A NOVEL APPROACH TO THE SELECTION OF SPATIALLY INVARIANT FEATURES FOR CLASSIFICATION OF HYPERSPECTRAL IMAGES

Claudio Persello and Lorenzo Bruzzone

Department of Information Engineering and Computer Science, University of Trento
Via Sommarive, 14 I-38100, Povo, Trento, Italy, Fax: +39-0461-882093, e-mail: claudio.persello@disi.unitn.it

1. INTRODUCTION AND BACKGROUND

Hyperspectral remote sensing images represent a very rich source of information for the analysis and automatic recognition of the land-cover classes. However, supervised classification of hyperspectral images is a very complex methodological problem due to many different issues [1]: i) the small value of the ratio between the number of training samples and the number of available spectral channels (and thus of classifier parameters), which results in the Hughes phenomenon; ii) the high correlation among training patterns taken from the same area, which violates the required assumption of independence of samples included in the training set (thus reducing the information conveyed to the classification algorithm by the considered samples); iii) the non-stationary behavior of the spectral signatures of land-cover classes in the spatial domain of the scene, which is due to physical factors related to ground (e.g., different soil moisture or composition), vegetation, and atmospheric conditions. In order to address the abovementioned problems, in the recent literature different promising approaches have been proposed for hyperspectral image classification. Among the others, we recall: i) the use of supervised kernel methods (and in particular of Support Vector Machines), which are intrinsically robust to the Hughes phenomenon [1]; ii) the use of semisupervised learning methods that take into account both labeled and unlabeled samples in the learning of the classifier [2]; and iii) the joint use of kernel methods and semisupervised techniques [3]. Nevertheless, the problem of the spatial variability of the features can be addressed (together with the sample size problem) at a different and complementary level, i.e., in the feature selection phase.

In this paper we address the aforementioned problem by proposing a novel approach to feature-selection that aims at identifying a subset of features that exhibit both high discrimination ability among the considered classes and high invariance in the spatial domain of the investigated scene. The proposed approach can be integrated in the design of any system for hyperspectral image classification for increasing the robustness and the generalization capability of the classifier.

2. PROPOSED METHODOLOGY

The novel contribution of this paper consists in the idea of explicitly considering the stationary properties of features in the investigated scene in the phase of feature selection. We propose a novel approach to feature selection (and the related methodological implementation), which is based on the idea that the features selected to be given as input to a classifier should exhibit two fundamental properties: i) they should be able to discriminate the information classes present in the considered problem as much as possible; ii) they should exhibit a high invariance in modeling each specific land-cover class on the analyzed scene. This approach is implemented by defining a novel criterion function that is made up of two terms: i) a standard separability measure, and ii) a novel invariance measure that assesses the stationarity of features in the spatial domain. The search algorithm, adopted for deriving the subsets of features that jointly optimize the two terms, is based on the optimization of a multi-objective problem for the estimation of the Pareto optimal solutions. For the assessment of the two terms of the criterion function we propose both a supervised and a semisupervised method that can be adopted according to the amount of available reference data.

2.1. Formulation of the proposed supervised criterion function

Let us consider a feature-selection problem for the classification of a hyperspectral image I, where each pixel, described by a feature vector $\mathbf{x} = (x_1, x_2, ..., x_n)$ in an n-dimensional feature space, is to be assigned to one of C different classes $\Omega = \{\omega_1, \omega_2, ..., \omega_C\}$. Let us further assume that the available training set T is made up of two subsets of labeled patterns T_1 and T_2 collected on disjoint (separate) areas on the ground. This property of the training set is exploited for assessing the spatial variability of the spectral signatures of the land-cover classes. We successively relax this hypothesis by proposing a semisupervised method that does not require the availability of a training subset T_2 spatially disjoint from T_1 and takes advantage of unlabeled samples. The proposed technique aims to select the subset $\boldsymbol{\theta}$ of m features (with $m<n$) that optimizes a novel criterion function based on two different terms:

a) *Discrimination term* Δ - this term is based on a standard feature-selection criterion function. In the proposed system we adopt the following definition: $\Delta(\boldsymbol{\theta}) = \sum_{i=1}^{C}\sum_{j>i}^{C} P(\omega_i)P(\omega_j)S_{ij}(\boldsymbol{\theta})$, where $S_{ij}(\boldsymbol{\theta})$ is a statistical distance measure (e.g., Bhattacharyya distance, Divergence, Jeffries-Matusita distance) between the distributions $p(\mathbf{x} \mid \omega_i)$ and $p(\mathbf{x} \mid \omega_j)$ of two classes ω_i and ω_j, respectively, and $P(\omega_i)$, $P(\omega_j)$ are the prior probabilities of the classes ω_i and ω_j in the considered scene. The subset of *m* features that maximizes this distance results in the best potential for discriminating land-cover classes.

b) *Invariance term* P - this term, which represents an important novel contribution of this paper, explicitly measures the invariance (stationary behavior) of features on each class in the investigated image. It can be defined as: $\mathrm{P}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{C} P^{T_1}(\omega_i)P^{T_2}(\omega_i)S_{ii}^{T_1T_2}(\boldsymbol{\theta})$ where $S_{ii}^{T_1T_2}$ is a statistical distance measure between the distributions $p^{T_r}(\mathbf{x} \mid \omega_i)$, $r = 1,2$ of the class ω_i computed on T_1 and T_2, and $P^{T_r}(\omega_i)$ represents the prior probability of the class ω_i in T_r, $r = 1,2$. We expect that a robust subset of features should minimize the value of $\mathrm{P}(\boldsymbol{\theta})$.

2.2. Semisupervised evaluation of the criterion function (invariance term estimation)

The collection of labeled training samples on two spatially-disjoint areas from the site under investigation can be difficult and/or very expensive. This may compromise the applicability of the proposed supervised method in some real classification applications. In order to overcome this possible problem, we propose a semisupervised technique to estimate the invariance term, which does not require the availability of a disjoint training subset T_2. This proposed method is based on the estimation, with the expectation maximization algorithm, of the distributions of classes in portions of the image separated from T_1. The estimation is carried out by exploiting the information captured from a set of unlabeled pixels.

2.3. Formulation of the proposed search algorithm

In order to select the final subset of features, we modeled our search problem as a multi-objective minimization problem, where the multi-objective function is made up of two different (and possibly conflicting) objectives, which express the discrimination ability Δ among the considered classes and the spatial invariance P of the subset of features, respectively. This problem is solved in order to obtain a set of Pareto optimal solutions, instead of a single optimal one. In our implementation we adopted a multi-objective genetic algorithm for the estimation of the Pareto front. The main advantage of this approach is that it avoids to aggregate metrics capturing multiple objectives into a single measure, but allows one to effectively identify different possible tradeoffs between values of Δ and P and to select the subset of features that simultaneously exhibits both properties.

3. EXPERIMENTAL RESULTS

In order to assess the effectiveness of the presented approach (with both the proposed supervised and semisupervised methods), we carried out several experiments on a hyperspectral image acquired over an extended geographical area. We considered a data set acquired by the Hyperion sensor of the EO-1 satellite in an area of the Okavango Delta, Botswana. Experimental results confirmed that the selection of robust features carried out with the proposed strategy involved a sharp increase in the accuracy and the generalization ability of the system with respect to the use of features selected according to standard strategies. (for space constraints the numerical results will be reported in the full paper).

REFERENCES

[1] F. Melgani, L. Bruzzone, “Classification of hyperspectral remote-sensing images with support vector machines”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42, No. 8, August 2004, pp. 1778-1790.

[2] B. M. Shahshahani and D. A. Landgrebe, “The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon,” *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 32, No. 5, pp. 1087–1095, September 1994.

[3] L. Bruzzone, M. Chi, M. Marconcini, “A Novel Transductive SVM for the Semisupervised Classification of Remote-Sensing Images”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 44, No. 11, 2006, pp. 3363-3373.