A CLASSIFICATION ALGORITHM FOR HYPERSPECTRAL DATA BASED ON SYNERGETICS THEORY

Daniele Cerra, Rupert Mueller, Peter Reinartz

German Aerospace Center (DLR)
Muenchner Strasse 20, Oberpfaffenhofen, 82234 Wessling, Germany
daniele.cerra@dlr.de, rupert.mueller@dlr.de, peter.reinartz@dlr.de

Commission VII/3

ABSTRACT:

This paper presents a new classification methodology for hyperspectral data based on synergetics theory, which describes the spontaneous formation of patterns and structures in a system through self-organization. We introduce a representation for hyperspectral data, in which a spectrum can be projected in a space spanned by a set of user-defined prototype vectors, which belong to some classes of interest. Each test vector is attracted by a final state associated to a prototype, and can be thus classified. 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, used to weight the prototype vectors through attention parameters and to produce a more accurate classification map through majority voting of independent classifications. Results are comparable to state of the art classification methodologies, both general and specific to hyperspectral data and, as each classification is based on a single training sample per class, the proposed technique would be particularly effective in tasks where only a small training dataset is available.

1 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 (Scott, 2008). Very often, not all the bands are useful for a given application. As a consequence, band selection can be performed. Alternatively, the data are often projected on a lower-dimensionality space to aid data exploration and improve computation performances (Bioucas-Dias and Nascimento, 2008). This is usually a preprocessing step aiding other operations such as classification and target detection. One of the most widely used dimension reduction 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 (Kaewpijit et al., 2003). If the user desires to differentiate different classes of interest, however, such approaches may not be optimal, as in general the dimensions in the subspaces 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 two decade old theory describing the spontaneous formation of patterns and structures in a system through self-organization. Applications based on synergetics have been derived in the pattern matching and image classification domains, but they have often been limited by the dependency of such systems on scaling, rotation and shifting of the images (Haken, 1991). These drawbacks can be discarded in applications to hyperspectral data performed in the spectral domain, as the study of the connections between synergetics and established methodologies and estimation techniques results in a novel representation for these data. Each pixel is represented as a data point projected in a subspace composed by a set of user-defined prototype vectors, belonging to some classes of interest. The pixel may then be represented as a particle on a potential surface, built as a manifold in this subspace, and is attracted by one of several possible final states, with each one being associated to a user-defined class, and hence classified. 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, used to weight the prototype vectors through attention parameters and to produce a final classification map through majority voting of independent classifications. The results obtained are comparable to state of the art classification methodologies, both general and specific to hyperspectral data, and could be easily improved by taking into account the spatial distribution of the data, by applying morphological filtering or segmentation.

The paper is structured as follows. Section 2 contains a brief reminder on synergetics theory, and analyzes the relation between synergetics and established concepts in estimation theory and data processing. Section 3 illustrates how the synergetics principles can be applied to hyperspectral data, to reduce dimensionality and perform pixel-wise classification. Section 4 reports experiments on an AVIRIS dataset and comparisons to state-of-the-art methods. We conclude in Section 5.

2 SYNERGETICS THEORY

Synergetics is an interdisciplinary science originally founded by Hermann Haken in 1969 (Haken, 2007). 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, and up to image elements. The term synergetics derives from the Greek "working together", indicating the cooperation of different parts in a system or different systems. Such system parts obey to an enslaving principle related to some
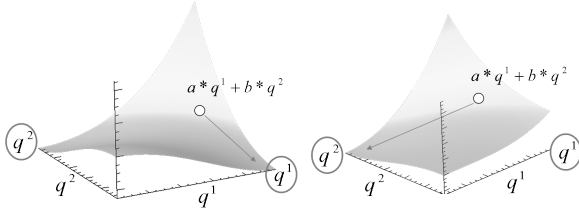
Figure 1: Energy function $E(q^1, q^2)$, for two prototypes q^1, q^2 in the order parameter domain and behaviour of a test vector projected on the plane in the parameter domain. The test vector is attracted towards a stable focus, corresponding for a final state for the vectors q^1 or q^2 (circled in 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 q^1, has been drastically decreased, and the test vector is then attracted by the final state associated to the vector q^2.

order parameter, which drastically reduces the degrees of freedom in the system.

After the first one by Haken himself (Haken, 1991), numerous pattern recognition algorithms based on synergetics have been described. Applications on image classification have been proposed, among others, in (Crounse and Chua, 1996, Hogg et al., 1998, Maeda et al., 1999, **?**). 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.

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. A dual space built using the adjoint vectors of the chosen prototypes forms a basis for a space of the kind described in (Haken, 1991). An input pattern belonging to an unknown class is then presented to the system, and represented as a linear combination of the prototype patterns. Subsequently, 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. The basic equation of synergetics for pattern recognition describes the time evolution of the feature vector $q(t)$ in the adjoint space built from M prototype vectors. In order to establish a dynamic system the following energy function is defined:

$$E(q^k) = -\sum_{k=1}^{M} \lambda_k \cdot (q^k)^2 + \frac{B}{2} \sum_{k=1_{k \neq j}}^{M} \sum_{j=1}^{M} (q^k)^2 \cdot (q^j)^2 + \quad (1)$$
$$\frac{C}{2} \left(\sum_{k=1}^{M} (q^k)^2 + |r|^2 \right)^2,$$

where the terms $q^k = (v^k, q)$ are the projections of q onto the adjoint basis vectors v^k and are called order parameters according to the theory of synergetics, r denotes the residual vector orthogonal to the subspace, B and C are two positive constants, and $\{\lambda_1, ...\lambda_M\}$ are positive values also called attention parameters.

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
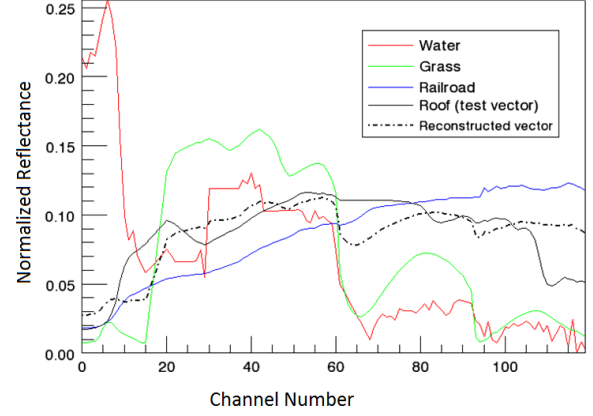

Figure 2: Prototype vectors as normalized spectra collected from a HyMap scene (water, grass, and railroad), plus reconstruction of a test vector (roof) as a linear combination of the prototype vectors.
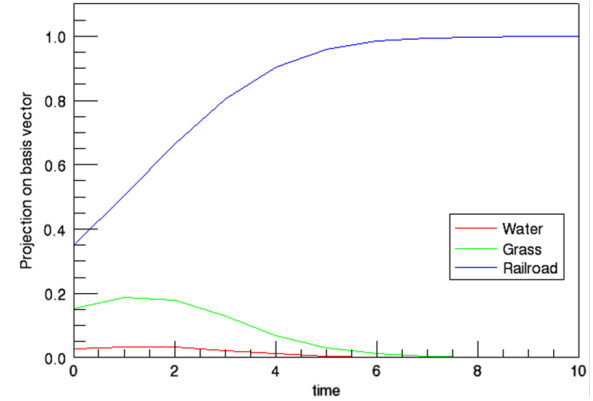

Figure 3: Simulation of the synergetics process for the test vector in Fig. 2, and related to the same prototype vectors. At $t = 0$ the test vector is projected into the space spanned by the three prototype vectors. The test vector is then attracted by one stable final state, corresponding to one of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

two prototypes q^1 and q^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 q^1 or q^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 q_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 q_2.

3 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) (Kruse et al., 1993), Orthogonal Subspace Projection (OSP) (Chang, 2005), and spectral unmixing techiques (Keshava and Mustard, 2002), and can be described as follows. Firstly, a set of prototype vectors (or classes) $v_k \in R^N$, $k = 1, ..., M$, is chosen. These vectors are formed by N-dimensional real valued components derived from spectral signatures (e.g. spectrum of one sample or mean value of sev-

eral samples, spectral derivates), which are linearly independent and normalized so that $|v_k| = 1$. The normalization suppresses illumination influences, as only the direction of the prototypes in the feature space is used, being in this similar to the SAM method. In a similar way to spectral unmixing techniques, the projected vector S_{test} is then expanded in the subspace of the adjoint prototypes S_v^k. Its representation is given in terms of the order parameters S_{test}^k, which are abundance values related to the composition of S_{test} in terms of the prototype vectors. Finally, a potential is computed for the unknown vector S_{test} in each dimension, and its evolution in time is tracked as S_{test} is pulled towards one of the k possible final states, in a similar way to the example sketched in Fig. 1. The synergetics principle (Haken, 1991) ensures that every prototype has an associated final state, and that no other final states exist safe from the ones associated to each vector in the adjoint vector space: this is the main advantage of such methodology over OSP, which shares many characteristics with the proposed approach. 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 the reconstruction in Fig. 2.

We illustrate this procedure through an example. We start by selecting three 128-dimensional spectra selected from a HyMAP scene acquired over Munich, Germany. The spectra S_1, S_2, and S_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 S_{test}. We greatly reduce the dimension of the system by building an n-dimensional space which uses as basis the adjoint prototypes S_i^k, with $i = 1, ..., n$. In our case this results in a 3-dimensional space, and the potential function will then 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: $S_{test} = aS_1^k + bS_2^k + cS_3^k$ (Fig. 2). We can observe in Fig. 3 the evolution of the potential function and of the prototype pattern retrieved, corresponding to the class "railroad". The final state for the test vector employed coincides with our expectations, as human-made objects spectra are often similar to each other, rather than to natural objects: we expect then the class "railroad" to prevail over the "water" and "grass" classes.

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 (Haken, 1991) 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 full differential equations, usually by selecting the largest initial order parameter (Wang et al., 1993, **?**, **?**, **?**, Crounse and Chua, 1996, Maeda et al., 1999). In this work we approximate then the synergetics equation 1 by its first term, which generates minima along the prototype vectors, only considered at time $t = 0$. Therefore, the higher abundance value defines the classification result.

For a pattern recognition system such as the one described in (Haken, 1991), 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 system 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.

To cope with this problem we propose a classification procedure as follows. In the first step, for each class $C_i, i \in 1...m$, n samples are selected. Then, n classifications are performed, in each of which a different training sample for each class is selected. Afterwards, each pixel p is assigned to a given class on a majority voting basis, i.e. for a pixel p the class C_k is chosen, with $k \in 1...m$ and $k = \arg\max_k C_k(p)$, where $C_k(p)$ is the number of classifications in which p has been assigned to C_k.

Up to this point the weighting for the prototype vectors, represented by the λ parameters, has not been taken into account (i.e., they all have been set to 1 in eq. 1). An award-penalty learning mechanism has been proposed in (Wang et al., 1993) 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.

In (Hogg et al., 1998) 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 often 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.

We propose an improvement over the methodology proposed in (Wang et al., 1993) 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 $N(i)$ 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:

$$\lambda_i = \lambda_i \left(1 + \alpha \frac{FN}{N}\right) \qquad (2)$$

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:
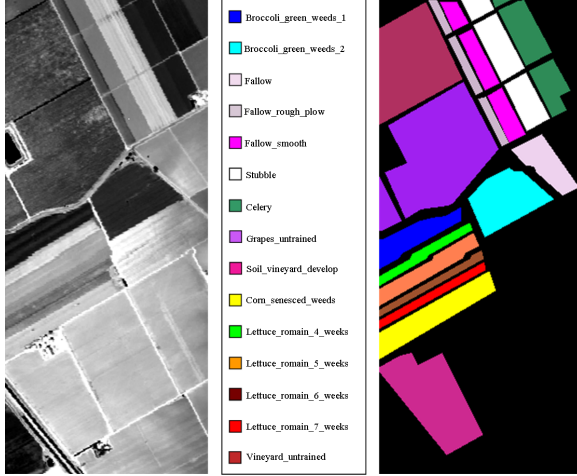
Figure 4: Salinas dataset (left), and available ground truth (right).

$$\lambda_i = \lambda_i \left(1 - \beta \frac{FP - FN}{N} \right) \qquad (3)$$

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. This 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 as chosen at $t = 0$ may now lead to misclassifications, as when the attention parameters are changed the winning vector can be different from v_0. According to (Maeda et al., 1999), 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 $t = 0$.

4 RESULTS

We analyze an AVIRIS hyperspectral scene acquired over the Salinas Valley. The full scene has a size of 512×217 samples with 192 spectral bands in the range 0.4 - 2.5 μm. The water absorption bands removed according to (Plaza et al., 2005), and the spatial resolution is 3.7 meters. 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. 4.

The test dataset has been analyzed with the described methodology. Twenty samples per class have been chosen (see Fig. 5) 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. An exception has been the class "corn", for which 40 samples have been collected. As this class is composed by two different homogeneous areas (see Fig. 4), two different classes were considered, which have been then merged in an unique class as a post-processing step common to each of the carried out classification procedures. For the overall scene results have been additionally improved by an additional step of attention parameters tuning, carried out with two different settings, both for 16 iterations. In the first run, with a similar approach to the one contained in (Wang et al., 1993), we used
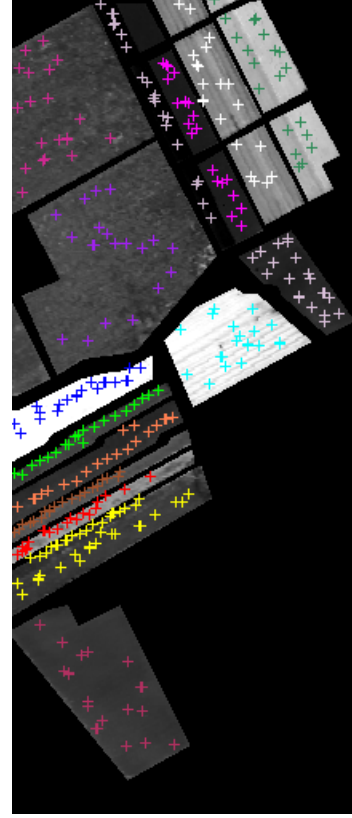

Figure 5: Training data collected over the scene.

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. Fig. 6 shows the classification results for the overall scene. It can be noticed how results benefit from this automatic tuning of the attention parameters, with this improvement being more obvious when the full ground truth is taken as a reference. 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. 7. 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, may converge to some local optimum.

In order to have a fair comparison with other techniques, we performed a classification on both sets and using the same training data using well-known methods in hyperspectral data analysis: the Spectral Angle Mapper (SAM)(Kruse et al., 1993) and the Spectral Information Divergence (SID) (Du et al., 2004). We also considered Support Vector Machine (SVM) (Joachims, 1999), a general classification methodology that 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, using a Gaussian Radial Basis Function (RBF) kernel defined as $K(u,v) = \exp(-\gamma|u - v|^2)$. Due to its natural con-
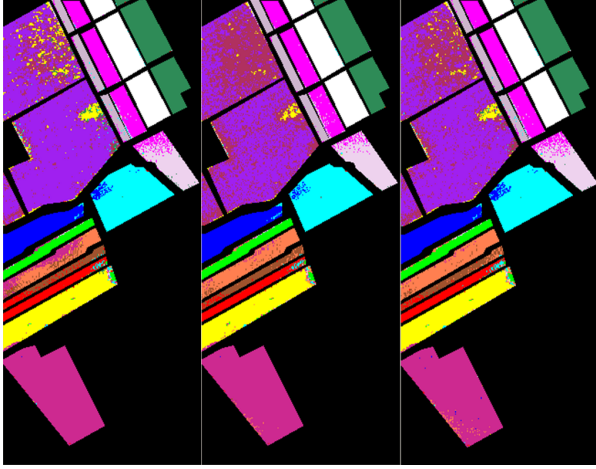
Figure 6: Synergetics classification with different attention parameters settings. From left to right the λ parameters are set as follows: all parameters set to 1, parameters adjusted on the basis of the full ground truth, parameters tuned using a training area for each class.
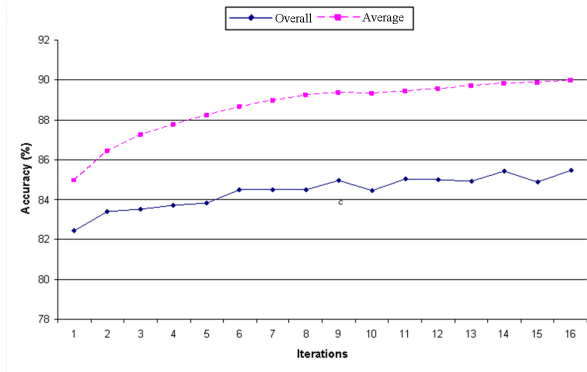

Figure 7: Increase in overall and average accuracy after automatic tuning of the λ parameters.

Table 1: Classification results for the Salinas dataset, in terms of Overall Accuracy (OA) and Average Accuracy (AA).

	OA	AA
Synergetics	82.43	84.96
Synergetics - λ tuned	**85.41**	89.81
Synergetics - λ trained	85.06	88.74
SVM	81.44	**90.6**
SID	83.71	90.2
SAM	78.64	85.4

nection to multidimensional data, SVM have been often applied to hyperspectral data (Demir and Erturk, 2010). Table 1 reports the classification accuracy on the dataset. The overall accuracy (OA) is computed as:

$$OA = \frac{100}{N} \sum_i p_{i,i}, \qquad (4)$$

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

$$AA = \frac{1}{I} \sum_i \frac{100 * p_{i,i}}{N_i}, \qquad (5)$$

where N_i is the number of pixels in class i and I is the total number of classes. The synergetics approach yields the best overall accuracy, achieved with a running time of around 10 minutes.

5 CONCLUSIONS

In this paper we presented a classification methodology for hyperspectral data based on synergetics theory. 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, selected areas belonging to the classes of interest can be used to tune the attention parameters in the synergetics equation, increasing the classification accuracy. Results on the AVIRIS Salinas scene show that such methodology can outperform traditional algorithms employed in the analysis of hyperspectral data.

Such synergetics-based analysis, carried out in the spectral domain spanned by the user-defined prototype vectors, is solely based on the spectral information of the images. Therefore, the information conveyed by the spatial interactions of the pixels in a scene are discarded, limiting the effectiveness of the method.

The results obtained could be then refined by adding a step in which such information is kept to some extent. This could be achieved by traditional methods, such as segmenting the image in a first step (Li et al., 2011) and then performing a region-based classification with the proposed method, with the average spectrum of each segment used as a test vector. Another typical solution is represented by an additional iterative step employing morphological filters (Plaza et al., 2005). Such technique has a twofold advantage, as it spatially regularizes the classification results while also removing salt and pepper noise. The synergetics-based method presented in this work allows looking at the embedding of spatial information in a different way. A more interesting and novel way to carry out such step would be by manipulating the data directly in the prototype vectors space. If we consider the hyperspectral image formation as a Markovian process, we could model the spatial relations between pixels through Markov Random Fields (Solberg et al., 1996), and then weight the projections of a test vector in the prototype vectors space. Each test vector should be displaced towards its neighbours projected in the same vectorial space spanned by the user-defined prototype vectors. 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 belongs, leading to a more homogeneous classification. Results could be further improved by chaining different classifiers instead of performing a simple majority vote. This could also lead to the definition of a hierarchical structure for the classes, where some classifiers are used to identify superclasses and others to separate individual classes from the superclasses.

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.

ACKNOWLEDGMENTS

REFERENCES

Bioucas-Dias, J. and Nascimento, J., 2008. Hyperspectral subspace identification. Geoscience and Remote Sensing, IEEE Transactions on 46(8), pp. 2435 –2445.

Chang, C.-I., 2005. Orthogonal subspace projection (osp) revisited: a comprehensive study and analysis. Geoscience and Remote Sensing, IEEE Transactions on 43(3), pp. 502 – 518.

Crounse, K. and Chua, L., 1996. 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, pp. 134 –137 vol.3.

Demir, B. and Erturk, S., 2010. Empirical mode decomposition of hyperspectral images for support vector machine classification. Geoscience and Remote Sensing, IEEE Transactions on 48(11), pp. 4071 –4084.

Du, H., Chang, C., Ren, H., Chang, C., Jensen, J. and D'Amico, F., 2004. New hyperspectral discrimination measure for spectral characterization. Optical Engineering 43(8), pp. 1777–1786.

Haken, H., 1991. Synergetics Computers and Cognition. Vol. 50, Springer Series Synergetics.

Haken, H., 2007. Synergetics. Scholarpedia 2(1), pp. 1400.

Hogg, T., Talhami, T. and Rees, D., 1998. An improved synergetic algorithm for image classification. Pattern Recognition 31(12), pp. 1893 – 1903.

Joachims, T., 1999. Making large scale SVM learning practical, Universitaet Dortmund Press.

Kaewpijit, S., Le Moigne, J. and El-Ghazawi, T., 2003. Automatic reduction of hyperspectral imagery using wavelet spectral analysis. Geoscience and Remote Sensing, IEEE Transactions on 41(4), pp. 863 – 871.

Keshava, N. and Mustard, J., 2002. Spectral unmixing. Signal Processing Magazine, IEEE 19(1), pp. 44 –57.

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

Li, J., Bioucas-Dias, J. M. and Plaza, A., 2011. Hyperspectral image segmentation using a new bayesian approach with active learning. Geoscience and Remote Sensing, IEEE Transactions on 49(10), pp. 3947 –3960.

Maeda, H., Okada, S. and Irie, T., 1999. 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, pp. 503 –508 vol.3.

Plaza, A., Martinez, P., Plaza, J. and Perez, R., 2005. Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations. Geoscience and Remote Sensing, IEEE Transactions on 43(3), pp. 466 – 479.

Scott, D. W., 2008. The Curse of Dimensionality and Dimension Reduction. John Wiley and Sons, Inc., pp. 195–217.

Solberg, A., Taxt, T. and Jain, A., 1996. A markov random field model for classification of multisource satellite imagery. Geoscience and Remote Sensing, IEEE Transactions on 34(1), pp. 100 –113.

Wang, F. Y., Lever, P. J. A. and Pu, B., 1993. A robotic vision system for object identification and manipulation using synergetic pattern recognition. Robot. Comput. Inegrated Manufacturing 10(6), pp. 445–459.