GMM Supervectors for Limited Training Data in Hyperspectral Remote Sensing Image Classification

AmirAbbas Davari, Vincent Christlein, Sulaiman Vesal, Andreas Maier, Christian Riess

Pattern Recognition Lab, Computer Science Department,
Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany

Abstract. Severely limited training data is one of the major and most common challenges in the field of hyperspectral remote sensing image classification. Supervised learning on limited training data requires either a) designing a highly capable classifier that can handle such information scarcity, or b) designing a highly informative and easily separable feature set. In this paper, we adapt GMM supervectors to hyperspectral remote sensing image features. We evaluate the proposed method on two datasets. In our experiments, inclusion of GMM supervectors leads to a mean classification improvement of about 4.6%.

Keywords: hyperspectral image classification, remote sensing, limited training data, GMM supervector

1 Introduction

Remote sensing plays an important role for various applications, including environmental monitoring, urban planning, ecosystem-oriented natural resources management, urban change detection and agricultural region monitoring [27]. Most of these monitoring and detection applications require the construction of a label map from the remotely sensed images. In these label maps, individual pixels are marked as members of specific classes like for example water, asphalt, or grass. The assignment of observations to labels is done via classification, which explains the importance of classification for remote sensing applications.

The availability of very high resolution hyperspectral remote sensing images (VHR-RSI) has drawn researchers' attention over the past decade. It has been shown that jointly exploiting spectral and spatial information improves classification performance compared to using spectral features alone [19]. To this end, extended multi-attribute profiles (EMAP) [8] are one of the most popular and powerful feature descriptors for such spectral-spatial pixel representations. By operating directly on connected components rather than pixels, EMAP allows to employ arbitrary region descriptors (e.g. shape, color, texture, etc.) and to support object-based image analysis. In addition, EMAP can be implemented efficiently via various tree-based hyperspectral (HS) image representation techniques

like max- and min-tree [23] or alpha tree [24]. EMAPs unique characteristics make it a popular tool in the hyperspectral remote sensing image analysis community [14, 21, 30].

A long-standing problem in hyperspectral remote sensing is image classification based on only a limited number of labeled pixels, as the process of labeling pixels for training data is a manual, time-consuming and expensive procedure. Additionally, popular descriptors like EMAP typically provide high dimensional features. These two factors together lead to a relatively small ratio of labeled data compared to the overall feature dimensionality, and can potentially cause a problem known as Hughes phenomenon [15]. Researchers have put considerable effort into developing different algorithms to address this problem. These approaches can be categorized into two groups: 1) developing new classifiers or reformulating existing ones in order to work well with the limited training data [3, 6, 12, 12, 16, 26, 28], and 2) using dimensionality reduction. In particular, supervised dimensionality reduction techniques like NWFE [18], DAFE [11], or DBFE [20] has been shown to oftentimes outperform unsupervised reduction techniques [4] like PCA. However, both approaches, i.e. designing classifiers capable of handling limited amount of training data and feature vector dimensionality reduction are challenged in extreme cases when training data is severely limited.

Gaussian Mixture Model supervectors have been successfully used to handle limited training data in several other applications, such as online signature verification [31], writer identification [7], or speech analysis [1, 5, 17, 25]. In this paper, we adopt GMM supervectors with a universal background model (GMM-UBM) to hyperspectral remote sensing image classification in the presence of limited training data. We apply the GMM-UBM approach to feature extraction. To deal with small size training sets, We concatenating all the means of the GMM components to form the supervectors. We show that such supervectors are highly effective in addressing the limited data problem.

Section 2 introduces the proposed workflow and tools used for feature extraction in the context of our framework for hyperspectral remote sensing images. Section 3 describes the experimental setup, datasets, feature extraction mechanism and classification algorithm and illustrates the results. Finally, our work will be concluded in Section 4.

2 Methodology

We first introduce Gaussian supervectors, then present the proposed workflow to limited data classification.

2.1 GMM Supervector

The computation of GMM supervectors consists of three main components, namely universal background model computation, adaptation to the data and normalization [22].

Universal Background Model The universal background model (UBM) essentially is a GMM fitted to the labeled training data. In detail, let $\lambda = \{w_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \,|k = 1, \ldots, K\}$ denote the parameters of a GMM with K mixture components, where w_k, $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$ denote the k-th mixture weight, mean vector and covariance matrix, respectively.

Given a feature vector $\boldsymbol{x} \in \mathcal{R}^D$, its likelihood function is defined as

$$p(\boldsymbol{x} \,|\, \lambda) = \sum_{k=1}^{K} w_k g_k(\boldsymbol{x}) \ , \tag{1}$$

where $g_k(\boldsymbol{x})$ is a function to evaluate the k-th Gaussian at position $\boldsymbol{x}$, i.e.,

$$g_k(\boldsymbol{x}) = g(\boldsymbol{x}\,;\, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}_k|}} \, \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_k)} \ . \tag{2}$$

The mixture weights are positive real numbers, i.e. $w_k \in \mathbb{R}_+$, and satisfy the constraint $\sum_{k=1}^{K} w_k = 1$.

Finally, the posterior probability of a feature vector $\boldsymbol{x}_j$ to be generated by the Gaussian mixture k is

$$\gamma_k(\boldsymbol{x}_j) = p(k \,|\, \boldsymbol{x}_j) = \frac{w_k g_k(\boldsymbol{x}_j)}{\sum_{l=1}^{K} w_l g_l(\boldsymbol{x}_j)} \ . \tag{3}$$

For estimating the GMM parameters, expectation maximization is used for a maximum likelihood estimation [10]. The parameters λ are iteratively refined to increase the log-likelihood $\log p(\boldsymbol{X} \,|\, \lambda) = \sum_{m=1}^{M} \log p(\boldsymbol{x}_m \,|\, \lambda)$ of the model for the set of training samples $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$.

In the original formulation of Gaussian supervectors, all parameters, i.e., weights w_k, means $\boldsymbol{\mu}_k$ and covariances $\boldsymbol{\Sigma}_k$ are used. However, in our case of severely limited training data, we found that supervectors consisting only of the means are much more robust. Thus, we will eventually only use these means for supervector formation.

GMM Adaptation and Mixing A key idea of the Gaussian supervectors is to adapt the GMM components to the distribution of the test set. Since we will only use the means, we only describe the adaptation of the GMM mean vectors.

Thus, let $\boldsymbol{X}_f = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T\}$ denote the D-dimensional feature representations of the T pixels in the test set. Let furthermore $n_k = \sum_{t=1}^{T} \gamma_k(\boldsymbol{x}_t)$. Then, the first order statistic for adaptation of the UBM to the test data is

$$E_k^1 = \frac{1}{n_k} \sum_{t=1}^{T} \gamma_k(\boldsymbol{x}_t) \boldsymbol{x}_t \ , \tag{4}$$

$$\tag{5}$$

where $E_k^1 \in \mathbb{R}^D$. The adapted mean vectors $\hat{\boldsymbol{\mu}}_k$ can then be computed as

$$\hat{\boldsymbol{\mu}}_k = \alpha_k E_k^1 + (1 - \alpha_k) \boldsymbol{\mu}_k \ , \tag{6}$$

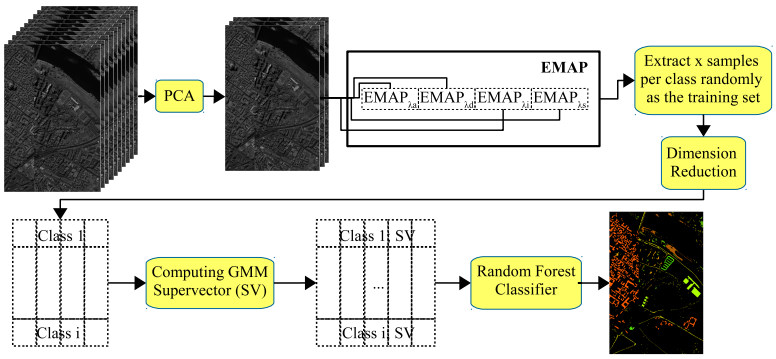


Fig. 1. Proposed workflow.

where

$$\alpha_k = \frac{n_k}{n_k + r} \quad , \tag{7}$$

and r denotes a fixed relevance factor that controls the strength of the adaptation.

Finally, the supervector $\boldsymbol{s}$ is formed by concatenating the adapted GMM parameters. As stated above, we use only the adapted mean components, leading to

$$\boldsymbol{s} = \left(\hat{\boldsymbol{\mu}}_1^\top, \ldots, \hat{\boldsymbol{\mu}}_K^\top\right)^\top \quad . \tag{8}$$

Normalization The purpose of normalization is to transform the $\boldsymbol{s}$ in Eqn. 8 into a common range. Several groups used for this task a feature mapping inspired by the symmetrized Kullback-Leibler divergence [7, 29] . This mapping is referred to as KL-normalization and is computed for the mean vectors as

$$\tilde{\boldsymbol{\mu}}_k = \sqrt{w_k}\boldsymbol{\sigma}_k^{-\frac{1}{2}} \odot \hat{\boldsymbol{\mu}}_k \quad , \tag{9}$$

where $\boldsymbol{\sigma}_k$ represents the GMM's k-th component standard deviation, $\tilde{\boldsymbol{\mu}}_k$ denotes the normalized adapted mean vector, and $\odot$ denotes the Hadamard product. In the case of mean adaptation, the normalized supervector $\tilde{\boldsymbol{s}}_\mathrm{m}$ is represented as

$$\tilde{\boldsymbol{s}}_\mathrm{m} = \left(\tilde{\boldsymbol{\mu}}_1^\top, \ldots, \tilde{\boldsymbol{\mu}}_K^\top\right)^\top \quad . \tag{10}$$

2.2 Proposed Workflow

A high-level overview of the proposed method is shown in Fig. 1. For hyperspectral remote sensing image classification, we use a standard dimensionality-reduction workflow. Here, spectral bands are first reduced via principle component analysis (PCA). Then, extended multi-attribute profile (EMAP) [8] is computed as the feature vector. From the computed EMAP, c samples per class are randomly

selected as the training feature set. These training features are again subject to dimensionality reduction, denoted by EMAP-reduced.

The key contribution of the method is injected right before classification: we propose to compute Gaussian mixture model supervectors from the EMAP/EMAP-reduced training features, denoted by EMAP-SV/EMAP-reduced-SV, and use in the classifier the supervectors as the features of the hyperspectral image. A GMM from such limited training data is necessarily only a coarse approximation of the underlying distribution. Nevertheless, we show that it is just good enough to support the classifier in better determining the class boundaries. The parametrization of the standard pipeline follows dataset-dependent recommendations from the literature, and is reported in Sec. 3.2.

3 Experiments

We first introduce the datasets that were used for the evaluation in Sec. 3.1. Feature extraction and classification are presented in Sec. 3.2 and Sec. 3.3, respectively. Quantitative and qualitative results are presented in Sec. 3.4 and Sec. 3.5, respectively.

3.1 Data Sets

In order to evaluate our method, we use two popular datasets that were acquired by two different sensors. First, Pavia Centre dataset has been acquired by the ROSIS sensor in 115 spectral bands during a flight campaign over Pavia, northern Italy. 13 of these bands are removed due to noise. Therefore, 102 bands are used in this work. The scene image is 1096×715 pixels with geometrical resolution of 1.3 m. For the computation of EMAP, we use the first three principle components of this dataset, containing 99.12% of the total spectral variance. Second, the Salinas dataset was acquired by AVIRIS sensor in 224 spectral bands over Salinas Valley, California. 20 of the water absorption bands were discarded and therefore 204 bands are used in this work. The scene image is 512×217 pixels with high spatial resolution of 3.7 meter pixels. We use the first three principle components of this dataset, containing 99.14% of the total spectral variance, for the computation of EMAP.

3.2 Feature Extraction

The host feature vector used in this work is the extended multi-attribute profile (EMAP) with four attributes and four thresholds, λ, per each attribute. For the Pavia Centre dataset we use the same threshold values as in [8]. For the Salinas dataset, we use the same threshold values as in [21]. The attributes and their corresponding threshold values for the Pavia Center are

- Area of the connected components: $\lambda_a = [100, 500, 1000, 5000]$;
- Length of the bounding box diagonal fit over the of the connected components: $\lambda_d = [10, 25, 50, 100]$;

– Standard deviation of the gray values of the connected components: $\lambda_s = [20, 30, 40, 50]$;
– Moment of inertia [13]: $\lambda_i = [0.2, 0.3, 0.4, 0.5]$;

The attributes and their corresponding threshold values for the Salinas dataset are

– Area of the connected components: $\lambda_a = [100, 500, 1000, 5000]$;
– Length of the bounding box diagonal fit over the of the connected components: $\lambda_d = [10, 25, 50, 100]$;
– Standard deviation of the gray values of the connected components: $\lambda_s = [15, 20, 25, 30]$;
– Moment of inertia: $\lambda_i = [0.1, 0.15, 0.2, 0.25]$.

For calculating the EMAPs, the Max-tree hierarchical image representation is used. We use max-filtering [9] for filtering the Max-tree with each value in λ.

We implemented two variants of the second dimensionality reduction (DR) to investigate its impact on our proposed approach. Specifically, we reduce the EMAP dimensionality in one variant with principle component analysis (PCA), and in another variant with non-parametric weighted feature extraction (NWFE) [18]. PCA is a popular unsupervised DR method. NWFE is supervised and very strong performance has been reported for this method [4]. For the Pavia Centre dataset, 7 PCA dimensions and 6 NWFE dimensions were used to preserve 99% of the variance of the input EMAP features. For the Salinas dataset, 4 PCA dimensions and 7 NWFE dimensions were used to preserve 99% of the variance of the input EMAP features.

The supervectors (SV) are computed over the aforementioned raw EMAP/ EMAP-reduced feature vectors. The number of GMM components is set to 3. However, in preliminary experiments we found that the choice of this parameter is not critical to this work. The relevance factor is set to $r = 16$ as it is commonly used in the literature [17,22]. Kullback-Leibler divergence is used for normalization of the supervectors. The supervectors are computed over EMAP results (denoted as EMAP-SV) and EMAP-reduced results (denoted as EMAP-reduced-SV), respectively.

3.3 Classification

We use the random forest classifier with 100 trees. The tree depth and the bagging number is set to be square root of the number of input variables by default as suggested in [2]. For training, we randomly select c pixels per class from the image as the training set. All the remaining pixels are used for testing. In order to simulate the severely limited training data case, we choose c to be 13 and 20 as two different training set sizes. For each experiment, this procedure was repeated 25 times. The performance metrics are overall accuracy, average accuracy and kappa, abbreviated as OA, AA, and Kappa:

– OA: The overall accuracy is the number of correctly classified instances divided by the total number of data points (pixels).

– AA: The average accuracy is the average of class-based accuracies.
– Kappa: The kappa statistic is a measure of how closely the instances classified by the classifier matched the ground truth. By measuring the expected accuracy, it gives a statistic for the accuracy of a random classifier.

3.4 Quantitative Results

For quantitative evaluation, we compare a total of twelve combinations of the proposed algorithm: we use PCA and NWFE as dimensionality reduction of the EMAP features, and we apply these two variants on both datasets. We compare the classification performance on raw EMAP, EMAP-PCA and EMAP-NWFE to the supervectors computed over each of them. For each of these sets, we choose two selections of training sets, namely 12 and 20 pixels per class. Table 1 and Tab. 2 show the classification results of the aforementioned feature sets computed over Pavia Centre dataset and Salinas dataset, respectively.

Both Tab. 1 and Tab. 2 show that the smaller the training size, the higher the performance gain achieved by using GMM supervectors. For example, consider the case of EMAP feature computed over Pavia Centre dataset in Tab. 1. In the case of 20 pixels per class, using supervectors results in a Kappa improvement of 0.0312. With a training data size of only 13 pixels per class, the Kappa improvement is even 0.0598. Thus, the proposed method has a bigger impact in applications with severely limited training data.

To study the effect of the second level dimensionality reduction algorithms on our idea, we reduced the dimensionality of EMAP by means of PCA and NWFE to a number that preserves 99% of the data variance. It turns out that the performance boost achieved by the supervectors over raw EMAP variants is consistent over variants of dimensionality reduction algorithms, i.e. PCA and NWFE. All metrics show improvement for both EMAP-PCA supervector and EMAP-NWFE supervector comparing to raw EMAP classification. This shows the consistency of the method over different dimensionality reduction techniques. We also note that the performance gained by NWFE is higher than for PCA. Furthermore, the standard deviation of the proposed method is consistently low. Thus, by using supervectors, robustness of the training set is increased with respect to the class-wise structure. By extension, the classifier becomes more robust and consistent on different training samples.

Finally, supervectors computed over EMAP variants using less training samples oftentimes lead to a comparable or sometimes even higher performance than raw EMAP variants using more training samples without synthetic samples. For example in Table 1, EMAP-PCA-SV on 13 training samples achieves a Kappa of 0.9198. This is higher than the Kappa of 0.8974 that is achieved by EMAP-PCA-raw on 20 training samples. Similarly, in Table 2, EMAP-NWFE-SV on 13 training samples obtain higher Kappa values than EMAP-NWFE-raw on 20 training samples.

Table 1. Classification performances of raw EMAP, EMAP-PCA and EMAP-NWFE vs. their supervector (SV) correspondences, computed over Pavia Centre dataset. This tables shows the results for two training data sizes namely 13 and 20 pixels per class.

Algorithm	Feature	AA% (± SD)	OA% (± SD)	Kappa (± SD)
13 Pix/Class				
EMAP	raw	77.87 (±2.97)	90.01 (±3.78)	0.8600 (±0.0495)
	SV	**88.73** (±1.30)	**94.28** (±0.94)	**0.9198** (±0.0129)
EMAP-PCA	raw	73.51 (±3.00)	86.38 (±3.61)	0.8089 (±0.0493)
	SV	**82.07** (±1.96)	**91.70** (±1.67)	**0.8838** (±0.0225)
EMAP-NWFE	raw	80.06 (±3.56)	91.37 (±2.67)	0.8787 (±0.0365)
	SV	**88.02** (±1.17)	**95.39** (±0.42)	**0.9349** (±0.0059)
20 Pix/Class				
EMAP	raw	81.80 (±2.07)	92.73 (±1.23)	0.8974 (±0.0171)
	SV	**90.43** (±1.22)	**94.92** (±0.77)	**0.9286** (±0.0106)
EMAP-PCA	raw	79.07 (±1.69)	90.89 (±1.22)	0.8717 (±0.0169)
	SV	**83.12** (±2.07)	**92.37** (±1.09)	**0.8928** (±0.0151)
EMAP-NWFE	raw	83.32 (±2.24)	93.28 (±1.30)	0.9053 (±0.0181)
	SV	**88.91** (±0.86)	**95.67** (±0.55)	**0.9389** (±0.0077)

3.5 Qualitative Results

Figure 2 shows example label maps corresponding to the classification results for training size of 13 pixels per class for the Salinas dataset. In this Figure, (a) shows the ground truth labeling, (b) shows the output of raw EMAP, (c) shows the EMAP supervector results. Analogously, (d) shows the output of raw EMAP-PCA, and (e) shows EMAP-PCA supervector, (f) shows the corresponding results for raw EMAP-NWFE, and (g) shows the output of EMAP-NWFE supervector, i.e., identical processing pipelines using NWFE dimensionality reduction, with raw and supervector feature sets. Comparing (e) and (g) confirms the superiority of EMAP-NWFE supervectors over the EMAP-PCA supervector. Comparing (b) with (c), (d) with (e) and (f) with (g), specially in the large homogeneous areas, clearly shows that using the supervectors avoids a number of misclassification that are present in the raw EMAP, EMAP-PCA and EMAP-NWFE.

4 Conclusion

Limited training data is a common issue in hyperspectral remote sensing image classification. This limitation severely challenges classifiers, particularly when using high dimensional feature vectors. We propose to use GMM supervectors

Table 2. Classification performances of raw EMAP, EMAP-PCA and EMAP-NWFE vs. their supervector (SV) correspondences, computed over Salinas dataset. This tables shows the results for two training data sizes namely 13 and 20 pixels per class.

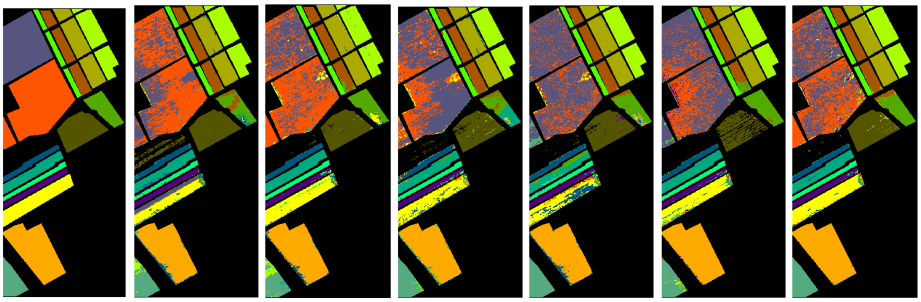
Algorithm	Feature	AA% (± SD)	OA% (± SD)	Kappa (± SD)
13 Pix/Class				
EMAP	raw	83.84 (±2.06)	76.30 (±2.74)	0.7380 (±0.0292)
	SV	**90.90** (±0.98)	**85.37** (±1.34)	**0.8378** (±0.0147)
EMAP-PCA	raw	82.50 (±2.06)	74.96 (±3.63)	0.7230 (±0.0378)
	SV	**86.84** (±1.77)	**78.30** (±2.90)	**0.7606** (±0.0311)
EMAP-NWFE	raw	88.68 (±1.20)	80.42 (±2.34)	0.7838 (±0.0247)
	SV	**91.43** (±1.00)	**83.09** (±1.72)	**0.8132** (±0.0186)
20 Pix/Class				
EMAP	raw	86.81 (±1.63)	79.74 (±2.56)	0.7756 (±0.0269)
	SV	**92.88** (±0.71)	**88.09** (±1.29)	**0.8680** (±0.0142)
EMAP-PCA	raw	86.59 (±1.06)	78.70 (±2.33)	0.7643 (±0.0249)
	SV	**86.99** (±1.28)	**79.40** (±1.88)	**0.7725** (±0.0205)
EMAP-NWFE	raw	90.56 (±1.26)	82.26 (±2.62)	0.8038 (±0.0280)
	SV	**91.91** (±0.81)	**83.40** (±1.94)	**0.8168** (±0.0208)

with a universal background model to address the limited data problem. In our results on real data, we show the performance gain on the Pavia Centre and the Salinas datasets. It turns out that supervectors consistently increase the overall accuracy, average accuracy, and kappa coefficient. Furthermore, the performance boost using supervectors is consistent over different dimensionality reduction algorithms and different training data sizes. It can also be observed that using supervectors decreased the standard deviations of the error metrics.

Quantitatively, the exact performance improvement depends on the details of the processing chain and on the dataset. The mean improvement in our experiments is almost 4.6%, with variations between one percent and almost ten percent. These results are encouraging, as the approach itself is quite straightforward, and can be smoothly integrated into any classification pipeline.

References

1. Bahari, M. H., Saeidi, R., hamme, H. V., and Leeuwen, D. V. Accent recognition using i-vector, gaussian mean supervector and gaussian posterior probability supervector for spontaneous telephone speech. In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (may 2013), Institute of Electrical and Electronics Engineers (IEEE).



(a) (b) (c) (d) (e) (f) (g)

Fig. 2. Example label maps on Salinas using 13 training samples per class. (a) ground truth (b) EMAP, (c) EMAP-SV (d) EMAP-PCA, (e) EMAP-PCA-SV, (f) EMAP-NWFE, (g) EMAP-NWFE-SV.

2. Breiman, L. Random forests. Machine Learning 45, 1 (2001), 5–32.
3. Bruzzone, L., Chi, M., and Marconcini, M. A novel transductive svm for semisupervised classification of remote-sensing images. Geoscience and Remote Sensing, IEEE Transactions on 44, 11 (2006), 3363–3373.
4. Castaings, T., Waske, B., Atli Benediktsson, J., and Chanussot, J. On the influence of feature reduction for the classification of hyperspectral images based on the extended morphological profile. International Journal of Remote Sensing 31, 22 (2010), 5921–5939.
5. Cerva, P., Silovsky, J., and Zdansky, J. Comparison of generative and discriminative approaches for speaker recognition with limited data. Radioengineering (2009).
6. Chi, M., Feng, R., and Bruzzone, L. Classification of hyperspectral remote-sensing data with primal svm for small-sized training dataset problem. Advances in space research 41, 11 (2008), 1793–1799.
7. Christlein, V., Bernecker, D., Hönig, F., Maier, A., and Angelopoulou, E. Writer identification using GMM supervectors and exemplar-SVMs. Pattern Recognition 63 (2017), 258–267.
8. Dalla Mura, M., Atli Benediktsson, J., Waske, B., and Bruzzone, L. Extended profiles with morphological attribute filters for the analysis of hyperspectral data. International Journal of Remote Sensing 31, 22 (2010), 5975–5991.
9. Dalla Mura, M., Benediktsson, J. A., Waske, B., and Bruzzone, L. Morphological attribute profiles for the analysis of very high resolution images. Geoscience and Remote Sensing, IEEE Transactions on 48, 10 (2010), 3747–3762.
10. Dempster, A., Laird, N., and Rubin, D. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39, 1 (1977), 1–38.
11. Fukunaga, K. Introduction to statistical pattern recognition. Academic press, 2013.
12. Hoffbeck, J. P., and Landgrebe, D. A. Covariance matrix estimation and classification with limited training data. IEEE Transactions on Pattern Analysis and Machine Intelligence 18, 7 (1996), 763–767.
13. Hu, M.-K. Visual pattern recognition by moment invariants. IRE transactions on information theory 8, 2 (1962), 179–187.

14. Huang, X., Guan, X., Benediktsson, J. A., Zhang, L., Li, J., Plaza, A., and Dalla Mura, M. Multiple morphological profiles from multicomponent-base images for hyperspectral image classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7, 12 (2014), 4653–4669.
15. Hughes, G. On the mean accuracy of statistical pattern recognizers. Information Theory, IEEE Transactions on 14, 1 (1968), 55–63.
16. Jackson, Q., and Landgrebe, D. A. An adaptive classifier design for high-dimensional data analysis with a limited training data set. Geoscience and Remote Sensing, IEEE Transactions on 39, 12 (2001), 2664–2679.
17. Kelly, F. Automatic recognition of ageing speakers. PhD thesis, Ph. D. thesis, Trinity College Dublin, 2014.
18. Kuo, B.-C., and Landgrebe, D. A. Nonparametric weighted feature extraction for classification. IEEE Transactions on Geoscience and Remote Sensing 42, 5 (2004), 1096–1105.
19. Landgrebe, D. A. Signal theory methods in multispectral remote sensing, vol. 29. John Wiley & Sons, 2005.
20. Lee, C., and Landgrebe, D. A. Feature extraction based on decision boundaries. IEEE Transactions on Pattern Analysis and Machine Intelligence 15, 4 (1993), 388–400.
21. Liu, T., Gu, Y., Jia, X., Benediktsson, J. A., and Chanussot, J. Class-specific sparse multiple kernel learning for spectral–spatial hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 54, 12 (2016), 7351.
22. Reynolds, D. A., Quatieri, T. F., and Dunn, R. B. Speaker Verification Using Adapted Gaussian Mixture Models. Digital Signal Processing 10, 1-3 (2000), 19–41.
23. Salembier, P., Oliveras, A., and Garrido, L. Antiextensive connected operators for image and sequence processing. Image Processing, IEEE Transactions on 7, 4 (1998), 555–570.
24. Soille, P. Constrained connectivity for hierarchical image partitioning and simplification. Pattern Analysis and Machine Intelligence, IEEE Transactions on 30, 7 (2008), 1132–1145.
25. Srinivasan, B. V., Zotkin, D. N., and Duraiswami, R. A partial least squares framework for speaker recognition. In 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (may 2011), Institute of Electrical and Electronics Engineers (IEEE).
26. Tadjudin, S., and Landgrebe, D. A. Covariance estimation for limited training samples. In Geoscience and Remote Sensing Symposium Proceedings, 1998. IGARSS'98. 1998 IEEE International (1998), vol. 5, IEEE, pp. 2688–2690.
27. Valero, S., Salembier, P., and Chanussot, J. Hyperspectral image representation and processing with binary partition trees. Image Processing, IEEE Transactions on 22, 4 (2013), 1430–1443.
28. Vatsavai, R. R., Shekhar, S., and Burk, T. E. A semi-supervised learning method for remote sensing data mining. In Tools with Artificial Intelligence, 2005. ICTAI 05. 17th IEEE International Conference on (2005), IEEE, pp. 5–pp.
29. Xu, M., Zhou, X., Li, Z., Dai, B., and Huang, T. S. Extended Hierarchical Gaussianization for Scene Classification. In Image Processing (ICIP), 2010 17th IEEE International Conference on (Hong Kong, Sept. 2010), pp. 1837–1840.
30. Xu, X., Li, J., Huang, X., Dalla Mura, M., and Plaza, A. Multiple morphological component analysis based decomposition for remote sensing image classification. IEEE Transactions on Geoscience and Remote Sensing 54, 5 (2016), 3083–3102.

31. Zapata-Zapata, G. J., Arias-Londoño, J. D., Vargas-Bonilla, J. F., and Orozco-Arroyave, J. R. On-line signature verification using gaussian mixture models and small-sample learning strategies. Revista Facultad de Ingeniería Universidad de Antioquia, 79 (2016), 86–97.