Application of Self Organizing Maps to multi-resolution and multi-spectral remote sensed images

Ferdinando Giacco[1], Silvia Scarpetta[1,2], Luca Pugliese[2], Maria Marinaro[1,2] and Christian Thiel[3]

[1] Department of Physics, University of Salerno, Via S. Allende, 84081 Baronissi (SA), Italy `giacco@sa.infn.it`
[2] INFN Gruppo Collegato di Salerno, INFM CNISM Unita' di Salerno and IIASS International Institute for Advanced Scientific Studies, Vietri sul Mare (SA), Italy
[3] Institute of Neural Information Processing, University of Ulm, 89069 Ulm, Germany, `christian.thiel@uni-ulm.de`

Abstract. In this paper we investigate the performance of the Kohonen's self organizing map (SOM) as a strategy for the analysis of multi-spectral and multi-resolution remote sensed images.
The paper faces the problem of data fusion, by extracting and combining multi-spectral and textural features. Moreover we address the problem of low-quantity and low-quality of labelled pixels in the training set, investigating a two-step strategy: in the first step (unsupervised training) we use a large unlabelled data set to train a SOM, in the second step a limited number of labelled data is used to assign each SOM node to one informative class. Self Organized Maps are shown to be effective way to discover the intrinsic structure of data.
When the SOM is used as a classifier, as here, a majority voting technique is usually used to associate nodes with informative classes. This procedure allows to obtain a SOM output grid which contains labelled and unlabelled nodes. Particularly in the framework of remote sensing, the unlabelled nodes may be important, since they are associated with new classes present in the image, or with the so-called mixed pixels, which represent an area on the ground composed of more then one land-cover class. Comparing the results of the proposed SOM-based strategy and the results of a supervised network such as SVM we show that the unlabelled nodes of the SOM are associated with high percentage to mixed pixels.

1 Introduction

The dimensionality, the amount, and the heterogeneity of the remotely sensed data available today requires advanced and innovative techniques to extract information and thematic maps useful for environmental monitoring. In the last years innovative methods, not strictly statistical, have been proposed, and among them neural network strategies are very promising [1, 2, 3]. They are especially useful for multisource data, since the whole multiple source data set is usually very difficult to model by statistical methods. Neural Network based classification methods allow to include as input both spectral and spatial (texture and context) features. Textural information was found to improve noticeably the classification ability in many problems, when the spatial scale of the texture is proper [4].

This study focuses on cluster detection, visualization, and land-cover classification of multi-spectral multi-source remote sensed images with two different spatial resolution: an high resolution image registered by IKONOS sensor (4 meters/pixel) and low-medium resolution image registered by ASTER sensor (15 meters/pixel). For the image taken from ASTER we also have a manually generated label map (ground truth dataset) for comparison. The labels indicate different land cover types, as detailed in Chap. 2.

To exploit the high-resolution image, we extracted from the high-spatial resolution images some textural features, using the Gray-Level Co-occurrence Matrix [5] [6], and merged them with spectral information of the middle-spatial resolution image. As we will see below, the combination of spectral and spatial information is especially valuable for land-cover classification systems in the areas with complex landscapes.

To discover and visualize the intrinsic cluster structure of the data, and to see how the extracted features can be related to the land cover classes present in the image, we first apply the Kohonen's Self-Organizing Maps (SOM) with a bi-dimensional lattice, and then using a small number of labelled samples, we merge SOM output nodes into meaningful classes. The SOM play a fundamental role, giving the possibility to detect relationships within large amounts of input patterns and to preserve as well as possible the topology of the original space in a lower dimensional output space. Moreover, the SOM algorithm and other neural approaches are suitable for the incorporation of both spectral and non-spectral data into the classification procedure.

This semi-supervised classification strategy, has some advantages over supervised strategies, when, as in the case under consideration, the available labelled samples have low accuracy or may be non-exhaustive.

Indeed, sensing applications the number of available labelled training samples is not large, since gathering reliable prior information is often too expensive both in terms of economic costs and time.

Besides, concerning the quality of training data, there are many problems in remote sensing applications, from the problem of mixed pixels, to the problem of the correlation among training patterns taken from the same area, to the exhaustive definition of the classes. Non exhaustive definition of the classes present in the image may happen when there is not enough a priori information on the territory composition, and in such a case an unsupervised strategy such as SOM, which does not use any a priori information, may be useful to detect the new classes. Mixed pixels, which are often abundant, are pixels that comprised more then a single class. As a pixel is an arbitrary spatial unit, it may represent an area on the ground which comprises more than one discrete land cover class, for example water and pine wood. Alternatively, it may happen that the classes overlap gradually with many areas of mixed class compositions, particularly near imprecise boundaries.

In this paper we address these problems investigating a two-step semisupervised strategy, which makes use of unlabelled data to train a Self Organized Map, and uses a limited number of labelled data to associate the nodes to informative classes. When the SOM is used as a classifier, a majority voting technique is usually used to associate its nodes with informative classes. This technique, however, cannot guarantee that every node in the output layer will be labelled, and thus will produce unclassified pixels in the final map.

We focus on these unlabelled nodes, that is nodes of the SOM map with which none of the labelled pixels is associated to.

These unlabelled nodes come from the presence of mixed pixels or non-exhaustive class definition. Our approach associate pixels with a high degree of mixing with unlabelled SOM nodes. Pure pixels fall into other nodes. The distinction between pure and mixed pixels is carried out both visually and with the help of a SVM strategy.

The reminder of the paper is organized as follows: in Sect. 2 we describe the multi-resolution and multi-spectral images and the feature extraction process. From the high-resolution image, textural features are extracted and fused with spectral features provided by the lower-resolution image. In section 3 the SOM is applied, nodes are labelled with a majority vote criterion, and results are discussed. In sec 4 the relations between SOM unlabelled nodes and the rejected pixels of a SVM with rejection threshold is studied. In sec 5 conclusions are drown.

2 Dataset and pre-processing

Satellite multi-spectral data from ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) and Ikonos 2 are used for the experiments. We considered the first nine bands of the Aster data set (from November 2004), spanning from the visible (bands 1-3) to the short wave infrared spectral region (bands 6-9). All the bands were georeferenced for this study and resized to a resolution of 15 m/pixel.

To determine the spatial information we realized some textural measures (detailed in Chap. 2.1) working on Ikonos 2 images from January 2004. The spatial resolution of the data is 4 meters for the visible bands (blue, green, red, near-infrared) and 1 meter for the panchromatic band (grey-level image). In our work we resized also the visible band at 1 meter/pixel resolution.

Both Aster and Ikonos images acquired have been radiometrically and geometrically corrected, cross-calibrated and co-registered to allow multi-scale analysis.

The interested area is located in the southern part of the alluvial plain of Salerno gulf, in the southern part of Italy. The area is primarily an agricultural land (no mountains are situated in the region) with major crops including corn, soybeans, grain, tobacco and canning vegetables. The coastal zone is mostly occupied by pine wood and urbanization consist of urban fabric and other human artefacts, especially for farming practice.

For the Aster image we have a manually generated label map (labelled dataset) for comparison purposes. The labels indicate different land cover types. Seven classes are considered: vegetated agricultural fields, buildings, pine forest, urban green, sea-shore, not vegetated agricultural fields, and water. Among the total number of 236985 pixels, each with a spatial resolution of 15 m, 1657 are labelled. The SOM will be trained on the whole dataset, composed of 236985 vectors, each with 11 dimensions (9 Aster spectral bands and 2 textural features coming from IKONOS). In the second step, the SOM nodes are labelled with majority vote, using about 2/3 of the labelled dataset. Specifically, we divide the labelled data set into a classification-set (1029 pixels), used to label the SOM nodes, and a test-set (628 pixels) used to evaluate the percentage of correct classification of our strategy. The composition of the labelled set is shown in the following table.

Table 1. The available set of labelled samples. The land cover classes are: Vegetated land (1), Built up area (2), Pine wood (3), Urban green (4), Greenhouse (5), Not vegetated land (6) and Water (7). Training set is used to label the nodes of the SOM with a majority vote strategy (see sect. 3), while the test set is used to evaluate the percentage of correct classification and the confusion matrix (fig. 1 and 2)

Class number	1	2	3	4	5	6	7	TOT
Labelled set	249	188	226	251	273	233	238	1657
Training set	145	90	162	159	166	144	163	1029
Test set	104	98	64	92	106	89	75	628

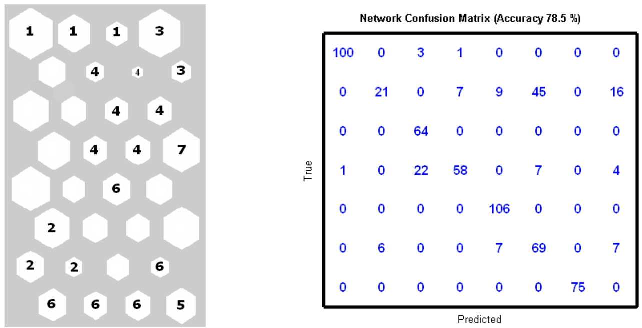
2.1 Feature extraction

In order to exploit the information that are available in two kinds of images we extract textural features from Ikonos, and add it to spectral features obtained from Aster.

Different textural features extracted from Ikonos images were introduced, in addition to the Aster spectral data, to take account for spatial information. The textural features were computed on two different Ikonos data: the panchromatic band (a grey-level imagery of 1 meter per pixel resolution, sensitive to all visible radiation) and the band ratio between near-infrared and red (4 meters per pixel resolution, resized at 1 meter per pixel), widely used in remote sensing to eliminate various albedo effects.

The spectral features were obtained from the well known Grey-Level Co-occurrence Matrix (GLCM), widely used in land-cover mapping [7]. The Gray

Fig. 1. On the left: SOM with only spectral features (9 Aster bands), without Ikonos information. Labelling of the SOM output nodes is based on a majority vote criterium looking at the training samples. On the right: confusion matrix resulting from labelled SOM. Columns, from left to right, corresponds to classes from 1 to 7, while the 8th column refers to pixels belonging to unlabelled nodes. As shown here, 4% of the test set, that is 27 pixels, are unclassified, since they belong unlabelled nodes.



Level Co-occurrence Matrix is a standard technique for extracting texture characteristics: distance as well as directional relationships among grey levels are summarized in a GLCM, obtaining a measure of the probability of occurrence of two grey levels separated by a given distance in a given direction. GLCM has been used successfully in a variety of applications, including land-cover mapping, crop discrimination and forest studies [4].

A moving window of 15 × 15 Ikonos pixels has been used in the computation of GLCM matrix and variance, in order to add textural information on a window having the same dimension as one Aster pixel [8]. In the computation of the GLCM, data are typically scaled to some fairly modest range of integers, (for example 0-7 in this work, such that the GLCM is a 8 × 8 matrix). After the GLCM is generated for each direction (horizontal, vertical, left diagonal, and right diagonal), the statistical measures are extracted and then the four directions are averaged to remove directional effects; this last choice is due to the absence of preferred direction in the geometry of investigated land-cover classes. We extracted from the GLCM four different statistical measures: Energy, Contrast, Homogeneity and Correlation [9].

3 Application of Self-Organizing Maps

The Kohonen's Self-Organizing Maps (SOM, [10]) used in this work is trained iteratively with a sequential algorithm, the distance measure chosen is the Euclidean one, the lattice is bi-dimensional with a local hexagonal structure, the weight vectors are updated with a gaussian neighborhood kernel, and the number of nodes is 32. All the setting parameters are empirically determined, we particularly have not found any critical dependence of the clustering procedure on the number of nodes, as well as on the lattice local structure.

The SOM algorithm achieves two important goals: (a) a clustering of the input data into nodes; and (b) a local spatial ordering of the map in the sense that the prototypes are ordered on the lattice such that similar inputs belong to topographically close nodes. Such an ordering of the data facilitates the understanding of data structures.

The first stage of the experiment was conducted with only the 9 Aster spectral bands as input for the SOM network. After the learning phase, an arrangement of the data on a 32-nodes output lattice is obtained. Looking a the distribution of the labelled classification subset into the SOM nodes, we found that pixels of different land-cover classes mostly falls in different nodes, and that some nodes are not associated to any of the labelled pixels of the classification subset. It means that the complete unlabelled dataset used to train the SOM contains also vectors whose characteristics are different from the ones of the seven land-cover classes. Such vectors are not present into the labelled dataset that we have used for the classification stage.

In the classification stage, we associate a label with those SOM nodes into which at least C labelled pixels fall. Association is accomplished using a majority vote technique. Setting threshold C equal to 3, there are 11 nodes of the lattice which cannot be associated with any label. To evaluate the classification performance, we compute the confusion matrix, on the test set, as shown in Fig. 1. The confusion matrix is computed using 8 classes, the 7 land cover classes and the unlabelled-nodes class.

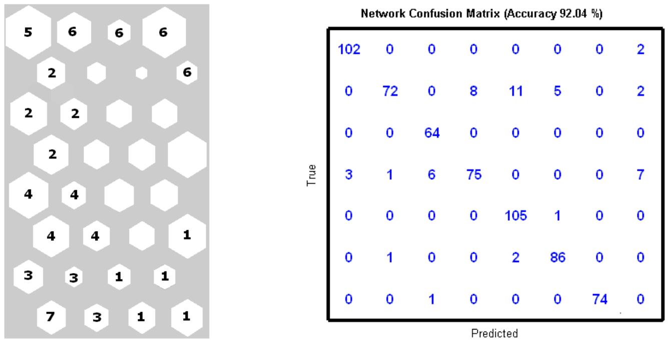
The percentage of correct classification achieved in this way is low, mainly because in this first experiment we used only the ASTER spectral bands, without exploiting the textural information extracted from Ikonos. The overall percentage of correct classification is 78.5% and the confusion matrix shows that main errors occurs for class 2 (built-up area) and class 4 (urban green).

When the two-stage strategy described above is applied while adding textural information to the spectral one, and in addition merging information at different spatial scales, significant improvements are achieved.

Among all the possible textural features, we selected the ones who better discriminate among the classes with major overlap in the spectral SOM map [11]. Therefore, we looked for texture measures which were more able to discriminate class 2 (built-up area) and 4 (urban green) from the remaining classes.

Introducing these two new inputs, in addition to the 9 Aster spectral bands, the resulting SOM map is shown in Fig 2. As described in the previous section, we project the labelled training subset on the SOM lattice, and label the SOM nodes with a majority vote criterium (with C=3) as before. We see that the confusion matrix, computed on the test subset, is considerably improved (compare Fig. 2), with an overall accuracy percentage of 92.04% on the test set. Among the 32 nodes, 10 are unlabelled, 6 of them are totally empty of labelled data and 4 of them contain less then C labelled pixels. The meaning of the unlabelled nodes is investigated in the next section.

Fig. 2. On the left: SOM trained with textural information, extracted from the Ikonos bands,in addition to the spectral ones (9 Aster bands). On the right: confusion matrix resulting from the labelled SOM.



4 SOM Unlabelled nodes and mixed pixels

In this section we compare the two-stage strategy based on the SOM presented above, with the results of a SVM used as a reference. The goal is to better understand the SOM results and particularly to characterize the unlabelled nodes.

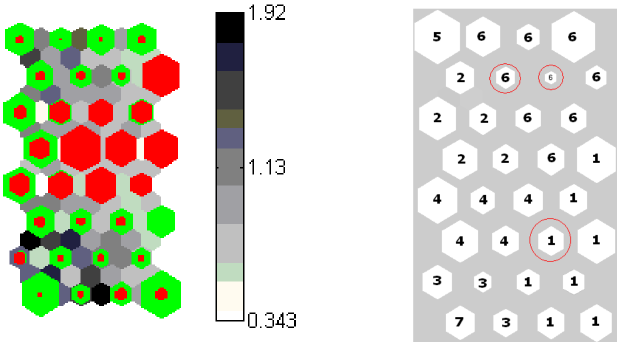
The baseline classification of the pixels was accomplished using Support Vector Machines (SVMs, introduced in [12]) with a linear kernel, and a weighting factor C for the slack variables of 0.1. Other than in previous ways [13] we employed the 1vs1-architecture, where a separate SVM is trained for each pair of classes. The distance-valued outputs of each of those machines were then converted into probabilities using a Fermi function, whose multiplicative variable was set to 2 uniformly for all machines. Finally, those pairwise probabilities were subjected to the iterative procedure of pairwise coupling according to the Bradley-Terry model [14], as suggested in [15]. This results in mixed answers for each pixel where the probabilities for each class are more distinguishable than with the initial estimation.

In the following we analyze the projection of the data onto the SOM lattice, using the output of the SVM as a fuzzy-label for pixels, where y_k^n represent the membership of pixel x^n to class k.

First of all, associating each pixel x^n with the class k which has the highest value of y_k^n, $k = 1 \dots l$, we get a "hard" labelling. Evaluating the performance of the hard SVM classification, using the confusion matrix on the set set, we find that it provides a very high performance, with an overall accuracy of 95.4%.

However, not all pixels have a high value of membership towards a specific class as output of the SVM. Usually one can introduce a threshold T, such that each pixel x^n is assigned to the class k if the SVM output vector $y_k^n > T$ and $y_k^n > y_j^n$ holds for all $j \neq k$. Indeed if we put such a rejection threshold on the SVM output, we see that, for example when $T = 0.3$, a total of 0.14% of the pixels is rejected. Indeed in our images many pixels are mixed pixels, especially when a part of an artificial surface (or built-up area) is mixed with crops, and

Fig. 3. On the left: Distribution of pixels rejected by the SVM classifier, on the SOM lattice. The hexagons depicting the (red/green) SOM nodes are colour-coded (with grey levels, as given below) to represent the distance between the node centres. The size of the central (red) hexagons is directly related to the number of rejected pixels. On the right: SOM labelled lattice, where each SOM node is associated with one of the 7 land cover class, taking into account the SVM hard answers on the whole dataset (236985 pixels). The circles emphasize the unlabelled nodes of Fig. 2 that, owing to the SVM analysis, could be assigned to one of the 7 pure land cover classes.



these mixed pixels than have low memberships in the 7 pure land cover classes. Fig.3 (left side) shows the distribution of rejected pixels on the SOM output grid, where the dimension of the central (red) hexagons is directly related to the total number of rejected pixels. By looking at the position of the big red hexagons in comparison to the white unlabelled nodes in Fig. 2, we find a large correspondence: at least six of the unlabelled nodes (in the middle of the grid) have large quantity of SVM-rejected pixels (red hexagons).

Also shown in Fig. 3 (right side) is the SOM lattice, where each SOM node is associated (using a majority vote criterion) to one of the 7 land cover classes, exploiting the hard SVM labels on the whole dataset. One can see then, when forced to have an hard classification on each pixel, pixels with same SVM-class form compact clusters on the SOM lattice.

Considering the pixels belonging to the unlabelled nodes of Fig. 2 (left) as unclassified pixels (class 8), we evaluate their distribution with respect to the pixels rejected by the SVM-classifier. Table 2 shows the confusion matrix between the answer provided by the SOM and the one provided by the SVM, evaluated on the whole unlabelled dataset. Columns, from left to right, corresponding to SOM information classes from 1 to 7, while the 8^{th} column refers to the unlabelled nodes. Matrix of table 2 shows an agreement of 86% on the 7×7 information classes, and the 8^{th} row shows that the 74% of the pixels rejected by the SVM (i.e. 6422 pixels) falls into the unlabelled nodes of the SOM.

However, in the 8th column there are high percentages of pixels of SVM-class 1, 4 and 6 assigned to the unlabelled nodes.

To understand this result one has to consider that the number of labelled pixels used in the classification of the SOM nodes is very small (see Table 1), thus some of the unlabelled nodes remain unlabelled due to the absence of sufficient labelled data of the corresponding cover class. This is confirmed by comparing the grid of Fig. 3 with the one in Fig. 2. It is evident that at least 3 nodes, circled in Fig. 3, are unlabelled due to the lack of labelled training data.

Table 2. Results of the classification procedure on the whole dataset for the seven investigated classes described in the text. Each column represents the answers provided by the SOM labelled nodes plus the unlabelled ones (see Fig. 2), while each row represents the SVM output for the same information classes plus the rejected pixels.

Class number	1	2	3	4	5	6	7	Unlabelled
1	36109	8	31	841	0	2195	0	15024
2	0	32211	0	4219	2140	35	0	2431
3	9	0	8105	181	0	0	504	0
4	2087	117	1583	26199	0	2310	6	11017
5	64	872	10	42	5128	1277	0	242
6	0	3810	4	35	1869	27635	0	22054
7	0	0	59	0	0	0	15294	0
Rejected	117	417	209	662	0	892	2	6422

5 Conclusions

This work focuses on the classification of satellite multi-spectral images starting from a limited number of labelled data through a two step semisupervised strategy based on the Self Organized Map algorithm. The unsupervised clustering provided by the SOM shows good capability to separate the 7 investigated land-cover classes on a two-dimensional output grid, where each node of the grid is labelled according to a majority vote technique. The resulting land cover map shows an overall accuracy of 92.05% with respect to a labelled test set.

Moreover, we investigate the meaning of the unlabelled nodes on the SOM grid exploiting the results of a Support Vector Machines with a rejection threshold.

According to our results, while some unlabelled nodes are related to the lack of sufficient number of labelled samples, the large part of the unlabelled nodes represent mixed pixels, that is pixels that have characteristic features of several land-cover classes.

References

1. P. M. Mather. *Computer Processing of Remotely-Sensed Images*, Jonh Wiley and Sons, Chichester, 1999.

2. T. Villmann , E. Merenyi and B. Hammer, Neural maps in remote sensing image analysis, *Neural Networks*, 16: 389-403, 2003.
3. G. M. Foody, Status of land cover classification accuracy assessment, *Remote Sensing of Environment*, 80: 185-201, 2002.
4. M. A. Shaban and O. Dikshit, Improvement of classification in urban areas by the use of textural features: the case study of Lucknow city, Uttar Pradesh, *International Journal of Remote Sensing*, 22: 565-593, 2001.
5. S. E. Franklin and D.R. Peddle, Classification of SPOT HRV imagery and texture features, *International Journal of Remote Sensing*, 11: 551-556, 1990.
6. P. Gong, D. J. Marceau and P. J. Horwarth, A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data, *Remote Sensing of Environment*, 40: 137-151, 1992.
7. R. M. Haralick, K. Shaunmugam and I. Dinstein, Textural features for image classification, *IEEE Transactions on Systems, Man and Cybernetics*, Vol 3, pages 610-621, 1973.
8. D. J. Marceau, P. J. Howarth, J. M. Dubois and D.J. Gratton, Evaluation of the grey-level co-occurrence method for landcover classification using SPOT imagery, *IEEE Transactions on Geoscience and Remote Sensing*, Vol 28, pages 513-519, July, 1990.
9. R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision: Vol. 1*, Addison-Wesley, 1992.
10. T. Kohonen. *Self-Organizing Maps. Series in Information Sciences*, Springer Verlag, 1997.
11. F. Giacco, L. Pugliese, S. Scarpetta, and M. Marinaro: Application of Neural Unsupervised Methods to Environmental Factor Analysis of Multi-spectral Images with spatial information, *Proceedings of Signal Processing, Pattern Recognition, and Applications*, SPPRA, ACTA press, Zurich (2008)
12. V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, 1995.
13. C. Thiel, S. Scherer and F. Schwenker, Fuzzy-Input Fuzzy-Output One-Against-All Support Vector Machines, *Proceedings of the 11th International Conference on Knowledge-Based and Intelligent Information Engineering Systems KES 2007*, 2007.
14. A. R. Bradley and M. E. Terry, Rank Analysis of Incomplete Block Designs: The Method of Paired Comparisons, *Biometrika*, Vol 39, pages 324-345, 1952.
15. T. Hastie and R. Tibshirani, Classification by Pairwise Coupling, *The Annals of Statistics*, Vol 26, Number 2, pages 451-471, 1998.