A Classification Algorithm for Hyperspectral Images based on Synergetics Theory

Daniele Cerra, Rupert Mueller, and Peter Reinartz

Abstract

This paper presents a classification methodology for hyperspectral data based on synergetics theory. Pattern recognition algorithms based on synergetics have been applied to images in the spatial domain with limited success in the past, given their dependence on the rotation, shifting and scaling of the images. These drawbacks can be discarded if such methods are applied to data acquired by a hyperspectral sensor in the spectral domain, as each single spectrum, related to an image element in the hyperspectral scene, can be analyzed independently. The spectrum is first projected in a space spanned by a set of user-defined prototype vectors, which belong to some classes of interest, and then attracted by a final state associated to a prototype. The spectrum can thus be classified, establishing a first attempt at performing a pixel-wise image classification using notions derived from synergetics. As typical synergetics-based systems have the drawback of a rigid training step, we introduce a new procedure which allows the selection of a training area for each class of interest, used to weight the prototype vectors through attention parameters and to produce a more accurate classification map through plurality vote of independent classifications. As each classification is in principle obtained on the basis of a single training sample per class, the proposed technique could be particularly effective in tasks where only a small training dataset is available. The results presented are promising and often outperform state of the art classification methodologies, both general and specific to hyperspectral data.

Index Terms

Hyperspectral image analysis, image classification, least squares approximation (LS), synergetics theory.

I. Introduction

Hyperspectral data are characterized by very rich spectral information, and as a consequence have strong discrimination power in detecting targets of interest. On the other hand, the very high

dimensionality of these data introduces several problems, summarized by the principle known as curse of dimensionality [1]. Very often, not all the bands are useful for a given application. As a consequence, band selection can be performed. Alternatively, the data can be projected on a lower-dimensionality space to aid data exploration or improve computation performances [2]. This is usually a preprocessing step aiding other operations such as classification, which is a task often constituting both the final objective and validation step of dimensionality reduction methodologies [3]. One of the most widely used among such techniques in remote sensing is the principal component analysis (PCA). PCA computes orthogonal projections that maximize the amount of data variance, and yields a dataset in a new uncorrelated coordinate system [4]. If the user desires to differentiate different classes of interest, however, such approach may not be optimal, as in general the dimensions in the subspace do not convey any semantics. Therefore, these may not match the user's needs, as information regarded as important for a given application may be considered secondary by the system, and thus discarded in the process.

This paper introduces a classification methodology for hyperspectral data based on synergetics theory, in which the subspace on which the data are projected is defined by the user. Synergetics is a four decade old theory describing the spontaneous formation of patterns and structures in a system through self-organization. Algorithms based on synergetics have been applied to pattern recognition in images, but they have often been limited by their dependency on scaling, rotation and shifting of the images [5]. These drawbacks would not affect applications to hyperspectral data performed in the spectral domain, as each image element, representing a spectrum, can be analyzed independently. Each spectrum is projected in a subspace composed by a set of user-defined prototype vectors, belonging to some classes of interest: this can be regarded as a semantic space, as the value of the test vector in each dimension quantifies the similarity to a given class of interest. The spectrum may then be represented as a particle on a potential surface, built as a manifold in this subspace, and attracted by one of several possible final states, with each one being associated to a user-defined class, and hence classified. The proposed approach can only be applied to overdetermined systems, i.e. it requires the data dimensionality to be much higher than the number of classes of interest, therefore hyperspectral data enable for the first time a pixel-wise classification methodology derived from these notions.

As typical synergetics-based systems have the drawback of a rigid training step, we modify it to allow the selection of user-defined training areas. In a first step, several independent

classifications are carried out on the basis of a single training sample per class, while the other samples belonging to the training set are used to weight the prototype vectors through attention parameters. Results are further improved by manipulating the data directly in the space spanned by the user-defined prototypes, where each image element is displaced towards the pixels which are its neighbours both in the image space and in the prototype vectors space. This increases the probability of a pixel to be attracted by the stable state related to the class to which the majority of its adjacent pixels belongs, leading to a more homogeneous data representation, and follows recent years' trend of exploiting contextual information to improve the results of pixel-wise classification in hypespectral data [3]. A final classification map is produced through majority vote of the independent classifications. The operations carried out in the prototype vectors space implicitly take into account the intraclass dependencies and similarities, and the results obtained are comparable to state of the art classification methodologies, both general and specific to hyperspectral data. The proposed technique is, for a given application, well capable of handling high-dimensional data, as the synergetics principle drastically reduces the degree of freedoms in a system. This step is represented by the transformation of a dataset from N to M dimensions, with $M << N$, where N is the original number of spectral bands in the image, and M is the number of classes of interest as selected by the user.

The proposed technique is able to exploit the same set of training samples twice: not only for the classification step, but also to iteratively improve the independent classifications' results. Therefore, it is able to produce satisfactory results even if a limited number of training samples is available, which is a desired property in recent hyperspectral analysis methods [6]. Furthermore, experiments suggest that this leads to a better discrimination between classes exhibiting similar spectral features with respect to traditional classification techniques.

The paper is structured as follows. Section II introduces the main ideas to the base of the synergetics theory, and illustrates how a pattern-recognition system can be defined according to this theory, highlighting its similarities and correspondences with the Least Squares (LS) approach for the approximate solution of overdetermined systems. Section III illustrates how the synergetics principles can be applied to hyperspectral data, to reduce dimensionality and perform pixel-wise classification, and analyzes the relation between synergetics and established concepts in estimation theory and data processing. Subsequently, different operations are carried out directly in a user-defined prototype vectors space to improve the initial results, and a final

classification is derived by plurality vote. Section IV reports experiments on the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) Salinas dataset and comparisons to state-of-the-art methods. We conclude in Section V with some final considerations and an outlook on future work.

II. SYNERGETICS THEORY

Synergetics is an interdisciplinary science originally founded by Hermann Haken in 1969 [7]. The synergetics theory tries to find general rules for the formation of patterns through self-organization, as new structures or processes spontaneously arise in macroscopic systems. Such rules should be valid for large classes of systems, being these composed of atoms, molecules, neurons, individuals, or image elements. The term synergetics derives from the Greek "working together", indicating the cooperation of different parts in a system or different systems. The fields of applications of synergetics range from biology to ecology, chemistry, cosmology, thermodynamics and up to sociology: countless self-organization phenomena have been explained through synergetics, including the formation of laser light and the origin of galaxies structures [8].

Although this theory originates from pattern formation, Haken links it to pattern recognition, by regarding the latter as "a sequence of symmetry-breaking events, where at each branching point new information is needed to break the symmetry, i.e., to make a unique decision possible" [9]. Such branching points are related to the reduced degrees of freedoms in a pattern formation process based on synergetics, as these obey to an enslaving principle related to some order parameter, which drastically reduces the degrees of freedom in the system. Henceforth we will restrict our analysis to pattern recognition and image analysis methods based on this theory, of which the first example is described in [5].

From here onwards, matrices are represented as bold upper case ($\mathbf{A}$), column vectors as bold lower case ($\mathbf{a}$), while all other quantities are scalar (a).

In the first step of a typical synergetics-based pattern recognition system the user selects some prototype patterns, each of which corresponds to a class of interest [5]. Let $\mathbf{v}'_k \in R^N$, $k = 1, ..., M$ be such prototype vectors (or classes) formed by N-dimensional real valued components, normalized by:

$$\mathbf{v}_k = \frac{\mathbf{v}'_k}{|\mathbf{v}'_k|} \tag{1}$$

so that $|\mathbf{v}_k| = 1$. This set of prototypes spans the subspace $\Omega_c \subset R^N$ with $M < N$. As the prototype vectors do not build in general an orthogonal set of basis vectors for Ω_c, a dual space having the same dimensionality M is constructed using the contravariant (or adjoint) basis vectors $\mathbf{v}^j \in R^N$,$j = 1, ..., M$ using the relation

$$\mathbf{v}^j = \sum_{k=1}^{M} g^{jk} \cdot \mathbf{v}_k \tag{2}$$

A unique set of contravariant prototype vectors can be derived employing the orthonormality relation $< \mathbf{v}^j, \mathbf{v}_k >= \delta_k^i$ with $< ., . >$ being the scalar product and δ_k^j the Kronecker delta. This leads to the metric tensor $\mathbf{G} \equiv (g^{jk}) = \mathbf{V}^{-1}$ with $\mathbf{V} = (< \mathbf{v}_j, \mathbf{v}_k >)$ the matrix of the scalar products of the prototype vectors. If an arbitrary feature vector $\mathbf{v} \in R^N$ belonging to an unknown class is then presented to the system, it can be expressed by

$$\mathbf{v} = \sum_{k=1}^{M} q^k \mathbf{v}_k + \mathbf{r}, \tag{3}$$

where the sum consists of the linear combination of the prototype patterns, and the remaining residual vector $\mathbf{r}$. The coefficients $q^k =< \mathbf{v}^k, \mathbf{v} >$ are the projection of the vector $\mathbf{v}$ onto the contravariant basis vectors $\mathbf{v}^k$ and are also called *order parameters* in the parlance of synergetics.

In synergetics theory an energy function is established with local energy minima for each of the prototype vectors. A prototype generates two symmetric minima but we consider here only the positive values. In order to establish a dynamic system the following energy function is defined:

$$E(q^1, ..., q^M) = -\frac{1}{2} \sum_{k=1}^{M} \lambda_k \cdot (q^k)^2 + \frac{1}{4} \sum_{k=1_{k \neq j}}^{M} \sum_{j=1}^{M} B_{kj} (q^k)^2 \cdot (q^j)^2 + \frac{1}{4} C \left(\sum_{k=1}^{M} (q^k)^2 + |\mathbf{r}|^2 \right)^2, \tag{4}$$

Where B_{kj} and C are positive constants (e.g. $B_{kj} = 1, \forall k, j = 1, ..., M$), and $C = 1$), and $\{\lambda_1, ...\lambda_M\}$ are positive values, also called attention parameters. Thus, local energy minima for each of the prototypes are formed (see figure 1). The first term generates minima along the prototype vectors, while the second term discriminates the prototypes in the landscape of the energy function, and the third is the saturation term, enveloping the energy system.

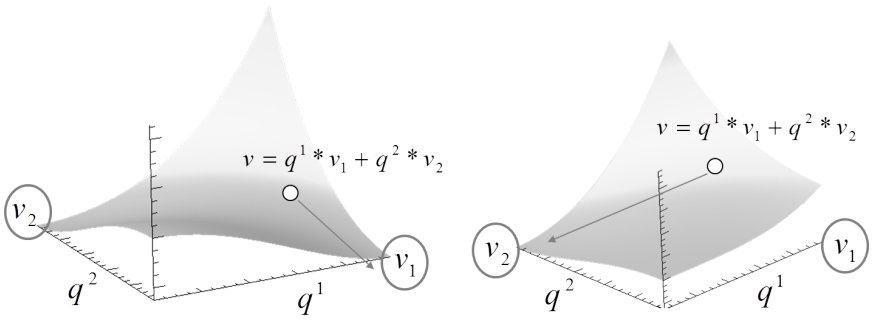


Fig. 1. Energy function $E(q^1, q^2)$, for two prototypes $\mathbf{v}_1$, $\mathbf{v}_2$ in the order parameter domain and behaviour of a test vector $\mathbf{v}$ projected on the plane in the parameter domain. The vector $\mathbf{v} = q^1\mathbf{v}_1 + q^2\mathbf{v}_2$ is attracted towards a stable focus, corresponding to a final state for the vectors v_1 or v_2 (circled in the figure). The final state in the two images differs as the attention parameters are changed. In the second image the parameter λ_{q^1}, associated to the vector $\mathbf{v}_1$, has been drastically decreased, and the test vector is then attracted by the final state associated to the vector $\mathbf{v}_2$.

The basic equation of synergetics for pattern recognition as formulated in eq. (4) describes the time evolution of the feature vector $\mathbf{v}(t)$. A coupled differential equation defines a dynamic system by describing the time evolution of the *order parameters* in the adjoint space built from M prototype vectors. Loosely speaking, the feature vector $\mathbf{v}(t)$ is moving in the landscape of energy towards a unique final state for $\mathbf{v}(t \to \infty)$:

$$\frac{d}{dt}\begin{pmatrix} q^1 \\ \vdots \\ q^n \end{pmatrix} = - \bigtriangledown_{q^j} \cdot E(q^1 \dots q^n). \tag{5}$$

Therefore, a pattern-formation process takes place as the initial pattern is pulled into one of the possible final states, each of which is linked to a prototype vector. The input is then assigned to the class of interest represented by the chosen prototype.

To better understand the evolution of a test vector in the parameter space and its relation with the final states associated to each prototype vector, consider the example in Fig. 1. The surface is a 3D-representation of the energy function $E(q^1, q^2)$, related to two prototypes $\mathbf{v}_1$ and $\mathbf{v}_2$ in the order parameter domain. An unknown test vector, expressed as a linear combination of the prototype vectors, is represented by a point projected on the potential surface in the parameter domain. The test vector is attracted towards a stable focus, corresponding to a final state for the vectors $\mathbf{v}_1$ or $\mathbf{v}_2$. If the attention parameters λ_{q^1} and λ_{q^2} are modified, the final state attracting

the test vector may differ: in the left image the attention parameters are set to $\lambda_{q^1} = \lambda_{q^2} = 1.0$ and the test vector is attracted to the final state $\mathbf{v}_1$, while in the image to the right λ_{q^1} is set to 0.5 while λ_{q^2} remains unchanged, and the test vector is attracted by the stable focus related to $\mathbf{v}_2$.

Haken himself applied his synergetics-based pattern recognition algorithm to image labeling, in the specific to face recognition and sketch categorization. He proposes to form the prototype vectors space with n user-selected gray-scale images of size $X \times Y$, which are converted to 1-dimensional vectors of size $X \times Y$ in a first step. Afterwards, a pattern-formation process takes place, in which an unknown image is presented to the system, and classified as it is pulled in the final state related to the most similar prototype. As each dimension of the prototype and test vectors is linked to the gray value of a single pixel, such system only works if the patterns have the same size, orientation and location in the test and retrieved images. To overcome this problem, the authors propose a similar algorithm based on Fourier analysis, making the patterns independent from rotation and shifting to some degree, but with limited success [5]. After the first one, numerous pattern recognition algorithms based on synergetics have been described. Applications on image classification have been proposed, among others, in [10], [11], [12], [13], and [14]. In [15] the authors perform 3D reconstruction of buildings. In the remote sensing field, an effort has been done to classify Synthetic Aperture Radar (SAR) image tiles and binary images derived from SAR scenes in [16]. Such notions have been popular especially in the 90's, with following years witnessing an interest decrease, probably due to the rigid training step and the great dependance on scale, rotation and shift typical of such methods.

A. Synergetics Theory and Least Squares

We will now show that the *order parameters* are closely related to the Least Squares (LS) approach for the approximate solution of overdetermined systems. Let $\mathbf{A} = (\mathbf{v}_1, ..., \mathbf{v}_M)$ be the matrix composed by the column vectors of the prototypes and $\hat{\mathbf{p}}_{OP} = (q^1, ..., q^M)^T = (< \mathbf{v}^1, \mathbf{v} >, ..., < \mathbf{v}^M, \mathbf{v} >)^T$ be the vector of the *order parameters*. Using the definition of the matrix $\mathbf{V} = \mathbf{A}^T \cdot \mathbf{A}$ and equation (2) for the row vectors $\mathbf{v}^j$ the following equation holds:

$$\hat{\mathbf{p}}_{OP} = \mathbf{G} \cdot \mathbf{A}^T \cdot \mathbf{v} = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{v}. \tag{6}$$

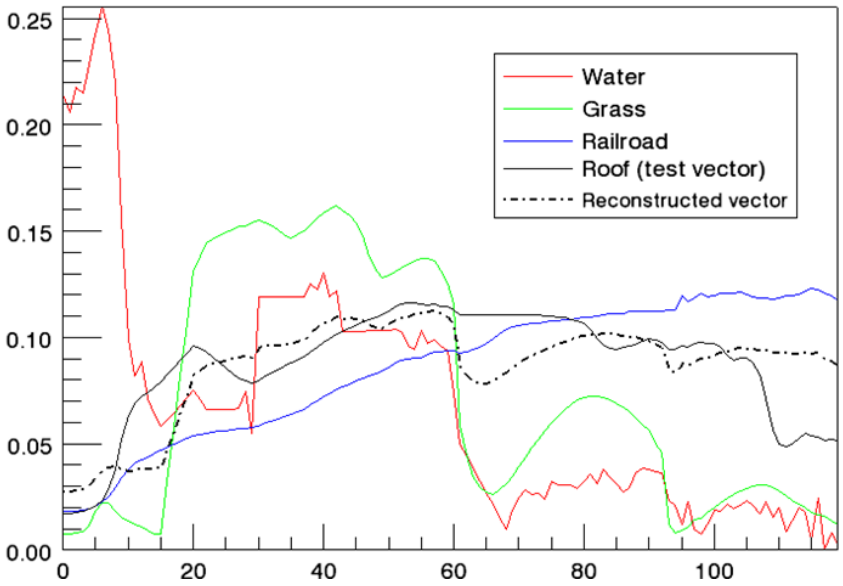


Fig. 2. Prototype vectors as normalized spectra collected from a HyMap scene (water, grass, and railroad), plus reconstruction of a test vector (roof) as a linear combination of the prototype vectors.

The *order parameters* are therefore the least squares solution for $\mathbf{p}_{OP}$ of the linear equation $\mathbf{v} = \mathbf{A} \cdot \mathbf{p}_{OP} + \mathbf{r}$ by minimizing the residuals $\mathbf{r}^T \cdot \mathbf{r} \rightarrow \min$ (note: the *hat* of $\hat{\mathbf{p}}_{OP}$ denotes the best solution in the sense of least squares). Equation (3) can be now rewritten as:

$$\mathbf{v} = \mathbf{A} \cdot \hat{\mathbf{p}}_{OP} + \mathbf{r} = \hat{\mathbf{v}} + \mathbf{r}, \tag{7}$$

where $\hat{\mathbf{v}} \in \Omega_c$ is the orthogonal projection of the feature vector $\mathbf{v}$ onto the subspace spanned by the prototype vectors and is given by the $\hat{\mathbf{v}} = \mathbf{P} \cdot \mathbf{v}$ with the symmetric and idempotent projection operator $\mathbf{P} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$. The vector $\mathbf{r} = (\mathbf{I} - \mathbf{P}) \cdot \mathbf{v}$, with $\mathbf{I}$ representing the unit matrix, denotes the residual vector which is orthogonal to the subspace Ω_c. It has to be remarked that each Hilbert space can be uniquely decomposed by orthogonal sub-spaces. To the best of our knowledge, it is the first time that such correspondence between LS and synergetics theory are made explicit.

III. Classification of Hyperspectral Data based on Synergetics Theory

For hyperspectral data the synergetics approach combines several characteristics typical of different well-known methods such as Spectral Angle Mapper (SAM) [17], Orthogonal Subspace

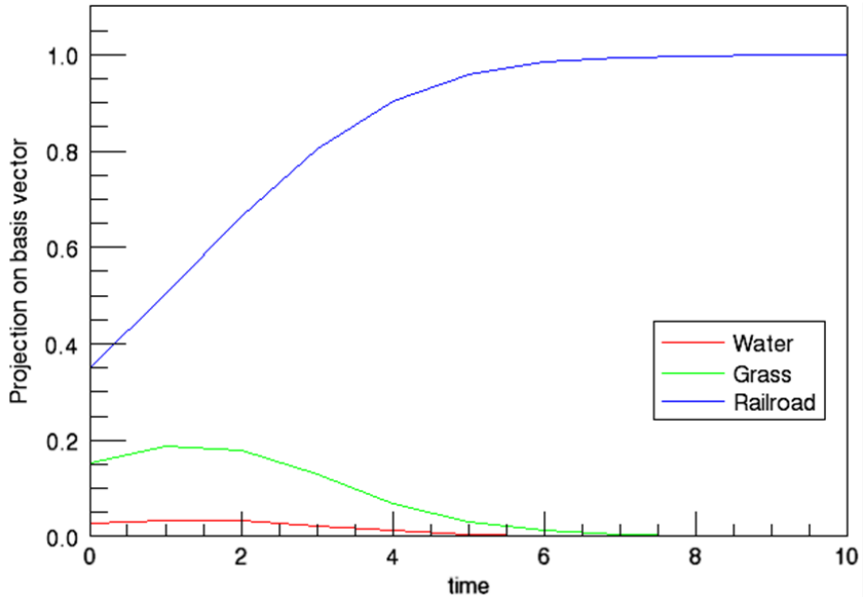


Fig. 3. Simulation of the synergetics process for the test vector in Fig. 2. At $t = 0$ the test vector is projected into the space spanned by the three prototype vectors, and can be reconstructed as in Fig. 2. The test vector is then attracted by one stable final state, corresponding to one of the basis vectors, which is selected as the winner.

Projection (OSP) [18], and spectral unmixing techiques [19], and can be derived from the description in Section II as follows.

Firstly, a set of N-dimensional real valued components derived from M spectral signatures is chosen to build an adjoint space $\mathbf{v}^k$, $k = 1, ..., M$, as in eq. (2). The normalization of the spectra suppresses illumination influences, as only the direction of the prototypes in the feature space is used, resembling in this a desirable property typical of the SAM method. A spectrum $\mathbf{s}$ belonging to an unknown class is then chosen as a test vector $\mathbf{v}$ in eq. (3), and represented as a linear combination of the prototype patterns. This resembles OSP which, based on the criteria of least squares minimization (see also Section II-A), finds an optimal representation of the feature vectors in terms of their projections onto the subspace of prototype vectors, by minimizing the length of the residual vector $\mathbf{r}$. In a similar way to spectral unmixing techniques, the projected vector $\mathbf{s}$ is expanded in the subspace of the adjoint prototypes $\mathbf{v}^k$. Its representation is given in terms of the order parameters q^k, which are abundance values related to the composition of $\mathbf{s}$ in terms of the prototype vectors.

Finally, the evolution in time of $\mathbf{s}$ is tracked, as the test spectrum is pulled towards one of

the M possible final states, in a similar way to the example sketched in Fig. 1. The synergetics principle [5] ensures that every prototype has an associated final state, and that no final states exist other than the ones associated to each vector in the adjoint vector space. The potential of the test vector at time $t = 0$ is equal to the projection of the test vector on each prototype vector, i.e. to the coefficient of each prototype vector resulting in the linear combination yielding a reconstruction of the kind sketched in Fig. 2. At time $t = \infty$ the test vector becomes a scaled version of the winning prototype.

We illustrate this procedure through an example. We start by selecting three 128-dimensional spectra selected from a airborne scene acquired by the HyMap sensor (HyVista Corporation) over Munich, Germany. The spectra $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are related to areas on ground containing water, grass, and railroad respectively. They have been chosen in order to be as pure as possible, and span an area on ground of approximately 4m X 4m. An additional spectrum related to a roof is then chosen as test vector $\mathbf{s}$. We greatly reduce the dimension of the system by building an n-dimensional space which uses as basis the projections on the adjoint prototypes $\mathbf{v}^i$, with $i \in 1, 2, 3$. The potential function will be modeled as a hyperplane in 4 dimensions, with the fourth dimension being the value of the potential in the 3d space. We can then represent the test vector as a linear combination of the prototype adjoint vectors: $\mathbf{s} = q^1\mathbf{v}_1 + q^2\mathbf{v}_2 + q^3\mathbf{v}_3 + \mathbf{r}$, with $\mathbf{r}$ the residual vector (Fig. 2). Afterwards, the coefficients of the linear combination are used to resolve the energy function in eq. (4), with the parameters $\lambda_k = B = C = 1$. We can observe in Fig. 3 the evolution of the potential function and of the retrieved prototype pattern, corresponding to the class "railroad". The final state for the test vector employed coincides with our expectations, as human-made objects spectra are often more similar to each other, rather than to natural objects: we expect then the class "railroad" to prevail over the "water" and "grass" classes. This example can easily be extended to the classification of a hyperspectral image, by building once a prototype vectors space as described above, and subsequently labeling each image element with one of the classes of interest, after resolving equation (5).

Typical algorithms based on synergetics theory for pattern matching, as in the example above, need to solve differential equations to estimate the dynamics of the test vector after being projected in the prototype vectors space. This makes difficult to apply such methods in real applications. Haken shows in [5] that the order parameter with the highest value at time $t = 0$ is related to the prototype that will be chosen by the system as winning final state, while all

others will eventually decay and assume a value of 0, if the attention parameters remain stable under certain limits. Based on these observations, many systems based on synergetics theory use approximations to avoid computing the differential equation (3), usually by selecting the largest initial order parameter [20], [14], [21], [22], [11], [13].

From here on we approximate then the synergetics equation (4) by its first term, which generates minima along the prototype vectors, only considered at time $t = 0$. It has to be remarked that the non-linear terms in the full potential equation (4) should be investigated in the future, along with a way of selecting the best values for the parameters B and C in the equation. The first non-linear term weighted by B represents the interactions between the chosen prototype vectors, while the second weighted by C is a general saturation component quantifying how fast does the projected test vector move in the prototype space. Therefore, the higher abundance value defines the classification result, and a class C_k is chosen among M possible classes for a test vector $\mathbf{v}$ with associated *order parameters* q^k, if

$$k = arg\max_k q^k. \tag{8}$$

The workflow detailed in the next paragraphs is summarized by the sketch in Fig.4.

A. Enabling the selection of a training area

For a pattern recognition system such as the one described in [5], the training step is quite problematic. As each training sample becomes a dimension in the adjoint vectors space, a classification in such space is strongly dependent on the selection of the base vectors, and such systems do not allow selecting different training samples for the same class. A spectrum averaged over a small, homogeneous area can reduce this dependance to some degree, but does not take into account intra-class variations. We could instead assign several samples to the same class of interest, but this would result in an over-determined adjoint vector space derived from a set of basis vectors with strong similarities between them. Such vectors form an ill-conditioned matrix, which is necessary to invert in eq. (2). This would introduce non-negligible numerical errors.

To cope with this problem we propose a classification procedure for an image H as follows. In the first step, for each class $C_i \in H$, with $i \in 1...M$, n samples are selected. Then, n

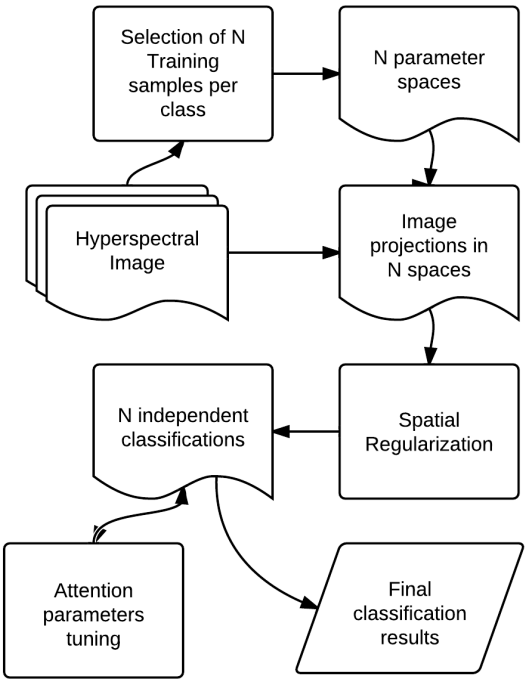


Fig. 4. Classification procedure. After the selection of N training samples per class as prototypes, N prototype spaces are created. The image is projected on each of such spaces, and each projection is independently classified after a step of spatial regularization. Each classification is then improved through an automatic tuning of the attention parameters related to the classes of interest. The final classification is derived by plurality vote.

classifications are performed, in each of which a different training sample for each class is selected. Afterwards, each pixel $p(x, y) \in H$ at coordinates (x, y) is assigned to a given class according to a plurality vote, which selects class C_k if:

$$k = arg\max_j \sum_{i=1}^{n} D(i, j), \tag{9}$$

where $D(i, j) \in 0, 1$, with $D(i, j) = 1$ if the classification i chooses C_j as the class to which p belongs, and 0 otherwise [23]. Ties are resolved by choosing k as the first number in standard enumeration. Such simple majority vote criterion is effective if the accuracy of each classifier is above 50%, as the Condorcet's jury theorem ensures that in this case the accuracy of the ensemble is monotonically increasing and approaches 100% for $n \to \infty$ [24].

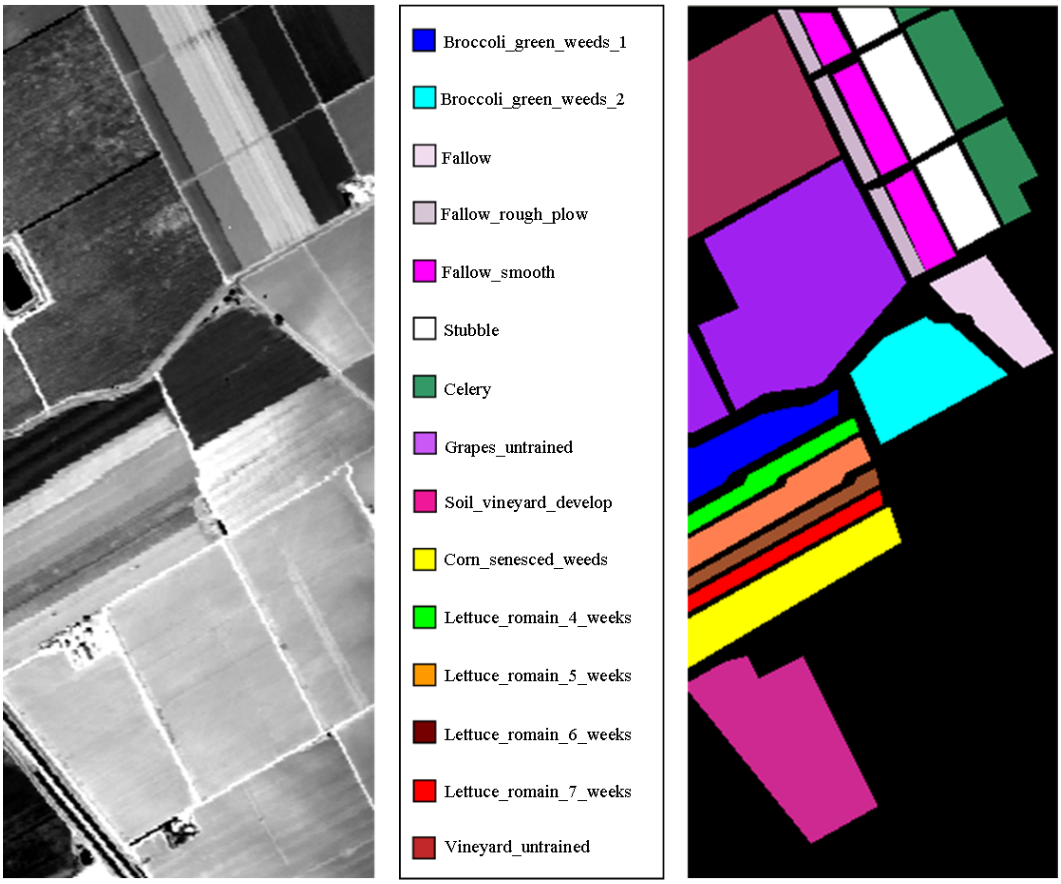


Fig. 5. Sample band from Salinas dataset (left), and available ground truth (right).

B. Adaptive spatial regularization

A classification based on synergetics as described so far is solely based on the data spectral information, as the analysis is carried out in the spectral domain spanned by the user-defined prototype vectors. Therefore, the information conveyed by the spatial interactions of the pixels in a scene are discarded, limiting the effectiveness of the method. In recent years, many pixel-wise hyperspectral classification algorithms have been defined which integrate spatial information in the analysis, often through the use of morphological operators [25] [6] [26].

The synergetics-based method presented in this work allows looking at the embedding of spatial information under a different point of view. An interesting way to carry out such step would be by manipulating the data directly in the prototype vectors space. Each test vector should be displaced towards its neighbours projected in the same vectorial space spanned by the user-

1. Given a square window w of side w_s, a distance measure d, a pixel $p(x, y)$ at coordinates x and y belonging to an image of size $m \times n$, & a threshold *t*:
2. if $(w_s/2 < x < (m + 1 - w_s/2)$ and $w_s/2 < y < (n + 1 - w_s/2))$ begin:
3. For $i = 0 \ldots w_s$ do
4. For $j = 0 \ldots w_s$ do
5. if $(d(p(x - [\frac{w_s}{2}] + i, y - [\frac{w_s}{2}] + j), p(x, y)) \leq t)$ $w(i, j) = 1$
6. else $w(i, j) = 0$
7. $w = \frac{w}{\text{total (where } w=1)}$
8. Return w

Fig. 6. Pseudocode to compute a convolution window w for the adaptive spatial filtering of a pixel p.

defined prototype vectors, but only if their spectral decompositions in terms of prototype vectors are similar. This would increase the probability of a pixel to be attracted by the stable state related to the class to which the majority of its adjacent pixels belong, leading to a more homogeneous classification. After the representation of the data in the parameter space, we employ then an adaptive low-pass filter, achieved through convolution in the prototype domain with a square window of size w_s, where w_s is an odd number, built as described by the pseudo-code in fig. 6, with the distance d representing Euclidean distance.

This introduces two additional parameters in the computation (the window size w_s and the threshold t), but improves results considerably and smooths intraclass variations while keeping edges information at the same time. Furthermore, the original spectral information is well preserved, as the value of each pixel is taken into account. This is an improvement over morphological operators, which alter in a more severe way the informational content of the data by replacing the value of a pixel with some value from its neighbourhood.

C. Attention parameters tuning

Up to this point the weighting for the prototype vectors, represented by the λ parameters, has not been taken into account. Therefore, for a prototype space composed by M classes of interest, $\lambda_i = 1, \forall i \in 1 \dots M$ in eq. (4).

In [12] the authors propose a weighting of the attention parameters through an explicit parameter learning phase, by choosing a decision boundary in the order parameter space. Such boundary divides the dataset into the classes of interest, and on its basis the attention parameters are derived. This approach finds an optimal parameters weighting, but has two major drawbacks. Firstly, it assumes that the data projected in the order parameter space falls along a smooth curve, i.e. that it is possible to perfectly separate the classes of interest by tuning the λ parameters, which is in general not the case. Furthermore, all the test set has to be used as training, since the algorithm requires to know a priori which objects are close to the decision boundary and how they are projected in the prototype vectors space. An award-penalty learning mechanism has been proposed in [20] to improve classification results based on synergetics theory. In this work, the attention parameters for a given class are iteratively increased or decreased by a small fraction δ, in presence of false negatives and false positives in the classification results, respectively. The system stops when an user-selected accuracy threshold is met. As this methodology uses the full test dataset as training, it is not feasible for real applications.

We propose an improvement over the methodology proposed in [20] for adjusting the attention parameters λ in the synergetics equation. Instead of using the complete data set as training, we select an additional training area T_i for each class C_i, and employ it to tune the overall λ values. The attention parameters are updated as follows. Let $FN(i)$, $FP(i)$, and $TN(i)$ be the false negatives, the false positives, and the total number of pixels in T_i. If, for a given class C_i, we have $FN(i) > FP(i)$, it means we must increase λ_i, as the class C_i is not correctly detected, and also is not dominant over the other classes. This means that in the order parameter space the final state(s) associated to C_i do not attract spectra belonging to other classes: therefore, if we increase the attention parameter λ_i, we expect the decrease of $FN(i)$ to be greater than the increase of $FP(i)$. So in this case we perform the following adjustment:

$$\lambda_i = \lambda_i \left(1 + \alpha \frac{FN(i)}{TN(i)}\right) \tag{10}$$

On the other hand, if $FP(i) > FN(i)$, than the class is an absorbing class, i.e. the predominant effect is the attraction into the final state related to C_i of objects belonging to other classes. But if the difference is small also the absorption effect is small, and it turns into confusion between classes instead. Therefore:

$$\lambda_i = \lambda_i \left(1 - \beta \frac{FP(i) - FN(i)}{TN(i)}\right) \tag{11}$$

Here α and β are two regularization constants, representing how much the λ parameters are modified, with β slightly higher than α, in order to balance the changes in the λ value. In the experiments contained in this section the values $\alpha = 0.1$ and $\beta = 0.15$ have been chosen. These parameters could be chosen to be smaller, but would need more iterations to converge to their final value. Such solution brings the process closer to the requirements of a real application, where often the user selects a restricted training area to perform the analysis. It has to be remarked that our approximation to the most similar base vector $v_{t=0}$ as chosen at $t = 0$ may now lead to misclassifications, since if the attention parameters are changed the winning prototype can be different from $v_{t=0}$. According to [13], for the case of a two dimensional prototype space, no attention parameter should be set to more than twice the value of the other to ensure that a test vector is attracted by the final state corresponding to the prototype vector of highest potential at $t = 0$.

The need of an additional training area T_i can be eliminated if T_i is composed for each classification by the training samples which do not take part in the formation of the prototype vectors space. It will be clear from the results achieved that such solution gives by far the best compromise between classification quality and size of the training area needed by the algorithm, at the cost of a negligible computational overhead.

D. Required computational resources

The computational complexity of the algorithm is driven by the projection of the image elements onto the adjoint vector space and by the spatial regularization step. A rough estimate of the computational resources needed by each step of the procedure is reported in Table I. The first step taken to produce the final classification map is the creation of the adjoint vector space, which requires a number of operations proportional to $N \times C^3 \times B$, where N is the number

TABLE I

ROUGH ESTIMATE OF OPERATIONS REQUIRED FOR EACH STEP OF THE CLASSIFICATION PROCEDURE (REFER TO SECTION I FOR A DESCRIPTION OF EACH ENTITY).

Step	Required Operations
Adjoint vectors space	$N \times C^3 \times B$
Data projection	$N \times C \times P \times B$
Spatial regularization	$P \times W^2 \times C^2 \times N$
Attention parameters tuning	$N \times C \times L$
Final classification	$P \times N \times C$

of training samples per class, C the number of classes, and B the number of bands in the hyperspectral image; this step is dominated by the N matrix inversions required. The projection step requires $N \times C \times P \times B$ operations, where P is the total number of pixels in the image. The spatial regularization step comprises $P \times W^2 \times C^2 \times N$ operations, with W the size of the square analyzing window employed, and the term C^2 deriving from the computation of the euclidean distances in the parameters space. The attention parameters tuning is less expensive, having a complexity proportional to $N \times C \times L$, with L being the number of iterations. Finally, the majority vote of independent classifications needs operations proportional to $P \times N \times C$. The total complexity of the algorithm is then proportional to $K_{tot} \rightarrow (N \times C)((C^2 \times B) + P(B + W^2 \times C + 1) + L)$. In a realistic scenario $O(N) \approx O(C) \approx O(L) \approx O(W^2) = O(A)$, and we can rewrite $K_{tot} \rightarrow A^2((A^2 \times B) + P(B + A^2 + 1) + A) \approx A^2P \times (A^2 + B)$, considering that $B << P$.

To give an idea, the experiments reported in next section had typical values (as described in this section) of: $P \approx 25.000$, $B \approx 200$, $C = 15$, $N = 20$, $W = 5$, and $L = 15$. The running time to produce a classification map with these settings on a machine with a double 2 GHz processor and 2GB of RAM, with a non-optimized code written in IDL, was less than 5 minutes.

IV. RESULTS

A. *Salinas AVIRIS dataset*

We analyze a hyperspectral scene acquired by the Airborne VisibleInfrared Imaging Spectrometer (AVIRIS) sensor of the Jet Propulsion Laboratory acquired over the Salinas Valley in

California, USA. The full scene has a size of 512 $\times$ 217 samples with 192 spectral bands in the range 0.4 to 2.5 μm, after removing water absorption bands and noisy bands according to [26]. The sensor has a spectral resolution of 10 nm and a spatial resolution of 3.7 m. The data are at-sensor radiance measurements and include vegetables, bare soils and vineyard fields. A sample band of the scene and the available ground truth are shown in Fig. 5.

The test dataset has been analyzed with the described methodology, summarized by Fig. 4. Twenty samples per class have been chosen (see Fig. 7(a)) and the same number of independent classifications have been carried out, with the final result derived from a majority voting as explained in the previous section. Figure 7(b) presents an RGB combination for the first three Principal Components extracted from the scene: as the class *corn* is composed by two different homogeneous areas (”corn” and ”senesced weeds” - see the different values for the pixels in the upper and lower part of the class in the image), two different classes have been considered and then merged in an unique class, as a post-processing step common to each of the carried out classification procedures. Therefore, 40 samples have been employed for this class. It can also be noticed that some classes look difficult to separate (e.g. *Broccoli1* and *Broccoli2*, or *Grapes* and *Vineyard untrained*).

Results of the independent classifications have been improved by an additional step of attention parameters tuning, carried out with three different settings, all of them for 16 iterations. In the first setting, with a similar approach to the one contained in [20], we used the full ground truth as a reference, and used it to tune the parameters as described in the above section. As this approach is not realistic in practical applications, where the classes of the test set are usually unknown, in a second setting we selected a separate training set, consisting of 100 samples per class. Finally, in the third setting we selected no additional training area, but used for every classification one sample per class to build the prototype vector space and the other 19 to tune the lambda parameters. This introduces a negligible computational overhead, as the λ parameters tuning is achieved in linear time for one iteration.

Fig. 8 shows the classification results for the overall scene. Fig.8(a) presents the results for a classification carried out on the basis of a single training sample per class, with confusion and salt and pepper noise being evident in many classes. Results improve considerably in Fig.8(b), after a plurality vote of 20 independent classifications. The improvement in overall accuracy achieved through majority vote agrees with the expected one of around 15% for a comparable

number of independent classifications and accuracy of a single classifier [23]. Salt and pepper noise in the classification is removed in Fig.8(c) after a preliminary step of adaptive spatial filtering, as detailed in Section III-B. Results benefit further from an automatic tuning of the attention parameters, with this improvement being more obvious when the full ground truth is taken as a reference (Fig.8(d) and 8(e)). Even though misclassifications are present, it has to be remarked that the confusion is almost exclusively limited to classes belonging to a same super-class. Therefore, we have confusion between vineyards and grapes, different fallow or broccoli fields, and lettuces of different age. The improvements obtained through the automatic tuning of the attention parameters for the case of the full ground truth adopted are reported in Fig. 9. As the algorithm tries to find the best parameters for all classes, the classes of interest containing a large number of pixels are not given priority, and may be penalized yielding a worse overall accuracy. On the other hand, the plot of the values for the average accuracy exhibits an increase up to an horizontal asymptotic value of approximately 90%. This suggests that the proposed training procedure, although empiric, converges to some local optimum.

The selection of an extra training area is unconvenient, as it increases the size of the training set (Fig.8(d)), or is not at all realistic for practical applications where no ground truth is available (Fig.8(e)). Therefore, the final results in Fig.8(f) represent the best compromise between accuracy and size of the training set, as an accuracy comparable to the classification in Fig.8(e) is achieved without the need of an additional training area for the λ parameters. This is justified by the fact that the majority vote benefits from having as input more accurate classifications, achieved through separate λ parameters tuning steps.

Table II reports the classification accuracy on the dataset. The overall accuracy (OA) is computed as:

$$OA = \frac{100}{N_{TOT}} \sum_{i=1}^{M} p_{i,i}, \tag{12}$$

where $p_{i,i}$ represents the number of the pixels from class i which are correctly assigned to i, M is the number of classes, and N_{TOT} is the total number of pixels in all classes. The average accuracy AA is computed as:

$$AA = \frac{1}{M} \sum_{i=1}^{M} \frac{100 * p_{i,i}}{N_i}, \tag{13}$$

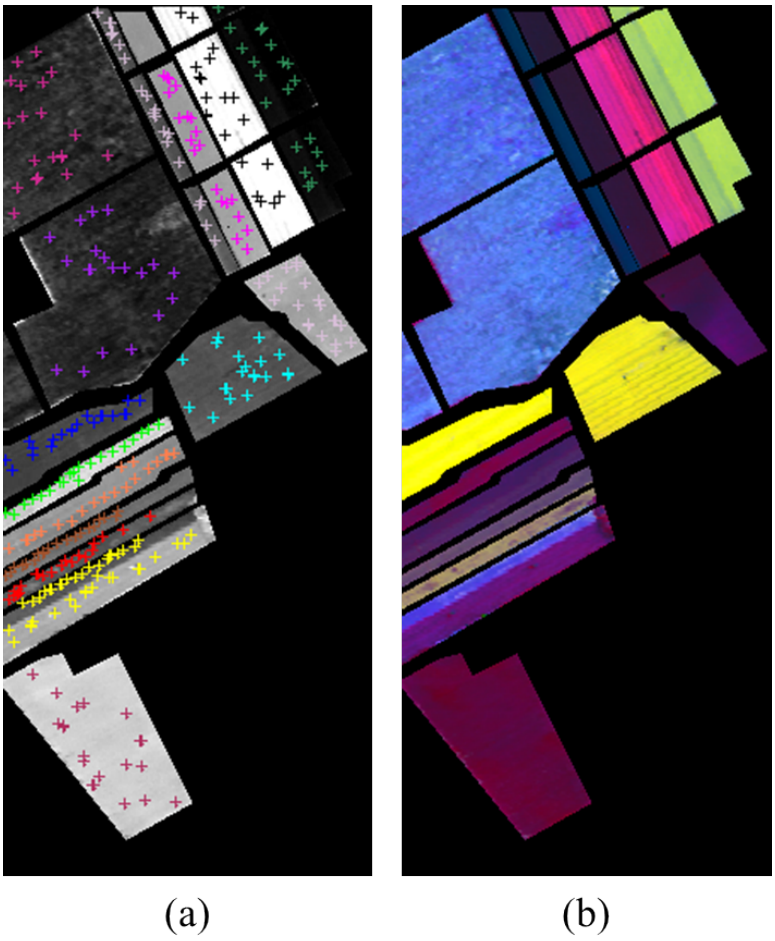


Fig. 7. Training data collected over the scene (a) and RGB composition of the first three Principal Components extracted from the scene (b).

where N_i is the number of pixels in class i.

To mitigate the influence of the training samples we produced four maps as in Fig.8(f). This resulted in an average overall accuracy of 88.12%, with a standard deviation in the results of $\sigma = 0.7$.

In order to have a fair comparison to other techniques, we performed a classification with the same training dataset using well-known distance measures and classification techniques widely used in hyperspectral data analysis. As distance measures, the Spectral Angle Mapper (SAM) [17] and the Spectral Information Divergence (SID) [29] have been applied in two different ways to produce a classification map: majority vote of 20 separate classifications and a single

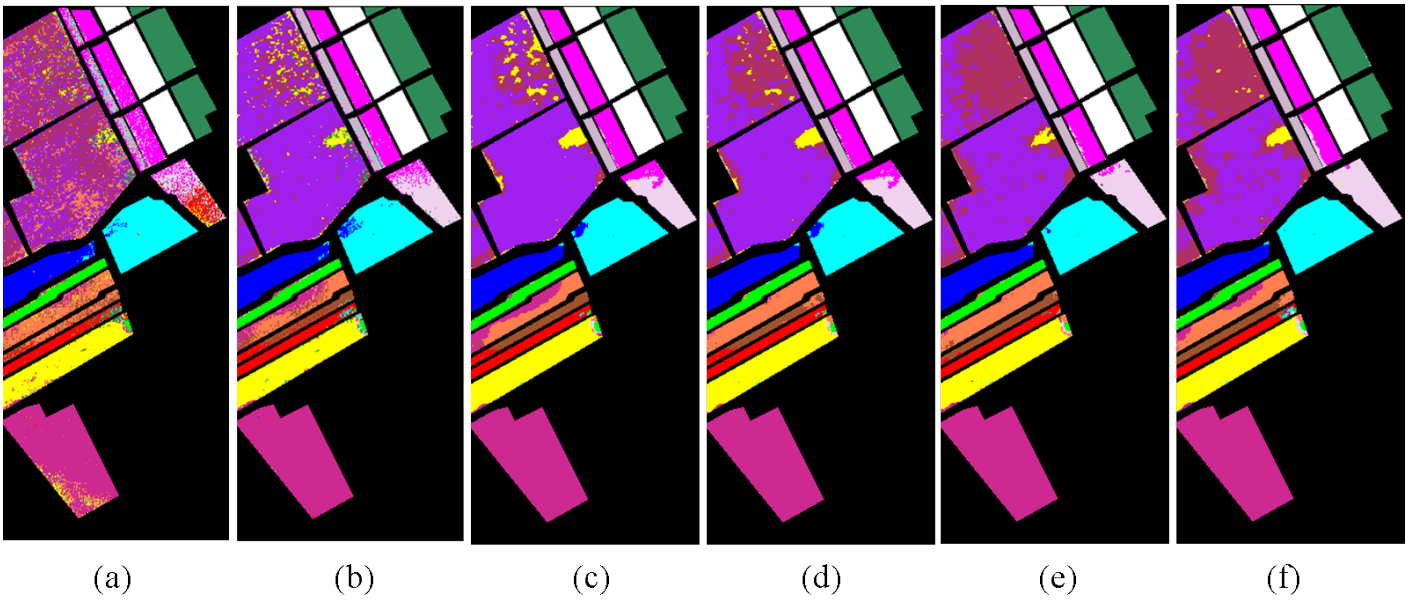


Fig. 8. Synergetics classification results with different settings. From left to right: (a) classification obtained with one training sample per class; (b) majority voting of 20 independent classifications, using a training dataset of 20 samples per class; (c) same as (b), carried out after a step of spatial regularization; (d) image (c) after λ parameters tuning, obtained with an extra set of 100 pixels per class; (e) image (c) after λ parameter tuning obtained using the full ground truth; (f) majority voting of 20 independent classifications using a training set of 20 samples per class. Each classification is obtained on the basis of 1 sample per class, with the other 19 used to tune the λ parameters. This image represents the best compromise between classification accuracy and traning dataset size.

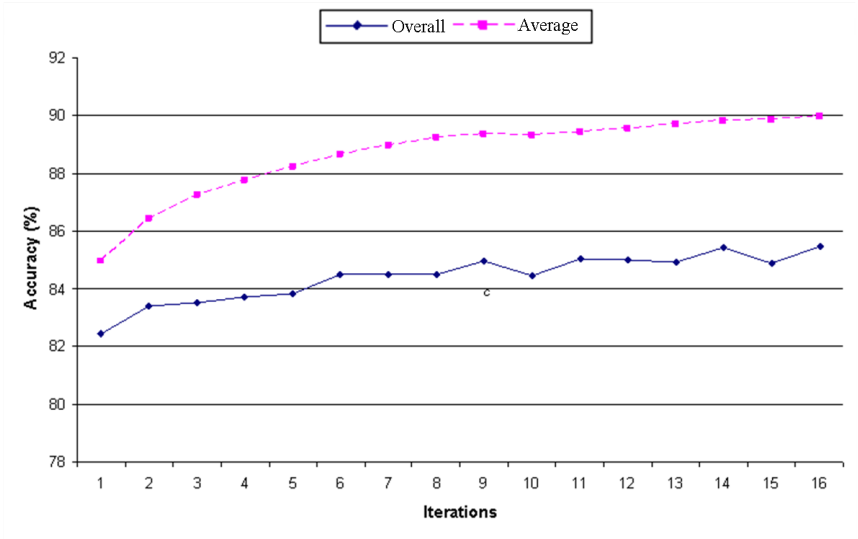


Fig. 9. Increase in overall and average accuracy after automatic tuning of the λ parameters.

TABLE II

CLASSIFICATION RESULTS FOR THE SALINAS DATASET: OVERALL ACCURACY (OA), AVERAGE ACCURACY (AA), AND NUMBER OF TRAINING SAMPLES PER CLASS (TS). THE HORIZONTAL LINES SEPARATE THE METHODS IN THREE GROUPS: CLASSIFICATION BASED ON SYNERGETICS, CLASSIFICATIONS WITH OTHER METHODS PERFORMED USING THE SAME TRAINING DATASET OF THE CATEGORY ABOVE, AND PREVIOUS WORKS BY OTHER AUTHORS. THE RESULTS BASED ON SYNERGETICS CONTAIN AS ADDED PARAMETER THE NUMBER OF ADDITIONAL TRAINING SAMPLES PER CLASS EMPLOYED TO TUNE THE ATTENTION PARAMETERS λ (TS_λ).

	OA	AA	TS	TS_λ
Synergetics (a)	69.57	74.26	1	0
Synergetics (b)	82.43	84.96	20	0
Synergetics (c)	84.1	87.8	20	0
Synergetics (d)	87.82	90.21	20	100
Synergetics (e)	90.15	94.4	20	all
Synergetics (f)	88.82	92.35	20	0
SVM	81.44	90.6	20	
SID	83.71	90.2	20	
SAM	78.64	85.4	20	
Factor Graphs [27]	85.32	90.91	20	
Neural Networks [26]	87.55	88.03	40	
Multinomial Logistic Regression [28]	86.49	NA	15	
SVM + Morphological Operators [25]	87.25	NA	65	

classification using the full training set, merging afterwards the classes of interest, following the criterion of minimizing the overall errors. The two techniques gave similar results and only the best results are shown, in which SID shows better discrimination power than SAM. We also performed a classification with Support Vector Machine (SVM) [30], which operates in implicit parameter hyperspaces by finding a manifold which divides the data of interest in two groups in the hyperspace, according to some criteria. In spite of being a general classification methodology, SVM have been often applied to hyperspectral data, due to their natural connection to multidimensional data [31]. We used a Gaussian Radial Basis Function (RBF) kernel defined as $K(u, v) = \exp(-\gamma|u - v|^2)$, which is found to yield the best results for the classification of a different AVIRIS scene (Indian Pines) in [32]. We found empirically the best parameters after several tests, and set γ to 0.01 and assigned a large penalty to errors $C = 100$. We assess the

TABLE III

CONFUSION MATRIX FOR CLASSIFICATION AS IN SYNERGETICS(F).

	Broc1	Broc2	Flw	Flw Rgh	Flw Smt	Stble	Clry	Grps	Soil	Corn	Ltc 4	Ltc 5	Ltc 6	Ltc 7	Vnyrd
Broccoli1	**99.4**	0.6	0	0	0	0	0	0	0	0	0	0	0	0	0
Broccoli2	0.03	**99.95**	0	0	0	0	0.3	0	0	0	0	0	0	0	0
Fallow	0	0	**98.89**	0	0.91	0	0	0	0	0.2	0	0	0	0	0
Flw Rough	0	0	0	**99.78**	0.22	0	0	0	0	0	0	0	0	0	0
Flw Smooth	0	0	5.53	0.37	**93.95**	0	0	0	0	0	0.04	0.04	0	0	0
Stubble	0	0	0.13	0	0	**99.87**	0	0	0	0	0	0	0	0	0
Celery	0	0	0.17	0	0	0	**99.83**	0	0	0	0	0	0	0	0
Grapes	0	0	0.25	0	0	0	0	**74.17**	0.1	2.84	0	0	0	0	22.62
Soil	0	0	0	0	0	0	0	0	**100**	0	0	0	0	0	0
Corn	0	0.12	2.44	0	0	0	0	0	1.25	**94.08**	1.13	0.98	0	0	0
Lettuce4	0	0	1.69	0	0	0	0	0	1.5	0	**96.82**	0	0	0	0
Lettuce5	0	0	0	0	0	0	0	0	20.03	0	0	**79.87**	0.1	0	0
Lettuce6	0	0	0	0	0	0	0	0	0	0	0	13.97	**85.7**	0.33	0
Lettuce7	0.28	4.02	1.03	0	0	0	0.84	0	0	0	0.84	0	4.39	**88.6**	0
Vineyard	0	0	0.56	0	0	0	0	23.82	0.04	1.17	0	0	0	0	**74.41**

statistical significance of the difference in classification accuracy against SVM by McNemar's test [33], which shows the two classifiers to be very different, with a probability for the differences in the results to be caused by random variations of some kind below 1%. Finally, we include results obtained through Factor Graphs [34] also on the base of the same training samples [27], after applying a median filter to the parameter space. The main difference in the results between the proposed approach and its competitors is the better discrimination between the classes *grapes* and *vineyard untrained*, which drastically improves after the λ-parameter tuning.

We also compared results obtained on the same dataset in recent works in literature [26] [25] [28]. As the first two make extensive use of morphological profiles, for sake of comparison we took into account results obtained on the original spectral information, achieved with neural networks [35] in the former and with SVM classifiers in the latter. In both cases, a sequence of 9 opening and closing morphological operations has been subsequently applied to regularize and

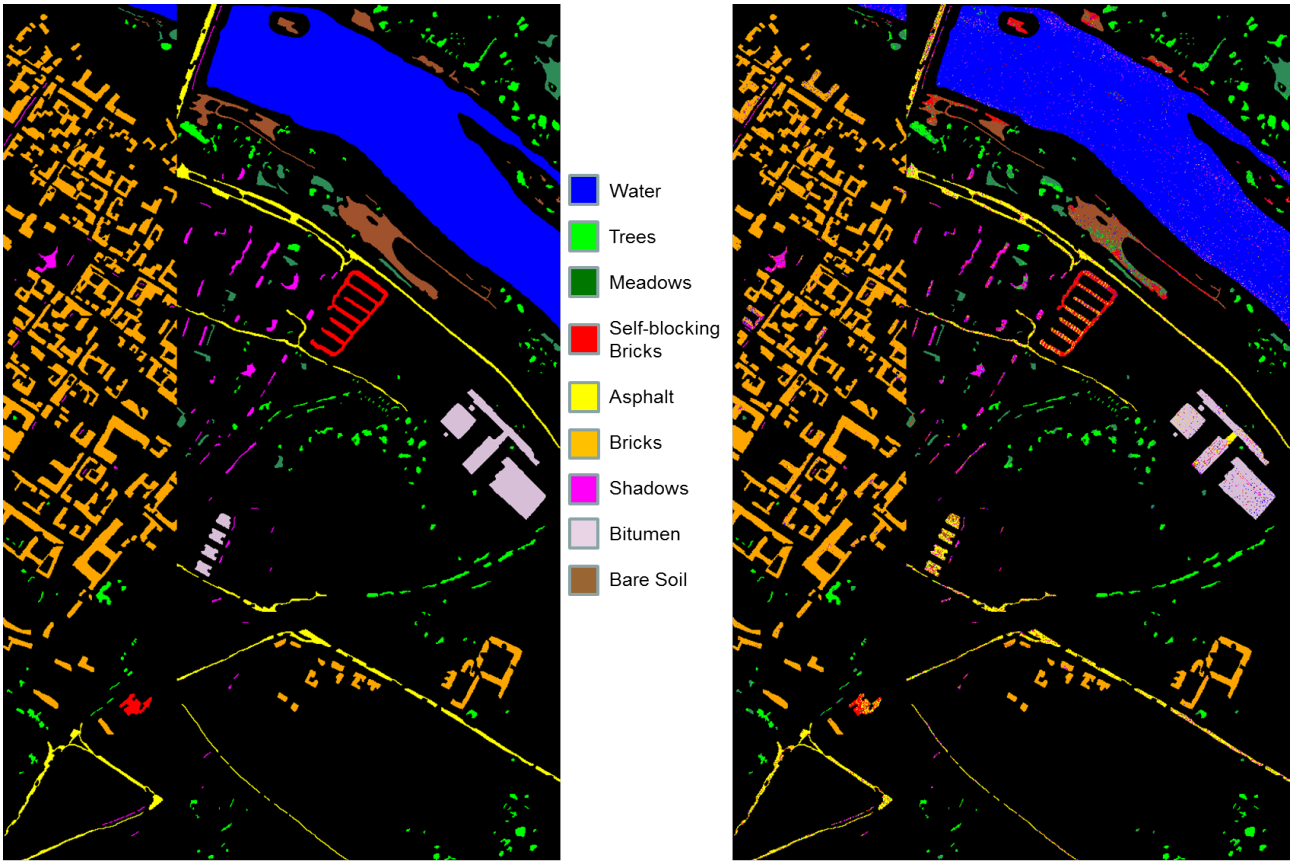
improve classification results, in a step which could be comparable to the spatial regularization described in Section III-B. In the case of [28] we take into account the results obtained with multinomial logistic regression [36] after the collection of both labeled (L) and unlabeled (U) samples, with $U = 2L$, before the integration of additional information through a segmentation step. Results summarized in Table II show the proposed approach to be competitive both in terms of classification accuracy and number of training samples needed, with results reported in Synergetics (f) outperforming the competitors. It has to be remarked that the use of extended morphological profiles and segmentation, which allows achieving superior results in the aforementioned works, could be also included to improve the results of the proposed technique.

B. Pavia Center ROSIS dataset

We analyze the Pavia Center image acquired by the Reflective Optics System Imaging Spectrometer (ROSIS). The dataset consists of 610 $\times$ 340 pixels and has 103 bands selected from the available 115, in the spectral range 0.43 to 0.86 μm, and with a spatial resolution of 1.3 m [6]. The ground truth is shown in Fig. 10 (a). We collect only 10 training samples per class and follow the same workflow leading to the classification reported in Fig. 8 (f), using the training data both to perform a set of independent classifications and to tune the λ parameters in each classification.

Results reported in Fig. 10 (b) have a satisfactory Overall accuracy of 91.25%. Some confusion between shadow and water is probably due to the noise which is accentuated after normalizing the spectra, and that is affecting ROSIS data more than AVIRIS, also because of the narrower spectral bands characterizing the former sensor. The class *shadows* is separated in a satisfactory way, but as the spectra are normalized this indicates that the materials within the shadow area are largely similar. If the spatial regularization step described in Section III-B is skipped, results are only slightly worse (1% in terms of Overall Accuracy). Better results could be obtained by employing morphological operators on the data or on the classification results.

Experiments in [37] obtain a higher Overall Accuracy, but employ a training dataset about 50 times larger.



(a) (b)

Fig. 10. Ground truth (a) and synergetics classification results (b) for the ROSIS Pavia center dataset.

V. Conclusions

In this paper we presented a methodology for hyperspectral data analysis based on synergetics theory, which represents the first attempt at producing a pixel-wise image classification derived from these notions. This method performs a focused dimensionality reduction, by representing the data in a vectorial space which uses a basis derived from user-defined prototype vectors. To overcome the lack in flexibility of synergetics-based systems in the training step we allow the selection of training areas, by performing classification on a majority voting basis. Furthermore, the same training areas can be used to tune the attention parameters in the synergetics equation, improving significantly the classification accuracy. The approach to the classification is discriminative, which represents an advantage for hyperspectral data processing, as it requires a smaller

training set and deals with Hughes phenomenon better with respect to generative models (see [28] and references therein). This is of particular interest in perspective to the future advent of spaceborn sensors such as the EnMAP mission [38], which will increase considerably the amount of exploitable hyperspectral remotely sensed data. As a consequence, it will be required to have available techniques which make such multidimensional data easier to handle. Results on the AVIRIS Salinas scene show that the proposed methodology can outperform traditional algorithms employed in the analysis of hyperspectral data. The main advantage of the proposed method resides in the representation of the data in the prototype vectors space, which enables operations in a semantic space, as the value of a projected test vector in each dimension represents the similarity to a given class of interest, and the attention parameters λ used to weight the projections are interdependent. As in traditional techniques the weighting of parameters is usually based on some property of the data, the proposed approach is capable of providing more accurate results. These could be further improved through a more refined technique to include spatial prior information, such as the one used in [28]. Alternatively, a segmentation step [39] could be performed, followed by a region-based classification with the proposed method, using the average spectrum of each segment as a test vector. An additional iterative step employing morphological filters could also improve results, with the drawback of removing at the same time information which could be relevant [26].

Correspondences between this approach and classical estimation methods such as Least Squares have been considered for the first time, opening interesting perspectives: the proposed approach could be slightly modified by employing Total Least Squares (TLS) [40]. This would allow taking into account the variability of the selected prototypes, regarded as measurements errors used to weight the TLS objective function.

The original formulation of synergetics for pattern recognition in eq. (4) presents some non-linear terms, representing the interactions between the chosen prototype vectors plus a saturation component. These quantities should be investigated in the future, as only the first term which is linear has been considered in the reported experiments. Additional future work includes the description of a rejection class for the described classification method, which should be linked to the residual vector $\mathbf{r}$ in equation (3). Furthermore, the supervision required by the algorithm could be reduced by defining an adaptive threshold for the spatial regularization step described in Section III-B.

The proposed technique would be a good choice for analysis of hyperspectral images of natural scenes, which usually are characterized by a limited intraclass variability. Suggested uses could include geological applications, acid mine drainage monitoring, and vegetation classification.

ACKNOWLEDGMENTS

SVM implementation courtesy of Humboldt University, Berlin [41]. The authors would like to thank P. Gamba for providing the Pavia city dataset and A. Makarau for fruitful discussions.

REFERENCES

[1] D. W. Scott, *The Curse of Dimensionality and Dimension Reduction*. John Wiley and Sons, Inc., 2008, pp. 195–217. [Online]. Available: http://dx.doi.org/10.1002/9780470316849.ch7

[2] J. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435 –2445, aug. 2008.

[3] J. Chanussot, M. M. Crawford, and B.-C. Kuo, "Foreword to the special issue on hyperspectral image and signal processing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 3871–3876, 2010.

[4] S. Kaewpijit, J. Le Moigne, and T. El-Ghazawi, "Automatic reduction of hyperspectral imagery using wavelet spectral analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 4, pp. 863 – 871, april 2003.

[5] H. Haken, *Synergetics Computers and Cognition*. Springer Series Synergetics, 1991, vol. 50.

[6] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, and et al., "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, no. Supplement 1, pp. S110–S122, 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0034425709000807

[7] H. Haken, "Synergetics," *Scholarpedia*, vol. 2, no. 1, p. 1400, 2007. [Online]. Available: http://www.scholarpedia.org/article/Synergetics

[8] M. Bushev, *Synergetics : chaos, order, self-organization*. World Scientific, 1994.

[9] H. Haken, "Synergetics," *Naturwissenschaften*, vol. 67, pp. 121–128, 1980, 10.1007/BF01073611. [Online]. Available: http://dx.doi.org/10.1007/BF01073611

[10] W. K. Konen, T. Maurer, and C. Von Der Malsburg, "A fast dynamic link matching algorithm for invariant pattern recognition," *Neural Networks*, vol. 7, no. 7, pp. 1019–1030, 1994.

[11] K. Crounse and L. Chua, "A synergetics approach to image processing in cellular neural networks," in *Circuits and Systems, 1996. ISCAS '96., Connecting the World., 1996 IEEE International Symposium on*, vol. 3, may 1996, pp. 134 –137 vol.3.

[12] T. Hogg, T. Talhami, and D. Rees, "An improved synergetic algorithm for image classification," *Pattern Recognition*, vol. 31, no. 12, pp. 1893 – 1903, 1998.

[13] H. Maeda, S. Okada, and T. Irie, "A new method for stereo matching problem in computer vision using synergetics," in *Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on*, vol. 3, 1999, pp. 503 –508 vol.3.

[14] F. G. Boebel and T. Wagner, "Theoretical foundations of synergetic image processing," in *Applied Synergetics and Synergetics Engineering Conference*, 1994, pp. 46–52.

[15] H. Kawano, S. Imamura, H. Maeda, and N. Ikoma, "3d rough reconstruction of buildings from streetscape by synergetic stereo matching," in *Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on*, vol. 3, oct. 2006, pp. 1992 –1999.

[16] S. Gou and L. Jiao, "Sar image recognition using synergetic neural networks based on immune clonal programming," in *Advances in Neural Networks ISNN 2004*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2004, vol. 3173, pp. 918–923.

[17] F. A. Kruse, A. B. Lefkoff, J. B. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz, "The Spectral Image Processing System (SIPS) - Interactive Visualization and Analysis of Imaging Spectrometer Data," *Remote Sensing of Environment*, vol. 44, pp. 145–163, 1993.

[18] C.-I. Chang, "Orthogonal subspace projection (osp) revisited: a comprehensive study and analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 502 – 518, march 2005.

[19] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44 –57, jan 2002.

[20] F. Y. Wang, P. J. A. Lever, and B. Pu, "A robotic vision system for object identification and manipulation using synergetic pattern recognition," *Robot. Comput. Inegrated Manufacturing*, vol. 10, no. 6, pp. 445–459, 1993.

[21] T. Wagner and F. G. Boebel, "Testing synergetics algorithms with industrial classification problems," *Neural Networks*, vol. 7, no. 8, pp. 1313–1321, 1994.

[22] T. Zhao, L. Tang, H. Ip, and F. Qi, "On relevance feedback and similarity measure for image retrieval with synergetic neural nets," *Neurocomputing*, no. 51, pp. 105–124, 2002.

[23] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004.

[24] K. K. Ladha, "The Condorcet Jury Theorem, Free Speech, and Correlated Votes," *American Journal of Political Science*, vol. 36, no. 3, 1992.

[25] J. Plaza, A. J. Plaza, and C. Barra, "Multi-channel morphological profiles for classification of hyperspectral images using support vector machines," *Sensors*, vol. 9, no. 1, pp. 196–218, 2009. [Online]. Available: http://www.mdpi.com/1424-8220/9/1/196/

[26] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 466–479, march 2005.

[27] A. Makarau, R. Mueller, G. Palubinskas, and P. Reinartz, "Hyperspectral data classification using factor graphs," in *to appear in ISPRS 12*, 2012.

[28] J. Li, J. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4085 –4098, nov. 2010.

[29] H. Du, C. Chang, H. Ren, C. Chang, J. Jensen, and F. D'Amico, "New hyperspectral discrimination measure for spectral characterization," *Optical Engineering*, vol. 43, no. 8, pp. 1777–1786, 2004.

[30] T. Joachims, "Making large scale SVM learning practical, Universitaet Dortmund Press," 1999.

[31] B. Demir and S. Erturk, "Empirical mode decomposition of hyperspectral images for support vector machine classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4071 –4084, nov. 2010.

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

[33] G. Foody, "Thematic map comparison: evaluating the statistical significance of differences in classification accuracy," *Photogrammetric Engineering and Remote Sensing*, vol. 70, no. 5, pp. 627–633, 2004.

[34] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, pp. 498–519, 1998.

[35] C. Lee and D. A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 4, pp. 388–400, Apr. 1993. [Online]. Available: http://dx.doi.org/10.1109/34.206958

[36] D. Boehning, "Multinomial logistic regression algorithm," *Annals of the Institute of Statistical Mathematics*, vol. 44, no. 1, pp. 197–200, 1992. [Online]. Available: http://www.springerlink.com/index/P661G851671713U2.pdf

[37] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Transactions Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3973–3985, 2011.

[38] T. Stuffler, K. Foerster, S. Hofer, M. Leipold, B. Sang, H. Kaufmann, B. Penne, A. Mueller, and C. Chlebek, "Hyperspectral imaging - an advanced instrument concept for the enmap mission (environmental mapping and analysis program)," *Acta Astronautica*, vol. 65, no. 7, pp. 1107–1112, 2009.

[39] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new bayesian approach with active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3947 –3960, oct. 2011.

[40] G. Golub and C. Van Loan, "Total least squares," in *Smoothing Techniques for Curve Estimation*, ser. Lecture Notes in Mathematics, T. Gasser and M. Rosenblatt, Eds. Springer Berlin / Heidelberg, 1979, vol. 757, pp. 69–76, 10.1007/BFb0098490. [Online]. Available: http://dx.doi.org/10.1007/BFb0098490

[41] "Imagesvm." [Online]. Available: http://www2.huberlin.de/hurs/projects/imageSVM.php