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, Department of Computer Science, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany

Introduction

- Labeling the HSRS images is a very expensive work.
- Though a lot of data is recorded, very little ground truth is available.
- We denote as "severely limited data" cases where less than 50 pixels per class are available.

Existing approaches to HSRS limited data classification:

- **Dimensionality reduction** via Unsupervised PCA and Supervised NWFE [1].
- **Specialized classifiers** [2].

However, these methods are challenged by severely limited training data.

Contribution: Use of GMM supervectors [3] to dynamically adapt to the limited data.

GMM Supervectors

- EMAP [4] is the base feature vector used in this work
- GMM Supervectors are built on top of EMAP.

GMM supervectors computation steps

1. Universal background model
   - Representative model of the data, i.e. GMM.
2. Adaptation to the data
   - Adapted mean for each component: $\hat{\boldsymbol{\mu}}_k = \alpha_k E_k^1 + (1-\alpha_k)\boldsymbol{\mu}_k$

     Where $E_k^1 = \frac{1}{n_k}\sum_{t=1}^{T}\gamma_k(\boldsymbol{x}_t)\boldsymbol{x}_t$, $\alpha_k = \frac{n_k}{n_k + r}$, $n_k = \sum_{t=1}^{T}\gamma_k(\boldsymbol{x}_t)$

     $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$ denote the D-dimensional features representations of the T pixels in the test set.

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

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

3. Normalization via symmetrized Kullback Leibler divergence
   - Purpose: to bring the supervectors into a common range.
   - Normalized adapted means: $\tilde{\boldsymbol{\mu}}_k = \sqrt{w_k}\boldsymbol{\sigma}_k^{-\frac{1}{2}} \odot \hat{\boldsymbol{\mu}}_k$

     where $\odot$ represents the Hadamard product.
   - Mean supervector $\tilde{\boldsymbol{s}}_{\mathrm{m}} = \left(\tilde{\boldsymbol{\mu}}_1^\top, \ldots, \tilde{\boldsymbol{\mu}}_K^\top\right)^\top$

Experimental Setup

EMAP's attributes and thresholds:

- Area: 100, 500, 1000, 5000
- Standard Deviation: 20, 30, 40, 50
- First Moment of Hu: 0.2, 0.3, 0.4, 0.5
- Bounding Box Diagonal: 10, 25, 50, 100

Classifier:

- Random forest classifier
- 100 trees
- the number of variables per node is square root of number of features.

Classification:

- Training set: 13, 20 pixels per class were randomly selected from the image as separate training sets.
- For each experiment, the random selection of the training set was repeated 25 times and the average of the overall accuracy, average accuracy and kappa statistics were calculated and reported.

Dataset 1: Pavia Centre

- acquired by the ROSIS sensor
- 610 * 340 pixels
- geometrical resolution of 1.3 m
- 103 spectral bands
- We used the first four of its PCs which contained 99.16% of the total variance

Dataset 2: Salinas Valley

- acquired by the AVIRIS sensor
- 512 * 217 pixels
- geometrical resolution of 3.7 m
- 204 spectral bands
- We used the first four of its PCs which contained 99.68% of the total variance

Results and Conclusions

- Mean classification improvement of about 4.6%.
- Supervectors consistently increase the overall accuracy, average accuracy, and kappa coefficient.
- Consistent over different dimensionality reduction algorithms and different training data sizes.
- Standard deviations of the error metrics are decreased.
- Easy to be smoothly integrated into any classification pipeline.

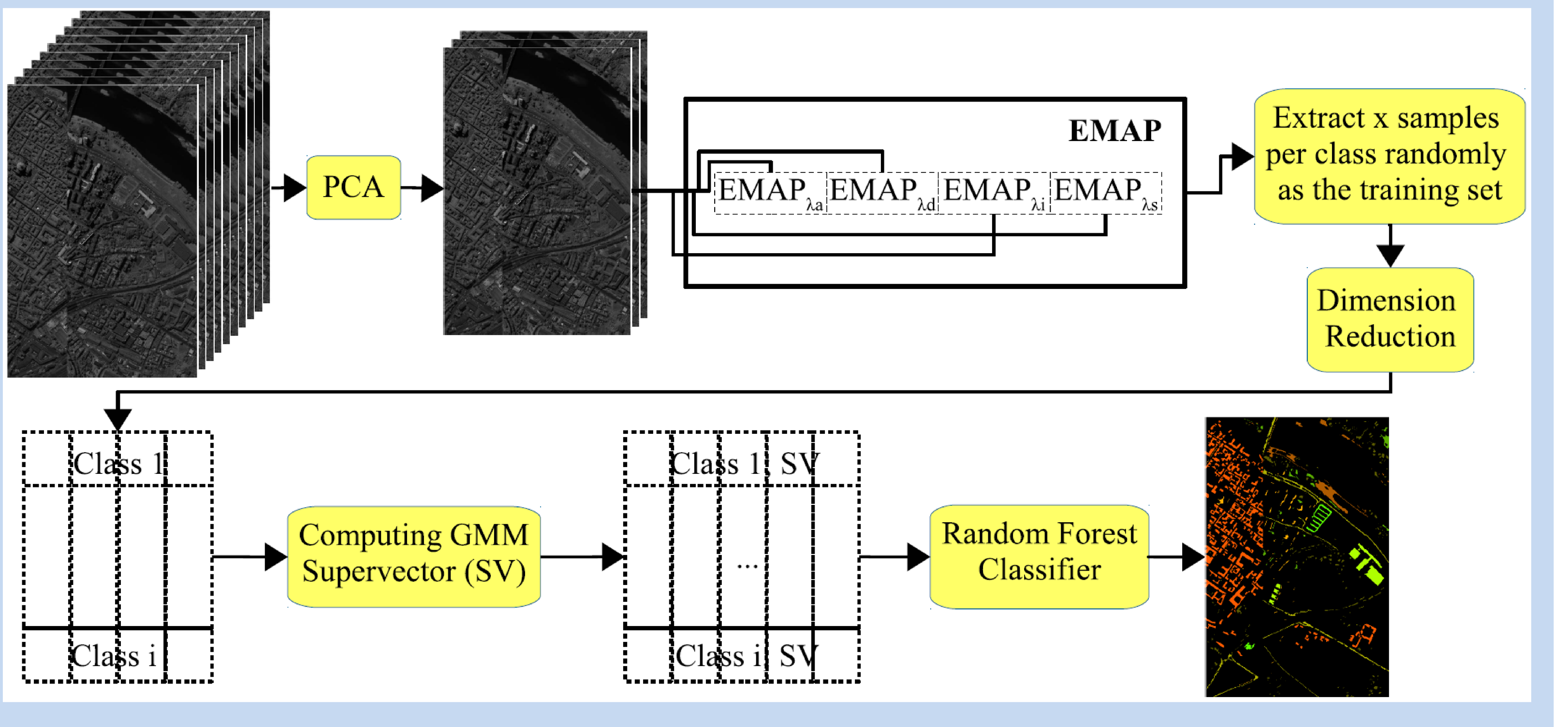

Figure 1: Our proposed HSRS image classification pipeline.

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)

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 training data size of 13 pixels per class. Further results can be found in the paper.

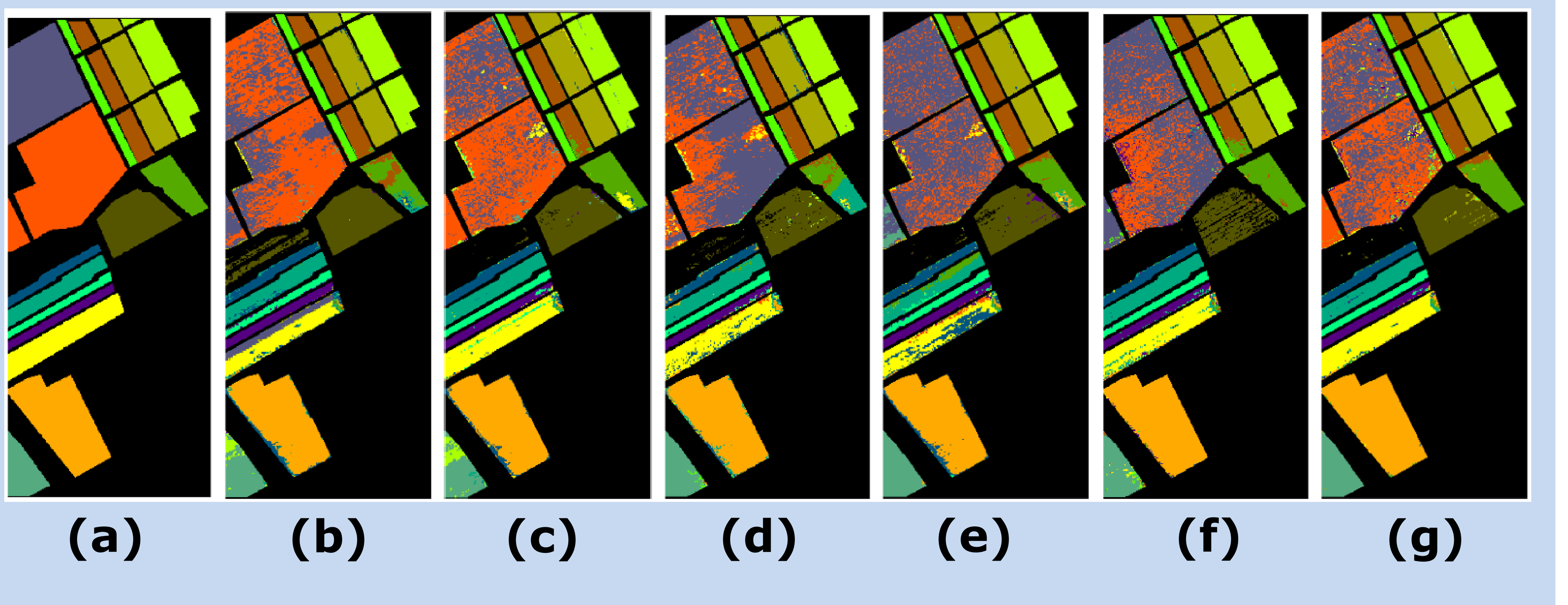

Figure 2: Example label maps on **Salinas valley dataset** using 13 training samples per class. **(a)** ground truth **(b)** EMAP, **(c)** EMAP-SV **(d)** EMAP-PCA, **(e)** EMAP-PCA-SV, **(f)** EMAP-NWFE, **(g)** EMAP-NWFE-SV.

References

[1] Kuo, B.-C., et al., Nonparametric weighted feature extraction for classification, TGRS 42:1096-1105, (2004)

[2] Chi, M. et al., Classification of hyperspectral remote-sensing data with primal svm for small-sized training dataset problem, Advances in space research 41:1793-1799 (2008)

[3] Reynolds, D. A. et al., Speaker Verification Using Adapted Gaussian Mixture Models DSP, 10:19-41, (2000)

[4] Dalla Mura, M. et al., Extended profiles with morphological attribute filters for the analysis of hyperspectral data, JRS 31:5975-5991, (2010)

Contact

amir.davari@fau.de

https://www5.cs.fau.de/en/our-team/davari-amirabbas