Classwise hyperspectral image classification with PerTurbo method

Laetitia Chapel, Thomas Burger, Nicolas Courty, Sébastien Lefèvre

CLASSWISE HYPERSPECTRAL IMAGE CLASSIFICATION WITH PERTURBO METHOD

L. Chapel[1], T. Burger[2], N. Courty[3], S. Lefèvre[3]

[1]Lab-STICC, Université de Bretagne-Sud, 56000 Vannes, FRANCE
[2] iRTSV (FR3425), CNRS, CEA (BGE), INSERM (U1038), Univ. Grenoble, 38000 Grenoble, FRANCE
[3] IRISA, Université de Bretagne-Sud, 56000 Vannes, FRANCE

1. INTRODUCTION

The classification of hyperspectral remote sensing images had been a subject of interest in the past few years, due to the recent advances in remote-sensors technology: hyperspectral data are composed of hundreds of images corresponding to different spectral bands. Classification of such images is a challenging task as it entails processing a huge amount of data that are high dimensional, leading to significant time and memory requirements. From a methodological viewpoint, many classifiers are not appropriate in this context as they suffer the dimensionality curse: the classification accuracy decreases with the dimension of the data when the number of available pixels is fixed. As an example, we can cite the underachievements of gaussian classifiers or neural networks techniques [1, 2]. More recently, Support Vector Machines [3] have received particular attention in this context [4] as they alleviate the dimensionality curse, and it had been shown that they generally outperform traditional classification techniques. Since then, some adaptations to the context have been developed, e.g. kernel function that take into account the spatial neighborhood information [5, 6].

Even though SVM are the state of the art classification technique for remote-sensing images, we would like to consider the use of a new non-parametric classification technique: PerTurbo [7]. It provides a *class-wise* classification: each class is defined by an intrinsic representation based on the Laplace-Beltrami operator approximation [8]. This geometric characterisation allows one to associate to each class a topologic information that describes its spatial distribution in the ambiant space. This information is then used to derive a perturbation measure for each new example. This example is then classified into the class whose perturbation is the smallest. In the paper where PerTurbo was first introduced [7], comparison with SVM is assessed on several examples and it shows similar, sometimes slightly better, performances.

In this work, we aim at investigating the performances of PerTurbo in the remote-sensing context. Indeed, the method possesses appealing characteristics:

- the classification *per class* avoids the use of settings like *one vs all* or *one vs one* in the multiclass formulation; hence there is no need to train several classifiers;
- it is a non-parametric method: there is no explicit formulation of the decision function like in SVM for instance. The method is very simple, easy to implement and involves few parameters to tune.

To assess the effectiveness of the classification method PerTurbo in the remote-sensing context, we run some tests on two classical datasets: an image taken from Airbone ROSIS of the *Pavia University* and of the *Centre of Pavia*, and compare the results with the ones obtained with SVM.

2. PERTURBO CLASSIFICATION APPROACH

We consider a classical machine learning problem, where one seeks for a function that best labels a set of unlabeled examples. We denote $\mathcal{S} = \left\{(\boldsymbol{x}^1, y_1), (\boldsymbol{x}^2, y_2), ..., (\boldsymbol{x}^N, y_N)\right\} \in \mathbb{R}^d \times \{u_1, ..., u_L\}$ the training set, and $\mathcal{S}_\ell$ the set of all the training examples with label u_ℓ belonging to $\mathcal{S}$. The idea to build the predictive function is the following: each $\mathcal{S}_\ell$ is embedded in a dedicated Riemannian manifold $\mathcal{M}_\ell$, whose geometric structure can be expressed in term of the Laplace-Beltrami operator. Despite the fact that it is generally not

possible to find an analytic expression of this operator, it can be approximated by the spectrum of $K(\mathcal{S})$, whose (ith, jth) term is the Gaussian kernel:

$$K_{ij}(\mathcal{S}) = k(\boldsymbol{x}^i, \boldsymbol{x}^j) = \phi(\boldsymbol{x}^i)^T \cdot \phi(\boldsymbol{x}^j) = \exp\left(-\frac{\left\|\boldsymbol{x}^i - \boldsymbol{x}^j\right\|^2}{2\sigma^2}\right), \quad (1)$$

where $\boldsymbol{x}^i$ and $\boldsymbol{x}^j \in \mathcal{S}$, ϕ is the mapping from the original space into the feature space (also called the Reproducing Kernel Hilbert Space RKHS), $\|\cdot\|$ is the Euclidean norm and $\sigma^2 \in \mathbb{R}$ tunes the variance of the Gaussian kernel.

When an example $\widetilde{\boldsymbol{x}}$ from the test set arrives, we compute

$$\tau(\widetilde{\boldsymbol{x}}, \mathcal{M}_\ell) = 1 - k_{\widetilde{\boldsymbol{x}}}^T K(\mathcal{S}_\ell)^{-1} k_{\widetilde{\boldsymbol{x}}}, \quad (2)$$

with $k_{\widetilde{\boldsymbol{x}}} = k(\mathcal{S}_\ell, \widetilde{\boldsymbol{x}})$ whose ith term is $k(\boldsymbol{x}^i, \widetilde{\boldsymbol{x}})$, $\boldsymbol{x}^i \in \mathcal{S}_\ell$, which quantifies the perturbation of the manifold $\mathcal{M}_\ell$ when $\widetilde{\boldsymbol{x}}$ is added to class u_ℓ. Each test sample $\widetilde{\boldsymbol{x}}$ is then classified into the class that provides the smallest perturbation, *i.e.*

$$\arg\min_\ell \tau(\widetilde{\boldsymbol{x}}, \mathcal{M}_\ell). \quad (3)$$

Thus, in some sense, PerTurbo can be seen as a subspace classifier *i.e.* a setting generalizing the principal component analysis, in which each class is modeled by a dedicated subspace of the input space, and in which a new item is classified into the class corresponding to the subspace the distance to which is the closest [9], but instead of working in the input space, the classifier operates in a kernelized space. PerTurbo works as long as the perturbation measure τ is defined, hence as long as $K(\mathcal{S}_\ell)$ is invertible $\forall \ell \leq L$. If $K(\mathcal{S}_\ell)$ is not invertible, it is always possible to compute its pseudo-inverse or to consider regularization techniques that find a close invertible matrix: for instance, in the case of Tikhonov regularization, one considers

$$\widetilde{K}(\mathcal{S}_\ell) = K(\mathcal{S}_\ell) + \alpha_\ell \cdot I \quad (4)$$

where I is the identity matrix, and $\alpha_\ell \in \mathbb{R}_+^*$. The main interest of such regularizations is that they make the Gram matrix (the spectrum of which is equivalent to that of the covariance matrix) less sensitive to outliers. Hence, even in the case where $K(\mathcal{S}_\ell)$ is invertible, it is possible to boost the performances by tuning α_ℓ to a value which is adapted to the covariance of the dataset.

3. PRELIMINARY EXPERIMENTAL RESULTS

In order to assess the performances of PerTurbo on remote sensing images, we use two hyperspectral scenes: *Pavia University* and *Pavia Centre*, both acquired by the ROSIS sensor. The first data set contains 102 spectral bands and is a 1096×715 pixels image; the second one is a 610×340 pixels image and contains 103 spectral bands. Both datasets have nine classes of interest, that are detailed in table 1. We select at random the training pixels and the remaining ones compose the test set.

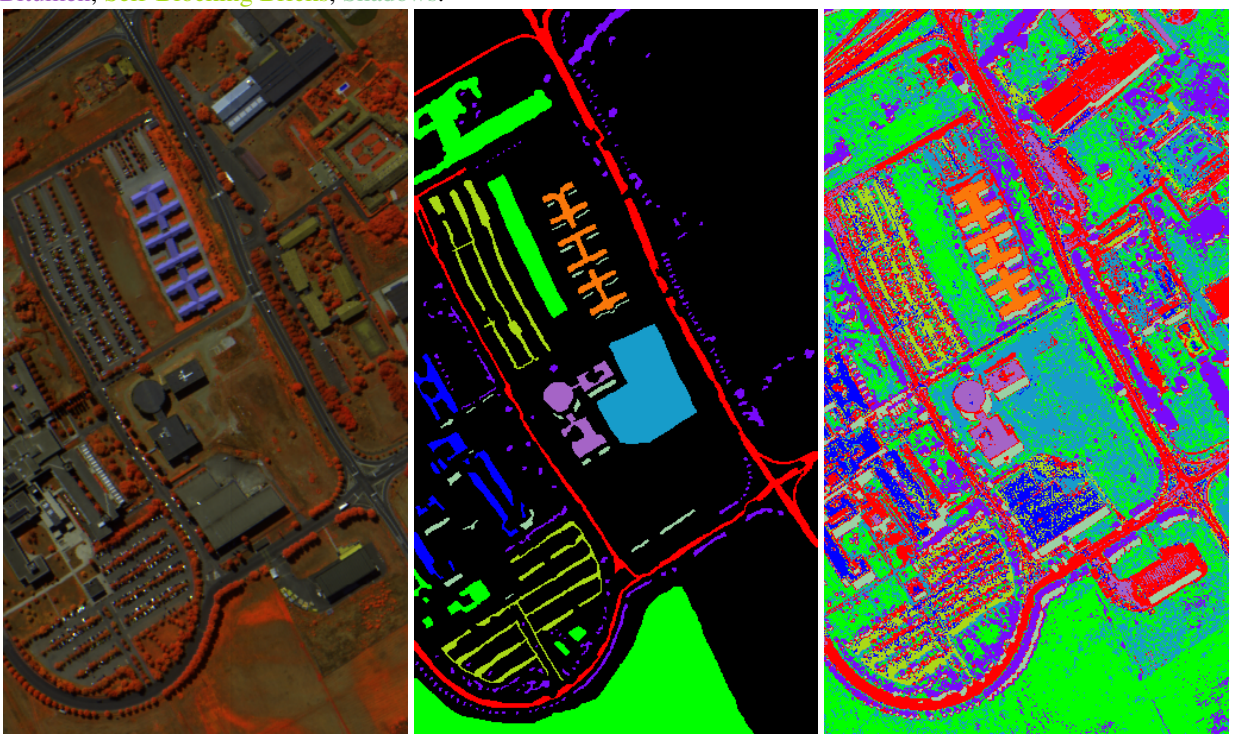
Two parameters need to be tuned for the SVM with Gaussian kernel: the Gaussian kernel width σ and the penalty term C. They were set using five-fold cross validation $\sigma \in \{0.5, 1, 1.5, 2, 3, 4, 5, 6, 10\}$ and $C \in \{1, 5, 10, 200\}$. We use the kernlab implementation for R [10] to run the experiments. For the PerTurbo algorithm, only the σ parameter needs to be tuned. We use a rule of thumb for its choice, coming from the nearest neighbor and spectral clustering community [11]: the minimum over all the classes of the mean distance of the training set points to their k-nearest neighbor, with $k = \log(N) + 1$ and N being the number of points in $\mathcal{S}$. Note that this rule of thumb is very *ad hoc* and data-dependent: accuracies are probably under-estimated as there is no guarantee that the chosen σ gives the best performances. Each experiment is repeated 20 times. Table 1 compares the average (and standard deviation in parenthesis) classification accuracies per class, the overall accuracy (OA) and the average accuracy (AA) obtained with SVM and PerTurbo. Figure 1 shows an example of results of PerTurbo classification with regularization on *Pavia University*.

We note that, for *Pavia Centre* dataset, SVM slightly outperforms PerTurbo but the accuracies are not significantly different. For *Pavia University* dataset, SVM clearly exhibits better performances, for every classes but three, especially for classes *bare soil* and *gravel* that are particularly badly predicted by PerTurbo. Thus, we tried the regularized version of the algorithm, tuning the $\alpha_\ell = \alpha, \forall \ell \in L$ parameter thanks to a logarithmic search (similarly to the C parameter of the SVM, which has roughly the same influence): it clearly improves the results, even if the OA and AA remain lower than SVM. In the near future, it is worth checking if this difference comes from an inappropriate rule of thumb for the sigma tuning (which is not optimized whatever the version

No class	1	2	3	4	5	6	7	8	9		
					PAVIA CENTRE						
Name	Water	Tree	Meadow	Brick	Bare soil	Asphalt	Bitumen	Tile	Shadow		
# train	824	820	824	808	820	816	808	1260	476		
# test	65148	6780	2269	1881	5769	8438	6486	41574	2396	**OA**	**AA**
PerTurbo	99.89	90.27	96.76	94.46	94.65	96.24	89.93	98.65	99.94	98.05	95.65
	(0.06)	(0.40)	(0.34)	(0.74)	(0.32)	(0.45)	(1.09)	(0.10)	(0.03)	(0.07)	
SVM	99.97	95.99	97.04	96.68	96.29	97.61	93.05	99.21	99.97	98.84	97.31
	(0.02)	(0.50)	(0.36)	(0.61)	(0.48)	(0.43)	(0.56)	(0.08)	(0.03)	(0.06)	
					PAVIA UNIVERSITY						
Name	Asphalt	Meadow	Gravel	Tree	Metal sheet	Bare soil	Bitumen	Brick	Shadow		
# train	548	540	392	524	265	532	375	514	231		
# test	6107	18101	1724	2672	1080	4798	816	3142	415	**OA**	**AA**
PerTurbo	76.08	94.53	67.34	95.77	99.44	60.21	95.13	91.45	100	86.24	86.66
(init)	(0.91)	(0.84)	(1.90)	(0.70)	(0.26)	(2.31)	(0.51)	(1.19)	(0.00)	(0.45)	
PerTurbo	81.17	94.24	71.73	95.97	99.43	63.59	93.74	91.66	100	87.48	87.95
(Tikhonov)	(0.01)	(0.00)	(0.02)	(0.01)	(0.00)	(0.02)	(0.01)	(0.01)	(0.00)	(0.27)	
SVM	87.47	93.86	84.23	97.80	99.86	95.19	94.82	92.58	99.93	92.97	93.97
	(0.71)	(0.50)	(1.03)	(0.38)	(0.18)	(0.52)	(0.92)	(0.53)	(0.10)	(0.26)	

Table 1. Information classes, training and test samples and classification accuracies in percentage for datasets *Pavia University* and *Pavia Centre*.

Fig. 1. *Pavia University* dataset. The color code is the following: Asphalt, Meadows, Gravel, Trees, Painted metal sheets, Bare Soil, Bitumen, Self-Blocking Bricks, Shadows.



(a) three-channel color composite (b) available reference data (c) PerTurbo classification results

of PerTurbo) that leads to degraded performances or if a grid search driven on PerTurbo could improve the results. In a similar way, we wonder on the interest of tuning σ and α parameters independently of each class. Also, it is worth noting that the PerTurbo classification procedure uses every samples in the learning database, and as such may be prone to errors in case of outliers or mislabeled samples. To balance these side-effects, two variations of PerTurbo are reported in [7]. The first one is the regularization procedure inspired by sparse learning that is used on *Pavia University*, but which could be generalized to any dataset after a short study of the impact of such regularization. The second one is based on a selection of the most adapted eigenvectors of $K(\mathcal{S})$, in a PCA-like manner (in the RKHS), so that only principal components are kept, leading to reduce the influence of outliers.

4. CONCLUSION

A new classification technique, PerTurbo, has been investigated in the context on hyperspectral remote sensing images context. In this framework, each class is characterised by its Laplace-Beltrami operator, then approximated by the spectrum of $K(\mathcal{S})$, whose terms are derived from the Gaussian kernel. The method is very simple, easy to implement and involves few parameters to tune. It also allows the definition of a simple multi-class strategy, and, as a class-wise classification method, the addition of a new class does not requires the re-training of the pre-existing class models. We conducted experiments on two datasets: results for *Pavia Centre* dataset are encouraging, while results obtained on *Pavia University* show that SVM clearly outperforms PerTurbo. Nevertheless, we believe that this difference comes from a bad parametrization of the algorithm (for which we used a rule of thumb, contrarily to the SVM procedure which was fully optimized). Hence, a systematic search for the optimal value of the parameter would improve the results. Moreover, there are several other possible improvements coming from the fields of regularization methods of dimensionality reduction techniques which let us think that this first experiment is promising. In the near future, we also plan to investigate rules leading to a better choice of σ. We are also interested in studying the behavior of PerTurbo when the classes are heterogeneous with only few training samples available, or when the classes in the training set are highly unbalanced.

5. REFERENCES

[1] G.H. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. on Information Theory*, vol. 14, pp. 55–63, 1968.

[2] K. Fukunaga and R.R. Hayes, "Effects of sample size in classifier design," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 873–885, 1989.

[3] V.N. Vapnik, *Statistical learning theory*, John Wiley & Sons, New York, 1998.

[4] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 42, pp. 1778–1790, 2004.

[5] B. Guo, S. Gunn, R. Damper, and J. Nelson, "Customizing kernel functions for SVM-based hyperspectral image classification," *IEEE Trans. on Image Processing*, vol. 44, pp. 2839–2846, 2008.

[6] M. Fauvel, J. Chanussot, and J.A. Benediktsson, "A spatial-spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognition*, vol. 45, pp. 381–392, 2012.

[7] N. Courty, T. Burger, and L. Johann, "PerTurbo: a new classification algorithm based on the spectrum perturbations of the laplace-beltrami operator," in *ECML/PKDD*, 2011, vol. 1, pp. 359–374.

[8] I. Chavel, *Eigenvalues in Riemannian geometry*, Academic Press, Orlando, 1984.

[9] H. Cevikalp, D. Larlus, M. Neamtu, B. Triggs, and F. Jurie, "Manifold based local classifiers: Linear and nonlinear approaches," *Journal of Signal Processing Systems*, vol. 61, no. 1, pp. 61–73, 2010.

[10] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, "kernlab – An S4 package for kernel methods in R," *Journal of Statistical Software*, vol. 11, no. 9, pp. 1–20, 2004.

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