained on two widely used data sets (*Indian Pines* and *Pavia University* data sets) reveal that the spectral and the EMP information fusion using the PLS method and the SVM as classifier (SpecEmpPLS-SVM) presents accuracy indexes greater than those achieved by the EMP feature extraction method alone with Multilayer Peceptron Classifier [8, 3, 13] and other baselines based on feature selection and kNN (FEGA-kNN) [16] and the raw data (pixelwise representation) using a SVM classifier.

The remainder of this work is organized as follows. Section 2 describes the main concepts of the extended morphological profile that is used to encode spectral-spatial information. Section 3 presents the proposed method, a classification approach based on the EMP, the PLS, and the SVM, referred to as SpecEmpPLS-SVM. Experiments are described in Section 4 and finally, Section 5 concludes this work.

2. EXTENDED MORPHOLOGICAL PROFILE

The *Extended Morphological Profile* (EMP), a technique based on Mathematical Morphology (MM) [17], has been studied by many researchers [8, 9, 18, 19, 3, 20, 13] and it has shown to be a successful method to encode spectral-spatial information for classification purposes. The MM has a set of operations that allows us to process the image to remove undesirable objects or even increase the size

The authors would like to thank CNPq, CAPES, FAPEMIG, UFMG, and UFOP for the financial support.

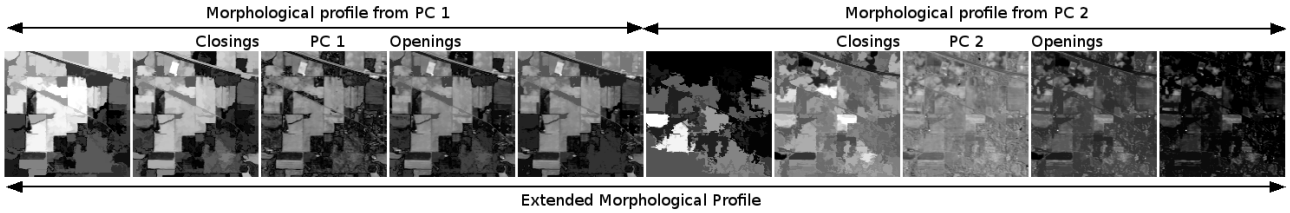

Fig. 1. Example of the EMP on the Indian Pines data set. Number of PCs = 2, discrete disk SE of radius and increasing step equal to 4, and number of openings/closings equal to 2. Total of 10 features in the EMP scheme.

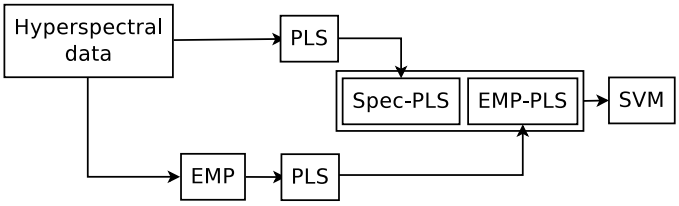

Fig. 2. Diagram of the SpecEmpPLS-SVM approach. The PLS is used to encode both spectral and spatial information which is interpreted by the SVM classifier later.

and shape of the interesting ones. This is done through an interaction with a known structured shape called *Structure Element* (SE), which may vary according to the shape of the object of interest. Two important operations for our concern are called *opening* and *closing*. While the opening has the ability of removing undesirable structures of a bright object without modifying its size, the closing is used aiming to group bright structures.

To build the EMP of an image, we first need to create the *Morphological Profile* (MP) of each channel. In a single channel image, the MP is built by applying successive *openings* and *closings* by *reconstruction* [17]. Each MP enlarges or reduces dark and bright structures using openings and closings by reconstruction, resulting in a more homogeneous version of the original image. At each iteration, the SE is increased to capture further spatial information [8]. In the context of hyperspectral images, we must initially reduce the number of channels in the image, otherwise, applying successive openings/closings over all available spectral channels would lead to a higher dimension feature vector and probably to an intractable classification problem. We recommend the reader to see [21] for some insights on the *Hughes* phenomenon (*curse of dimensionality*).

The traditional EMP approach uses the *Principal Component Analysis* (PCA) to project a hypercube onto a lower dimensional space, keeping the spectral information as much as possible. In [8], the total number of *Principal Components* (PC) used in the EMP scheme should be able to maintain at least 95% of the accumulated variance in the projected image. Fig. 1 illustrates the construction process of an EMP and how the spatial information is captured.

Although the example shown in Fig. 1 corresponds to 10 features, for building a baseline classification approach, we follow the recommendation in [8] that suggests to use four openings/closings, the structuring element with discrete *disk* shape, radius equal to 4 and increasing step equal to 2, resulting in a feature vector of 18 dimensions. A variability threshold number equal to 95% is chosen to select the number of PCs used to build the EMP.

3. PROPOSED APPROACH

In this section, we introduce a new classification approach, referred to as *SpecEmpPLS-SVM*, based on the EMP and the PLS transformation for feature representation and the SVM as classifier. Note that the EMP approach is a spectral-spatial feature representation, and as a consequence the proposed representation as well.

The use of MM has proven its potential to encode spatial information through the EMP approach [3, 13]. In this work, we also propose a new method based in [9] to include both spectral and spatial information into a more compact and discriminant vector of features. It is based on the available spectral information and the EMP approach represented in a single stacked vector of features. This is done by applying a *supervised* feature extraction algorithm on both spectral and EMP feature representation and joining them into a stacked vector [9], as illustrated in Fig. 2. The novelty here is the use of the PLS algorithm [14, 15] to reduce the dimensionality of the data. The new feature representation is then named as SpecEmpPLS.

The PLS algorithm has been used with great success in the field of computer vision to perform some tasks as face identification [22, 23] and pedestrian detection [24, 25, 26], being very effective for dimension reduction, e.g., PLS was able to reduce the feature space from more than 160,000 to 20 dimensions and achieved accurate results [24]. Its classical form is based on the *Nonlinear Iterative Partial Least Squares* (NIPALS) algorithm [14] to model relations among features on the data through *latent variables*. Basically, new features (latent variables) are generated as a linear combination of the original feature set considering its respective classes or categories. In this way, the PLS algorithm searches for an orthogonal projection in which the original features are correlated with the respective class, maximizing the covariance in the latent space as much as possible. More details regarding PLS can be found in [14, 27, 15, 22].

To build the SpecEmpPLS feature representation scheme, we first need to build an EMP fully loaded of spatial information. This is done by using a larger variance threshold to select the projected features, *i.e.*, number of PCs of a PCA. In this case, Fauvel *et al.* [9] suggest to use 99% of variability of the data as threshold in order to select the number of PCs. We keep the other parameters in the traditional EMP described in Section 2. After building the EMP, the PLS algorithm is applied to the data to reduce the dimensionality. Only 5% of training samples were randomly chosen to build the PLS models. We then select a number of features (projected features or latent

(a) *AVIRIS - Indian Pines* - on false color composition

(b) Ground truth

(c) Pixelwise-SVM classification map: OA = $80.24(\pm 0.20)$.

(d) SpecEmpPLS-SVM classification map: OA = **93.95**(± 0.17).

Fig. 3. Results for the *Indian Pines* data set.

variables) which represents 99% of variance of the data, as suggested in [9]. The same procedure is applied to the hyperspectral data, then all projected features of each feature scheme are stacked in a single vector of features. Once the SpecEmpPLS feature representation is built, the SVM learning algorithm [28] with RBF kernel is applied to classify the data. We have chosen such kernel due to its success when dealing with hyperspectral data [9, 3].

4. EXPERIMENTAL RESULTS

In this section, we perform experiments in order to validate our proposed approach. Initially, a description of data sets used and their setup for the experiments is presented. Then, the accuracies obtained for our proposed approach and for other 3 baselines aiming a comparison are presented and finally discussed.

4.1. Evaluation Data Sets

Experiments were performed on two widely used data sets, *Indian Pines* and the *Pavia University*, taken by the *Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)* [4] and *Reflective Optics System Imaging Spectrometer (ROSIS)* [29], respectively. The two data sets are described below (more details regarding these data sets can be found in [30, 31]).

The Indian Pines data set. It is an agricultural area recorded over North-western Indiana, USA, with dimensions of 145×145 pixels, a spatial resolution of $20m$ per pixel, and 220 channels covering a spectral range of $0.4\mu m$ to $2.5\mu m$. Fig. 3(a) shows a RGB false color composition of this area using the following channels: R(50), G(20), and B(10). The same twenty noisy bands (channels 104-108, 150-163, and 220) mentioned in [32] are removed from this data set, remaining 200 spectral bands for the experiments. Sixteen classes of interest are considered and the number of samples per class is shown in the second column in Table 1.

The Pavia University data set. It is an urban area recorded over the University of Pavia, Italy. The image is composed of 610×340 pixels with a spatial resolution of $1.3m/pixel$, spectral range of $0.43\mu m$ to $0.86\mu m$ along its 115 channels. Twelve noisy channels were removed from the remaining 103 bands for the experiments[1]. In the second column in Table 2, the number of samples for each of the 9 classes are presented.

[1]The available data set already contains only 103 bands, and so it is not possible to discriminate which channels were removed from it.

Table 1. Accuracy and $F-score$ results for the *Indian Pines* data set using only 10% for training.

		Individual Classification Approaches			
	# samples	Pixelwise SVM (200 features)	FEGA kNN (95 features)	EMP MLP (18 features)	SpecEmpPLS SVM (16+16 features)
OA(%)		$80.24(\pm 0.20)$	$61.24(\pm 0.36)$	$84.60(\pm 0.78)$	**93.95**(± 0.17)
AA(%)		$67.56(\pm 0.72)$	$50.03(\pm 0.87)$	$80.42(\pm 1.00)$	**85.82**(± 0.90)
Classes		$\mathbf{F-score}$			
Alfalfa	54	$0.22(\pm 0.08)$	$0.21(\pm 0.06)$	$0.72(\pm 0.08)$	**0.93**(± 0.01)
C.-notill	1434	$0.77(\pm 0.00)$	$0.50(\pm 0.01)$	$0.78(\pm 0.03)$	**0.90**(± 0.00)
C.-min	834	$0.68(\pm 0.01)$	$0.41(\pm 0.01)$	$0.77(\pm 0.01)$	**0.93**(± 0.01)
Corn	234	$0.53(\pm 0.04)$	$0.40(\pm 0.03)$	$0.53(\pm 0.05)$	**0.86**(± 0.01)
G./Pasture	497	$0.89(\pm 0.01)$	$0.65(\pm 0.02)$	$0.83(\pm 0.02)$	**0.95**(± 0.00)
G./Trees	747	$0.93(\pm 0.00)$	$0.67(\pm 0.01)$	$0.90(\pm 0.01)$	**0.98**(± 0.00)
G./p.-mowed	26	$0.54(\pm 0.08)$	$0.17(\pm 0.06)$	**0.79**(± 0.06)	$0.67(\pm 0.12)$
Hay-windr.	489	$0.93(\pm 0.00)$	$0.91(\pm 0.00)$	$0.91(\pm 0.02)$	**1.00**(± 0.00)
Oats	20	$0.02(\pm 0.02)$	$0.01(\pm 0.01)$	**0.38**(± 0.06)	$0.12(\pm 0.06)$
S.-notill	968	$0.73(\pm 0.01)$	$0.53(\pm 0.01)$	$0.77(\pm 0.01)$	**0.89**(± 0.00)
S.-min	2468	$0.78(\pm 0.00)$	$0.64(\pm 0.00)$	$0.87(\pm 0.01)$	**0.94**(± 0.00)
S.-clean	614	$0.73(\pm 0.01)$	$0.31(\pm 0.01)$	$0.73(\pm 0.04)$	**0.91**(± 0.01)
Wheat	212	$0.96(\pm 0.01)$	$0.81(\pm 0.01)$	$0.99(\pm 0.00)$	**1.00**(± 0.00)
Woods	1294	$0.94(\pm 0.00)$	$0.82(\pm 0.00)$	$0.98(\pm 0.00)$	**0.99**(± 0.00)
Bld.-G.-T.-D.	380	$0.67(\pm 0.01)$	$0.24(\pm 0.02)$	$0.94(\pm 0.01)$	**0.95**(± 0.01)
St-st towers	95	$0.83(\pm 0.02)$	$0.83(\pm 0.01)$	$0.91(\pm 0.03)$	**0.92**(± 0.02)

Table 2. Accuracy and $F-score$ results for the *Pavia University* data set using only 10% for training.

		Individual Classification Approaches			
	# samples	Pixelwise SVM (103 features)	FEGA kNN (34 features)	EMP MLP (18 features)	SpecEmpPLS SVM (4+8 features)
OA(%)		$93.12(\pm 0.07)$	$88.32(\pm 0.08)$	$95.73(\pm 0.25)$	$98.94(\pm 0.03)$
AA(%)		$90.54(\pm 0.12)$	$85.53(\pm 0.13)$	$93.67(\pm 0.39)$	$98.18(\pm 0.06)$
Classes		$\mathbf{F-score}$			
Asphalt	6631	$0.94(\pm 0.00)$	$0.89(\pm 0.00)$	$0.97(\pm 0.00)$	**0.99**(± 0.00)
Meadow	18649	$0.97(\pm 0.00)$	$0.94(\pm 0.00)$	$0.98(\pm 0.00)$	**1.00**(± 0.00)
Gravel	2099	$0.78(\pm 0.00)$	$0.69(\pm 0.00)$	$0.80(\pm 0.02)$	**0.95**(± 0.00)
Trees	3064	$0.94(\pm 0.00)$	$0.88(\pm 0.00)$	$0.97(\pm 0.00)$	**0.99**(± 0.00)
M. Sheets	1345	$0.99(\pm 0.00)$	$0.99(\pm 0.00)$	$0.98(\pm 0.01)$	**1.00**(± 0.00)
Bare Soil	5029	$0.90(\pm 0.00)$	$0.79(\pm 0.00)$	$0.96(\pm 0.01)$	**1.00**(± 0.00)
Bitumen	1330	$0.85(\pm 0.00)$	$0.77(\pm 0.00)$	$0.86(\pm 0.01)$	**0.96**(± 0.00)
Bricks	3682	$0.85(\pm 0.00)$	$0.78(\pm 0.00)$	$0.90(\pm 0.01)$	**0.97**(± 0.00)
Shadow	947	$0.99(\pm 0.00)$	$0.99(\pm 0.00)$	$0.98(\pm 0.01)$	**1.00**(± 0.00)

4.2. Evaluation criteria

As the size of the training data is an important issue in remote sensed image classification, experiments were carried out by using different training sizes to evaluate the robustness of our proposed approach. We setup our experiments as follows:

Training phase. We train the classification approaches with training sample sizes of 10%, 15%, 20%, and 25% (full training set), where the samples were randomly chosen. The models produced here will be used in the final testing phase.

Testing phase. For each classifier, we ran the experiments 30 times, and then Confidence Intervals (CI) (obtained from the mean and standard deviation) of the Overall Accuracy (OA), Average Accuracy (AA), and $F-score$ ($2\frac{recall \times precision}{recall+precision}$) for each class were reported. The respective sample size of the test sets are 90%, 85%, 80% and 75%.

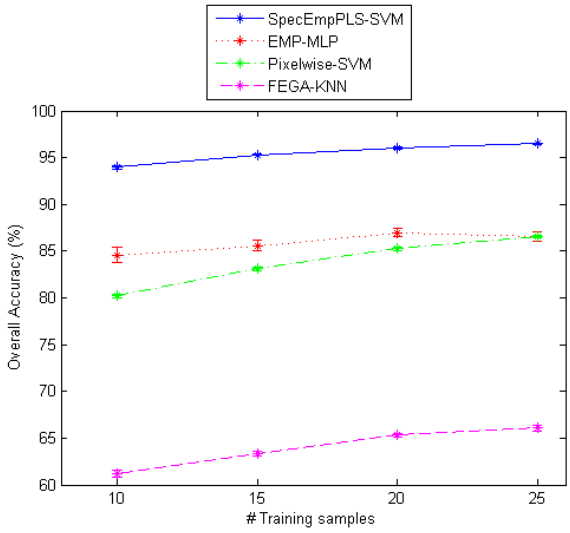

Fig. 4. Results for the *Indian Pines* data set as a function of the samples used for training (in percentage). The star symbols (*) are the averages of 30 runs and the vertical bars are the respective CI.

4.3. Results and Comparisons

In this section we present the results achieved by the proposed method and compare them to other three classification methods: Pixelwise-SVM, FEGA-kNN and EMP-MLP. First, the *Pixelwise-SVM* approach is the simplest one in terms of representation, since it uses the raw data and is based only on spectral information. That is, all the hyperspectral bands of each data set are taken as features. Moreover, the SVM algorithm with RBF kernel is used for classification. The *FEGA-kNN* approach is based on feature selection by means of Genetic Algorithms and it uses the kNN algorithm for classification [16]. This approach is also based only on spectral information. Finally, the *EMP-MLP* uses the well-know EMP feature representation, which is a spatial-spectral representation, and the MLP classifier as suggested in [9].

Due to space constraints, only the thematic maps for the Indian Pines data set produced by the Pixelwise-SVM and the SpecEmpPLS-SVM classification approaches (Fig. 3) is shown for comparison purposes. By observing the values presented in the caption of this figure, we can see that the error accuracy achieved by the SpecEmpPLS-SVM approach is reduced from about 20% (Pixelwise-SVM) to only 6%. Moreover, visually we can notice that the classified regions (Fig. 3(d) *vs* Fig. 3(c)) are quite more consistently defined when compared to the ground truth in Fig. 3(b).

Fig. 4 shows plots of the overall accuracies obtained by our proposal and the three other approaches used in this comparison as a function of the portion of samples used for the training set (*i.e.*, 10%, 15%, 20%, and 25%). It is noticeable that in all cases, the accuracy obtained by our approach is statistically greater than the other approaches. The EMP-MLP and Pixelwise-SVM approaches present similar results when using 25% of training samples and the FEGA-kNN classification approach reaches the worst results in all scenarios, mainly due to the use of kNN classifier. Even though we only show results for the Indian Pines data set, the kNN algorithm does not achieve promising accuracy for both datasets been considered. It is important to remember that the goal of this last approach is to use few features/bands.

Finally, we discuss the results presented in Tables 1 and 2, which show the overall and average accuracies (OA and AA), and the $F-score$ by classes obtained by the proposed approach and the three other baselines for the Indian Pines and Pavia University data sets, respectively. According to both tables, the proposed approach achieves greater overall accuracies when compared to the other methods. Regarding the $F-score$, for the Pavia University data set (Table 2) our approach also achieved the greatest, and best, results for all classes. For the Indian Pines data set (Table 1), the best $F-score$ is achieved by our approach in 14 out 16 classes. The EMP-MLP approach has achieved better results than our approach in only two classes, i.e., *Oats* and *Grass/pasture-mowed* (G./p.-mowed) classes, two small classes (with 26 and 20 samples, respectively in the entire data set). Despite that fact, it reasonable to state that the proposed SpecEmpPLS-SVM approach outperformed the other approaches in terms of accuracy for the two data sets considered in these experiments.

Note that the worst results for the Indian Pines dataset (Table 1) are obtained for the two smallest classes of Indian Pines, which use only one sample for PLS model building (corresponds to 5% of training samples from total of 20 and 26 samples for which class). If a fixed and greater number of samples per class (e.g., a minimum of 10) were used for building the PLS model, the accuracy for these classes would be significantly improved.

5. CONCLUSIONS

In this work, we proposed a novel classification approach based on the EMP, the Partial Least Squares (PLS), and the SVM learning algorithm called SpecEmpPLS-SVM that uses both spectral and spatial information analyzed by PLS to emphasizes the importance of the more discriminative features. The experimental results showed that the proposed approach achieved the best results in two well-known and widely employed data sets when compared with three other methods. In addition, our method also demonstrated to be robust when just a part of the training data is used to learn the classification models, making it useful for a large range of applications, mainly those with restrict number of samples. We can attribute such a feature to the Partial Least Squares since it was designed to work when few samples are available for learning.

6. REFERENCES

[1] C. I. Chang, *Hyperspectral data exploitation: theory and applications*, Wiley-Blackwell, 2007.

[2] A.F.H. Goetz, G. Vane, J.E. Solomon, and B.N. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147, 1985.

[3] A. Plaza, J.A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing Environmet*, vol. 113, pp. 110–122, 2009.

[4] R.O. Green et al., "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sensing of Environment*, vol. 65, no. 3, pp. 227–248, 1998.

[5] S. Prasad, L. M. Bruce, and J. Chanussot, *Optical Remote Sensing: Advances in Signal Processing and Exploitation Techniques*, vol. 3, Springer, 2011.

[6] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, MA, USA, 2001.

[7] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.

[8] J.A. Benediktsson, J.A. Palmason, and J.R. Sveinsson, "Classification of hyperspectral data from urban areas based on extrended morphological profiles," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480–491, 2005.

[9] M. Fauvel, J. Chanussot, J.A. Benediktsson, and J.R. Sveinsson, "Spectral and spatial classification of hyperspectral data using svms and morphological profiles," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, pp. 3804–3814, 2008.

[10] Y. Tarabalka, J.A. Benediktsson, J. Chanussot, and J.C. Tilton, "Multiple spectral-spatial classification approach for hyperspectral data," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4122–4132, 2010.

[11] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng, "Adaptative markov random field approach for classification of hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 5, pp. 973–977, 2011.

[12] L. Shen and S. Lia, "Three-dimensional gabor wavelets for pixel-based hyperspectral imagery classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 49, no. 12, pp. 5039–5046, 2011.

[13] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 652–675, 2013.

[14] H. Wold, "Partial least squares," *Encyclopedia of Statistical Sciences*, vol. 6, pp. 581–591, 1985.

[15] R. Rosipal and N. Krämer, "Overview and recent advances in partial least squares," in *Int. Conf. on Subspace, Latent Structure and Feature Selection (SLSFS'05)*, 2006, pp. 34–51.

[16] A. B. Santos, C. S. F. de S. Celes, A. de A. Araújo, and D. Menotti, "Feature selection for classification of remote sensed hyperspectral images: A filter approach using genetic algorithms and cluster validity," in *IPCV*, 2012, pp. 675–681.

[17] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press., 1982.

[18] M. Dalla Mura, A. Villa, J.A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 3, pp. 542–546, 2011.

[19] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas," *EURASIP J. Adv. Signal Process*, vol. 2009, no. 11, pp. 1–14, 2009.

[20] G. Licciardi, P.R. Marpu, J. Chanussot, and J.A. Benediktsson, "Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 3, pp. 447–451, 2012.

[21] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. on Information Theory*, vol. 14, no. 1, pp. 55–63, 1968.

[22] W. R. Schwartz, H. Guo, J. Choi, and L. S. Davis, "Face Identification Using Large Feature Sets," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2245–2255, 2012.

[23] W. Schwartz, H. Guo, and L. Davis, "A robust and scalable approach to face identification," in *European Conference on Computer Vision (ECCV)*, 2010, pp. 476–489.

[24] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, "Human detection using partial least squares analysis," in *IEEE 12th International Conference on Computer Vision (ICCV)*, 2009, pp. 24–31.

[25] W. R. Schwartz and L. S. Davis, "Learning discriminative appearance-based models using partial least squares," in *XXII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, 322–329, p. 2009.

[26] W. R. Schwartz, L. S. Davis, and H. Pedrini, "Local Response Context Applied to Pedestrian Detection," in *IberoAmerican Congress on Pattern Recognition (CIARP)*, 2011, pp. 181–188.

[27] M. Barker and W. Rayens, "Partial least squares for discrimination," *J. of Chemometrics*, vol. 17, no. 3, pp. 166–173, 2003.

[28] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification and Scene Analysis*, John Wiley & Sons, 2 edition, 2000.

[29] P. Gege, D. Beran, W. Mooshuber, J. Schulz, and H. Van Der Piepen, "System analysis and performance of the new version of the imaging spectrometer ROSIS," in *The First EARSeL Workshop on Imaging Spectroscopy*, 1998, pp. 29–35.

[30] M. Fauvel, *Spectral and Spatial Methods for the Classification of Urban Remote Sensing Data*, Ph.D. thesis, University of Iceland and Grenoble Institute of Technology, 2007.

[31] Y. Tarabalka, *Classification of Hyperspectral Data Using Spectral-Spatial Approaches*, Ph.D. thesis, University of Iceland and Grenoble Institute of Technology, 2010.

[32] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 93–97, 2006.