

## An efficient neural classification chain of SAR and optical urban images

P. GAMBA

Department of Electronics, University of Pavia, Via Ferrata, 1, 27100 Pavia, Italy; email: gamba@ele.unipv.it

and B. HOUSHMAND

Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109-8099, USA

(Received 9 February 1999; in final form 11 November 1999)

**Abstract.** In this paper a suitable neural classification algorithm, based on the use of Adaptive Resonance Theory (ART) networks, is applied to the fusion and classification of optical and SAR urban images. ART networks provide a flexible tool for classification, but are ruled by a large number of parameters. Therefore, the simplified ART2-A algorithm is used in this paper, and the neural approach is integrated into a classification chain where fuzzy clustering for merging of classes is also considered. The interaction between the two methods leads to encouraging results in less CPU time than classification with fuzzy clustering alone or other classical approaches (ISODATA). Examples of classification are provided using C-band total power AIRSAR data and optical images of Santa Monica, Los Angeles.

### 1. Introduction

The remote sensing analysis of urban structures is gaining increasing interest for both the academic and industrial sectors because of the number of applications that it may provide, such as settlement detection, population estimation, mappings of land use and changes, assessment of urban activities on the landscape, and so on (see Henderson and Xia 1997 or Xia 1996 for a detailed list). In particular, urban data analysis by airborne sensors is now already operational and provides observation at flexible spatial and temporal scales without interference with the urban environment. Moreover, it enables the monitoring of the same environment at rates that are clearly impossible for a study 'on the ground'.

Many different sensors have been used for this task, from cameras to radar, from multispectral mappers to hyper-spectrometers. However, if we may say that the photointerpretation of urban areas is a relatively well known problem (Gruen *et al.* 1995, Gruen 1998), the use of a high definition SAR to obtain similar results still holds some open questions. They attain to the useful information which can be extracted, and to the results of a characterization based on the fusion of the information from both sources (optical and radar). In Weydhal *et al.* (1995), for instance, it has been observed that SAR images may improve the identification of buildings and

hard targets within a built-up area. Similarly, in Dousset (1995) a detailed comparison between the classes of detectable objects in Synthetic Aperture Radar (SAR) and SPOT images of Los Angeles reveals that SAR detects residential areas, while it provides limited information for the estimation of soil moisture in undeveloped blocks and parks. Hyperspectral sensors can provide accurate land cover classifications. For urban area analysis, the compilation of spectral signatures of urban areas is currently ongoing (Hepner *et al.* 1998).

In this paper an unsupervised classification is proposed for segmentation of urban aerial optical and SAR images. The classifier is based on the use of competitive neural networks to exploit the spectral and spatial properties of optical and SAR data. The choice of a neural network approach relies on the already proven (Benediktsson 1990, Paola and Scowengerdt 1995) consideration that they have a superior performance over maximum likelihood classifiers due to their nonparametric inherent structure. When the classes to be identified are strongly overlapping in the feature space (i.e. we have mixed pixels classification problems), the neural analysis provides fewer errors. In other words, if we do not have any *a priori* knowledge about the data, quantitative assessments have shown that neural networks result in a better classification map. However, the main disadvantage of the feed-forward network structure is in terms of training time: it could take very long to get reliable results. Our approach, instead, is based on an unsupervised approach, directly comparable in term of CPU time consumption, as shown in the result section, to other classical clustering algorithms (ISODATA and Fuzzy C-Means).

Other interesting kinds of classifier undoubtedly useful in the urban context are knowledge-based ones (see, for instance, Dobson *et al.* 1996), that exploit the *a priori* information available for a given environment to organize hierarchical decision rules to differentiate land-cover classes. Even if these kinds of procedures are very complex and usually extremely useful for well-defined applications, they could be in some way connected to neural classifiers. Indeed, they could be considered systems that learn from their experience to understand the environment they are working on, and that use the acquired knowledge to perform better and better.

Many neural approaches have been proposed in the technical literature for remote sensing image analysis (Ritter and Hepner 1990, Liu and Xia 1991, Bishop *et al.* 1992, Hermann and Khazenie 1992, Dreyer 1993). A few were designed to deal with SAR data (Hara *et al.* 1994, Chen *et al.* 1996, Tzeng and Chen 1998), and a smaller number have been applied to urban environments (Dousset 1995, Hrikkonen *et al.* 1997, Silva and Caetano 1997). In this paper we analyse in detail the fusion of SAR and optical images by means of competitive neural networks based on the Adaptive Resonance Theory (ART). Competitive neural networks have been used for data classification in Baraldi and Parmiggiani (1995), Levine and Penz (1990) and Carpenter *et al.* (1991). In Baraldi and Parmiggiani (1995) the authors pointed out the large number of parameters governing the behaviour of this network; the solution was found in a simplified ART neural network (called Simplified ARTNN or SARTNN). The same kind of analysis was performed in Levine and Penz (1990), where the ART 1.5 network was introduced. We instead integrate the ART2-A algorithm, already introduced in Carpenter *et al.* (1991) to simplify the ART implementation, into a more efficient classification chain taking into account both spectral and spatial analysis. The proposed classification chain is presented, and the results indicate that it can achieve better results than simpler ART-based approaches. Moreover, in this work we analyse the fusion of airborne SAR and aerial photo

information by a neural classification approach, in order to show the capability of the proposed classification chain to combine the data and to extract valuable insights into the urban structure.

## 2. The neural detection algorithm

In this paper we propose an unsupervised neural approach to the classification of urban multisensor images. The aim is to simplify the detection of the urban features by means of an automatic analysis of the data. We employ the neural competitive algorithms for urban classifications by looking for a consistent grouping of the input data in clusters where, possibly, erroneous and/or noise-affected values have been discarded.

### 2.1. The ART networks

ART networks are self-organizing nets based on biological concepts and implemented by means of differential equations (Carpenter and Grossberg 1987a,b). They were developed to provide a solution to the *stability–plasticity dilemma*. In other words, the proposed ART network structure is designed to find relationships between input patterns in an adaptive way (*plasticity*), but without losing the knowledge obtained by means of the previous training phases (*stability*). This result is accomplished by a bi-directional neural network structure that compares iteratively the input pattern and the most similar one in its memory searching for a *resonance* between them. When the difference is too large, the network introduces a new prototype, therefore maintaining the results of the training steps already performed and, at the same time, introducing the new knowledge brought by the different input pattern.

The ART network family was started by the ART1 network designed to recognize binary patterns: two layers connected by bi-directional weights compose its structure. The first layer introduces a query to the second one, computed by means of a set of *bottom-up* weights, searching for a correspondence. The second layer responds to the query by computing a possible input pattern compatible with the memory acquired by the preceding inputs and the new situation (*top-down* weights). The procedure iteratively continues until a stable situation is found and a match is provided between input and memory (i.e. the network system is *resonating* at a stable point). Resonance is detected by comparison with a *vigilance parameter* that rules a test constituting the *orienting subsystem* of the network, so called because it orientates the networks toward the above-mentioned resonance. The input and internal layers (called *input representation* and *category representation fields*, respectively) constitute the *attentional subsystem*, whose name is due to the fact that it pays attention to the input and searches for the system *memory* with the greatest similarity (see Carpenter and Grossberg 1987a).

To improve the previously presented structure, the ART2 network was introduced in Carpenter and Grossberg (1987b): ART2 is tailored to handle analogue input patterns by means of the same concepts as ART1. To understand its behaviour, consider the most usual ART2 structure in figure 1 (many different architectures with equivalent network dynamics are possible). The network is defined again by means of the orienting and attentional subsystems. The *input representation field*  $F_1$  and the *category representation field*  $F_2$  that constitute the attentional subsystem have a common structure aimed to normalize (within a loop) their in–out vectors. Network dynamics proceeds as follows.

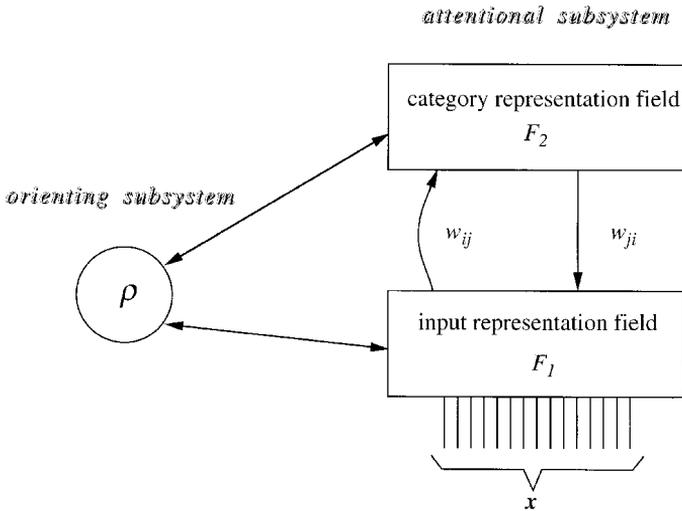


Figure 1. A very simple scheme of the architecture of an ART2 network. The dynamics of the net and the role of all the parameters in this figure are explained in the text (see §2.2).

First the network acts, in response to a generic input pattern  $x = (x_1, x_2, \dots, x_i, \dots, x_N)$ , by progressively enhancing the pattern contrast (while maintaining the normalization). Then, this pattern progressively enhances the value (initially set to 0) of each top-down weight  $w_{ij}$  (see next subsection). The maximum combination of inputs and weights activates a pattern in the  $F_2$  field, called *category pattern*, whose function is to be representative of a certain group of inputs. If the activated pattern is initially uncommitted, i.e. not defined, it becomes proportional to the input, and when the activation overcomes the vigilance parameter  $\rho$ , a reset signal is sent. On the other hand, if the category pattern was already committed (i.e. there is memory of other inputs activating this particular pattern), a similarity test is taken as input to the orienting subsystem. If the test is passed, the input pattern is assigned to that category, and the category pattern is updated taking into account the new information. Otherwise, a new category prototype is selected. The network performance is further complicated by the dynamics of the weight connecting  $F_1$  and  $F_2$ . Bottom-up and top-down weights change following a so-called *outstar* or *instar learning* (basically, an exponential growth or decay) ruled by two different time constants. Furthermore, their interaction is sent to the vigilance test following a third, different time constant.

## 2.2. The ART2-A algorithm

We note from the preceding short outline of ART2 networks that one of the most problematic aspects is the complexity of the parameters that affect their performance. It is not clear what are the effects of each parameter on the overall behaviour of the net. That is, how to change parameters in order to obtain a given performance. In particular, in the ART2 structure there are many parameters and it is difficult to identify the contribution of each parameter to the final system output. The problem has been clearly stated in Baraldi and Parmiggiani (1995) where the original structure is simplified by considering a monodirectional weighting in the *fast learning* limit (any input gives in one time step a classification result), and a qualitative performance

assessment is performed. A different solution to the same problem was proposed also in Levine and Penz (1990), where the so-called ART1.5 network was introduced. However, a simplified version of the ART2 network was already studied in Carpenter *et al.* (1991), where the ART2-A algorithm was introduced. The major advantage of this enhancement is that it has the same behaviour as the original ART2 network in fast learning mode: in this situation ART2-A is quicker by two or three orders of magnitude than ART2. Moreover, ART2-A works well also in a larger range of situations where *intermediate learning* is required, thanks to its *fast commitment* feature, i.e. the possibility to define very rapidly the activation weights and choose if there is the need to activate a new node in the category representation field  $F_2$ .

In Carpenter *et al.* (1991) it is also shown that the ART2-A architecture allows results to be obtained that are comparable with those of the original ART2, while simplifying the processing steps. In particular, only four parameters should be considered (called  $\alpha$ ,  $\beta$ ,  $\rho$  and  $\theta$ , see next paragraphs), even if the original structure is maintained, in particular the bi-directionality of the connections. Indeed, this advantage comes from the fact that ART2-A is not a new network structure, but an algorithm that directly provides the results of the ART2 learning process when a hypothesis on the net behaviour can be done.

The following is a brief excursus on the ART2-A algorithm and working implementation. Interested readers should refer to Carpenter and Grossberg (1987a) or (1987b) for mathematical details.

The algorithm can be subdivided into four steps (see also the flow chart in figure 2):

1. Given an  $n$ -dimensional input vector  $x$  (different from a null pattern), the output of the representation field  $F_1$  (corresponding to the input to the  $F_2$  field) is computed by the nonlinear normalized relationship:

$$x_1 = N(f(N(x))) \tag{1}$$

where  $N(\cdot)$  represents the normalization function ( $x/\|x\|$ ),

$$f(x_i) = \begin{cases} x_i & \text{if } x_i > \theta \\ 0 & \text{otherwise} \end{cases} \quad i = 1, \dots, n \tag{2}$$

and  $\theta$  belongs to  $(0, 1/\sqrt{n})$ .

2. Then,  $x_1$  is fed into the  $F_2$  field and activates all its nodes (numbered by the subscript  $j$ ) by means of the vector  $T$  (that combines the effects of both the bottom-up and the top-down weights), whose components are

$$T_j = \begin{cases} \alpha \sum_i x_{1i} & \text{if } j \text{ is an uncommitted node} \\ x_1 \cdot w_j & \text{if } j \text{ is a committed node} \end{cases} \tag{3}$$

where the positive constant  $\alpha$  is upper bounded by  $1/\sqrt{n}$  and  $w_j$  are the top-down weights, each one initially set to a null vector.

3. The  $F_2$  node with maximum  $T$  (say, the  $J$ -th node) is activated (if there are more than one, a random choice solves the conflict). If, furthermore, this  $J$ -th node was uncommitted or it was already committed and

$$T_j \geq \rho \tag{4}$$

the choice remains stable. Otherwise, a new node is considered.

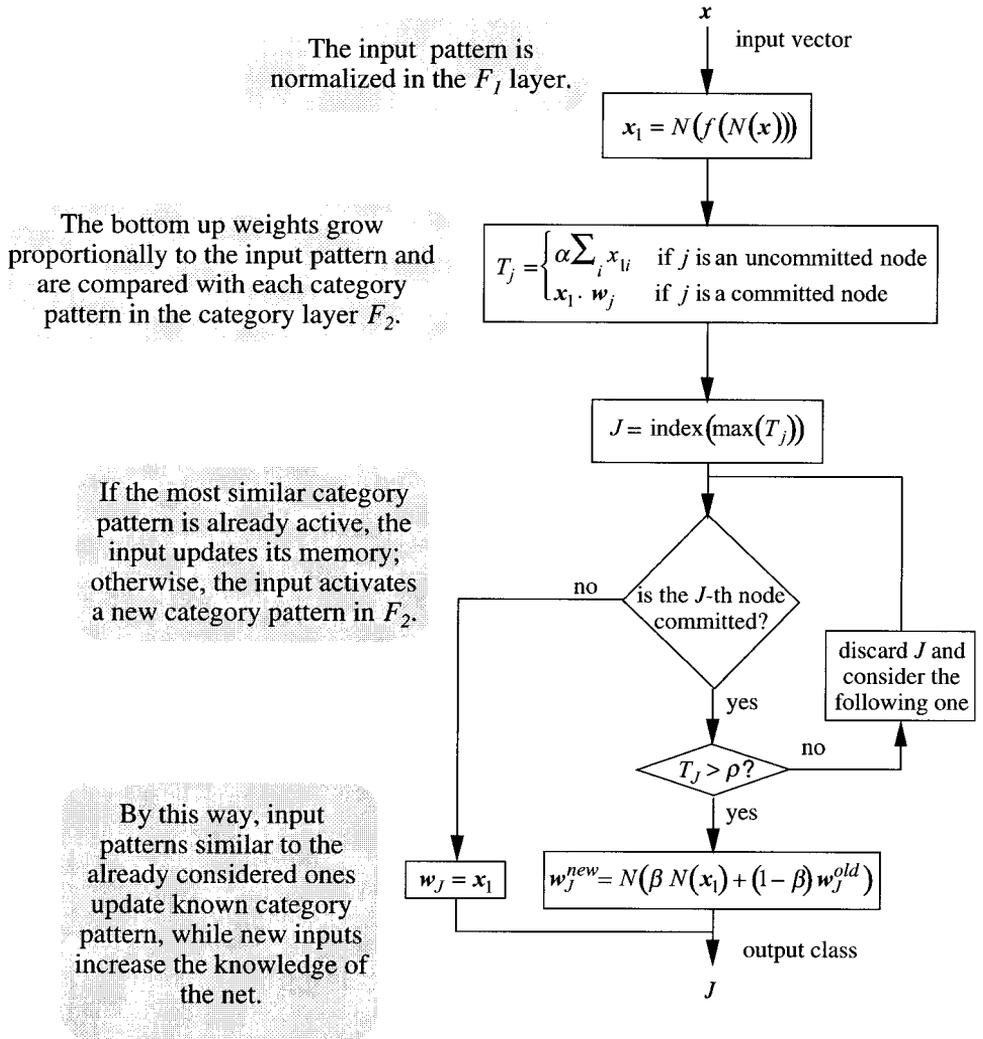


Figure 2. A simplified flow chart of the ART2-A algorithm (see also figure 1 and the text).

4. Finally, at the end of the input presentation,  $w_j$  is updated to  $w_j'$  where

$$w_j' = \begin{cases} x_1 & \text{if } J \text{ is an uncommitted node} \\ N(\beta N(x_1) + (1 - \beta)w_j) & \text{if } J \text{ is a committed node} \end{cases} \quad (5)$$

where  $\beta$  belongs to  $[0, 1]$ .

First, we observe that there are now only four parameters governing the equations, i.e.  $\alpha$ ,  $\beta$ ,  $\rho$  and  $\theta$ . The parameter  $\alpha$  refers to the activation of the  $F_2$  field: the larger its value, the more easily new nodes of  $F_2$  are committed. Parameter  $\beta$  is related to the learning stage of the network: it rules what part of the old weight value is retained while updating the *knowledge* of the net. Parameter  $\rho$  is the *vigilance* parameter. It defines a threshold for activation of a committed node to assign the input pattern to the corresponding class: small  $\rho$  values cluster the inputs in few class, a larger value results in more new committed nodes after each input set.

Finally,  $\theta$  is critical for noise suppression: the larger its value, the larger the difference among input patterns needed to consider them different. We will show in the next section, while applying this algorithm to our classification problem, how these parameters affect the overall behaviour of the proposed classification chain.

### 2.3. The classification chain

The most recent example of the use of ART networks has been proposed for a fast and reliable analysis of satellite images taken over a town (Lisbon, Portugal) (Silva and Caetano 1997). For this application, ART1 networks were used for a preliminary spectral analysis, followed by a spatial one performed by an ART2 network. The results showed a remarkable dependency of the number of classes from the network parameters, especially the vigilance parameter (Carpenter *et al.* 1991), due to the large variability of the backscattered field in urban areas.

In this work, we build a classification chain involving two steps (the structure of the chain is depicted in figure 3): an ART2 network performs both the spectral and spatial classifications. The ART1 structure needs binary inputs, and this is consistent with the digital images to be analysed. However,  $n \times k$  input nodes are required for an  $n$ -band multispectral image coded with  $k$  bits. To reduce the inputs and simplify the network, in Silva and Caetano (1997) it is suggested to subdivide the feature space in an arbitrary number of intervals. Of course, the best results are obtained considering a subdivision that takes into account the relative occurrences of the features, but it is impossible to know them *a priori*. Without this choice the ART1 results are worse than ART2 ones even after optimizing the values of the parameters (see next section for a quantitative assessment). An improvement in the classification chain depicted in figure 3 is the introduction of a clustering algorithm to merge the network outputs after the spectral and spatial ART2 classification. This *double-face* approach (neural network and fuzzy clustering) is due to the necessity to balance the need to retain the extreme variety of urban environments, and simultaneously to overcome the noise errors. Indeed the noise effect (especially in SAR images) is that the ART outputs splits the same class in more than one group. The clustering is performed by means of a very fast Fuzzy C-Means (FCM, Davé and Krishnapuram 1997) procedure.

With this processing chain, input data are first classified by a neural approach with respect to their spectral appearance in the optical and microwave regions of

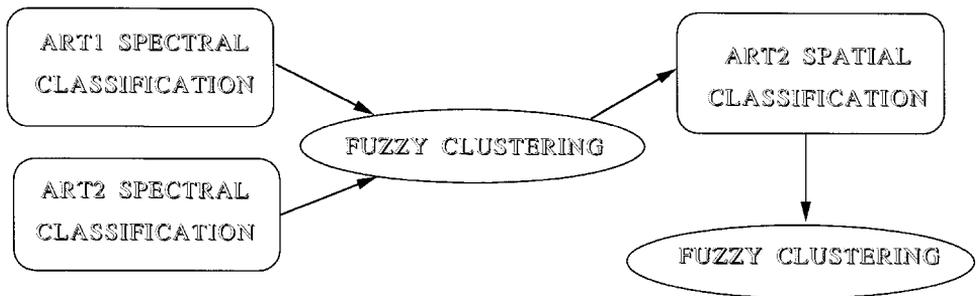


Figure 3. The classification chain proposed to classify urban SAR and optical images: the first step (spectral classification) is performed either by ART1 or ART2 networks, while the spatial classification is always obtained by means of an ART2-A algorithm. The fuzzy clustering steps are used to reduce the final number of categories.

the spectrum. Then, similar classes (i.e. classes with similar prototypes) are merged by the FCM clustering algorithm. The third step involves a spatial analysis again by neural networks, where a pixel is characterized also by its neighbourhood. This step, too, usually produces a larger number of classes than expected. Therefore, again to merge similar classes, a last FCM clustering is performed. The reason for twice interleaving the network approach and the fuzzy clustering is related to the general structure of the ART networks. They perform better when the number of output categories is left large (that is, when  $\rho$  is large, or  $\theta$  near to 0). This is due to the fact that the network updates its knowledge by a local analysis of the image to be analysed (exactly, pixel by pixel in the spectral analysis, in a limited neighbourhood for the spatial one). This could lead, when few classes are needed, to slightly different results for different raster scanning of the image matrix. The problem can be overcome by using the network with a parameter set less robust to noise and introducing FCM for a more global analysis of the data.

Finally, as the last consideration, it is useful to justify here the guidelines used in the choice of the four parameters  $\alpha$ ,  $\beta$ ,  $\rho$  and  $\theta$  in the results of the next section.

As already noted,  $\alpha$  is directly proportional to the ability of the net to recognize a new input as a new category. However, small  $\alpha$  values do not mean exactly rough classifications: we must take into account the dependence from vigilance  $\rho$ . Similarly, setting  $\beta$  (the learning rate) to a value near or equal to 1 corresponds to train the network with a very fast learning. This procedure decreases CPU times, but it may cause unwanted output oscillations. In any case, training results obtained with this choice are worse than those with slow-learning mode. Instead, setting  $\beta$  to 0.01 or 0.001 (there is no appreciable difference between these two values in our experience) grants a sufficiently stable convergence of the learning phase. The value of  $\theta$ , instead, as clearly shown from the first step of the ART2-A algorithm, corresponds to a threshold. We set it to an intermediate value in its range (i.e. around 0.5) so that some noise can be tolerated, but smaller pattern changes remain recognizable. The parameter  $\rho$  is the most critical parameter: larger values give rise to few output categories, but may result in oscillations and widening of the learning time of the net. According to our experience, to obtain stable outputs in reasonable time it is necessary to set  $\rho$  to a sufficiently large value while maintaining  $\beta$  small.

In summary, in our classifications  $\rho$  is set to 0.6, and  $\beta$  to the above-considered values (0.01–0.001), while  $\alpha$  and  $\theta$  may change to obtain the best results, but usually assume values in the middle of their respective ranges.

### 3. Experimental results

The classification chain introduced in the previous paragraphs was applied to four images of West Los Angeles (see figure 4) obtained by combining the R, G and B bands of the orthophoto with the C-band vertical polarization SAR image.

The SAR image was taken by a polarimetric–interferometric SAR operated by the Jet Propulsion Laboratory. This system operates in polarimetric and interferometric modes at C and L bands, with additional P-band polarimetric mode. The SAR image used in this study is acquired by the C-band interferometric mode with a 40 MHz bandwidth. The radar intensity is orthorectified as part of the interferometric processing and reported on a 5-m grid. The SAR data we worked on were obtained on 5 August 1994, from the height of 11 000 m. The flight path was from 33.97 latitude,  $-118.47$  longitude to 33.97 latitude,  $-118.41$  longitude, but we present here only a very small part of these data. The accompanying aerial photo is

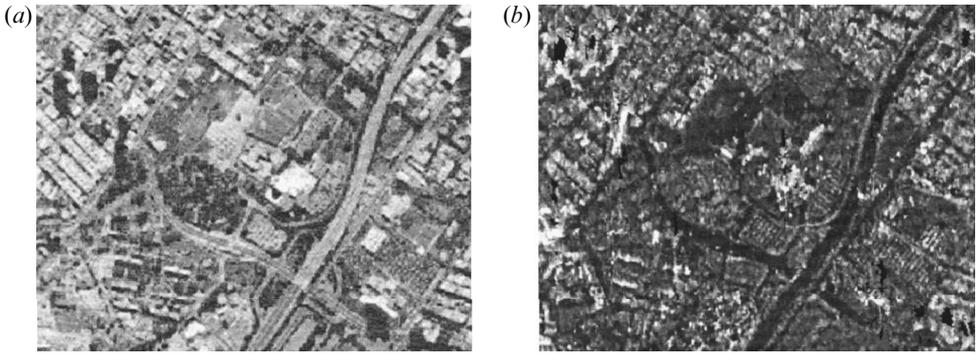
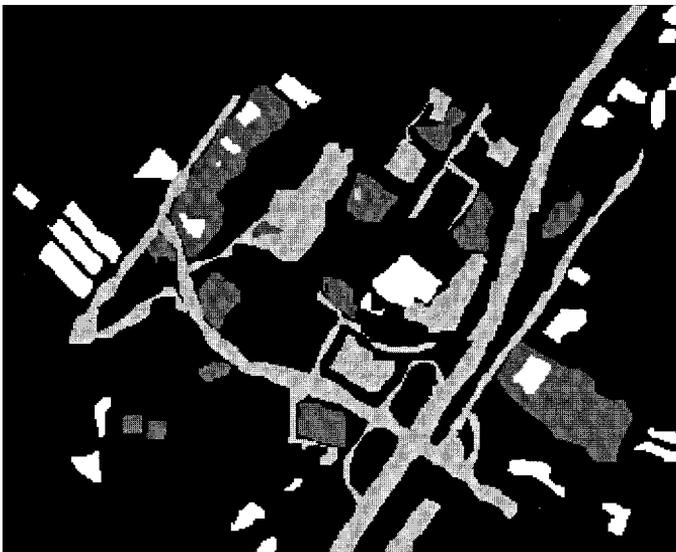


Figure 4. The images of Santa Monica, Los Angeles, used for the analysis; (a) optical image; (b) C-band SAR image.

co-registered to the SAR image using ground control points (GCPs) such as road crossings and other dominant features observable in both imagery. The resolution of the aerial photograph is 2 m, and the root mean square registration error with respect to the GCPs is less than a 2-m pixel.

The ground data was provided by visual inspection of the orthophoto and, for some parts of the original image, by direct classification of the image pixels 'on the ground'. Figure 5 presents the ground data set: white pixels represent buildings, light grey pixels bare soil, grey pixels vegetation, while the black background corresponds to the area where no ground data was collected. Note that the bare soil areas do not belong only to roads, but also to parking lots: this choice is due to the need to counteract the so-called *cardinal effect*, i.e. the well-known different visibility



■ Vegetation   □ Buildings   ■ Bare soil

Figure 5. The ground data set used for comparison of the classification algorithm over Santa Monica. Only three classes were considered: vegetation, bare soil and buildings.

of oriented features in a SAR image for different look directions. By choosing parking lot points, we introduce in the ground data features that should not give rise to look-dependent classifications; the same reason suggests to pick up built areas with as many different orientations as possible in the studied area.

To understand how the neural classification performs, we present the best classification results after each processing step on this small subset of the analysed image.

### 3.1. Spectral classification results

As explained in §2, the first step of our processing chain is a spectral aggregation of the original data performed by an ART1 or ART2 network. The four-band data are aggregated in an input vector in order to match the ability of ART networks to discriminate against different patterns. The output of the ART classification is fed to the FCM algorithm, assuming that only three classes must be present. The final classification map is shown for the ART1 in figure 6(a), and the ART2 in figure 6(b). The corresponding confusion matrices are reported in tables 1 and 2, respectively.

First of all, please note that the FCM algorithms may provide unclassified data, i.e. parts of the original data that it was not possible to classify and aggregate to any of the final classes. These pixels appear in the classification map in black, while the other grey levels have the same meaning as in figure 5. Similarly, in all the tables we have an *unclassified* data column that provides information on how many pixels of the ground data belonging to each class were not classified.

From a comparison of the two results, we observe that the best one corresponds to the ART2 classification, as we expected due to this network's capability to handle analogue inputs. The overall class performance is around 72%, which is not a completely satisfying result. In particular, we do not have good values of the omission accuracy for all the classes, and in particular vegetation is as low as 65%. Correspondingly, vegetation has the largest rate of unclassified sample (around 11%). A reason for this behaviour could be the large variance of the backscattered power for different vegetated areas: FCM is not able to aggregate the large number of vegetation classes provided by ART2, and is forced to leave unclassified many pixels. Note also that the building class is the most discriminated against the other ones, and its pixels are almost completely classified (the unclassified pixels are only 2.5% of the class total).

The results obtained by means of the ART1 network are definitely worse: the overall class performance is only 65%. This value corresponds mainly to an erroneous classification of the pixels belonging to the building class, since almost 40% are classified as bare soil. This is a consequence of the simple discretization scheme used: indeed, in this analysis we used a discretization of the input values into five equal ranges (see §2.3), and a vigilance parameter set to 0.35. A different choice of the discretization levels may improve the classification results, but the problems are that more levels mean more CPU time, and an adaptive discretization should be differently designed for each input dataset.

### 3.2. Spectral plus spatial classification results

The second phase of the data analysis procedure proposed in this paper again employs the ART2-A algorithm, but to perform a spatial analysis of the classified image: to this aim, the spatial similarity among neighbouring pixels is exploited. Therefore, a small ( $3 \times 3$ ) window around each input pixel is considered, and the



(a)



(b)

Figure 6. The classification results for (a) spectral analysis by ART1 or (b) spectral analysis by ART2, followed by the FCM clustering step.

relative weight of each class after the spectral classification in the window is computed. For instance, in our case of a three-class output from the first two steps, the percentages of pixels belonging to each of these three classes is computed. In turn,

Table 1. Spectral classification with ART1. Overall class performance (12 364/19 144)= 64.6% .

Class name	Omission accuracy (%)	Number of samples				Unclassified
			Vegetation	Bare soil	Buildings	
Vegetation	76.6	5654	4330	645	216	463
Bare soil	63.2	9804	2347	6201	898	358
Buildings	49.7	3686	291	1494	1833	68
Total		19 144	6968	8340	2947	889
Commission accuracy (%)			62.1	74.4	62.2	

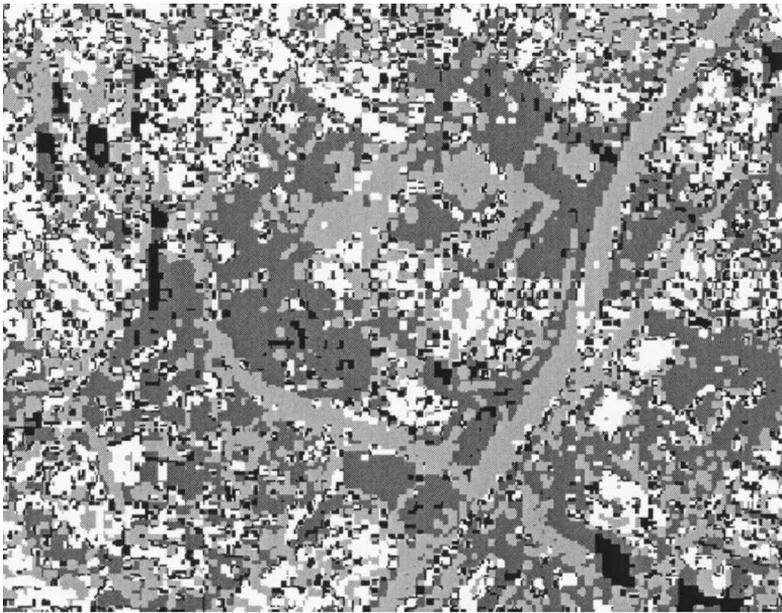
Table 2. Spectral classification with ART2. Overall class performance (13 768/19 144)= 71.9% .

Class name	Omission accuracy (%)	Number of samples				Unclassified
			Vegetation	Bare soil	Buildings	
Vegetation	65.1	5654	3680	1200	153	621
Bare soil	76.4	9804	1554	7489	380	381
Buildings	70.6	3686	221	778	2599	88
Total		19 144	5455	9467	3132	456
Commission accuracy (%)			67.5	79.1	83.0	

this three-valued vector is fed as input to the neural spatial classifier. Finally, after the neural clustering, the FCM is applied for consistent redundancy reduction. Note that this spatial analysis is similar to Gong and Howarth (1992), where the so-called *cover frequencies* for different land cover classes in a urban environment were computed. Our approach is simpler, and we do not attempt in this paper to recognize land covers by means of the spatial distribution of pre-classified pixels, but only to improve the initial spectral analysis.

As a first trial, we applied the fuzzy clustering to the ART1 spectral analysis presented in the preceding section to find if it improves its performance, and we obtained the results in figure 7(a) and the confusion matrix in table 3. The overall class performance has increased from 65% only to 67.5%, and the unclassified pixels have increased (from 889 to 1209). The small improvement of the class performance depends on the higher building omission accuracy (from 49.7% to 63.7%) but also on the lower vegetation results (from 76.6% to 71.2%). The reason is mainly the large vegetated area at bottom right in the ground data: the pixels erroneously classified as buildings in the first step have grown to larger groups, lowering the class performance. As for the other negative note, the higher number of unclassified pixels, we should note that it has a different distribution than in the initial ART1 spectral analysis: the largest part of them now belongs to the bare soil class. The reason for this behaviour may be the pixels on the boundaries between roads and built areas: the  $3 \times 3$  window used for the spatial analysis performs conceptually a low-pass filtering of the image, and it loses some of the smaller details.

Anyway, the confusion matrix has lower nondiagonal elements than in table 1, because there is slightly less confusion among all the classes than before: the membership of some pixels has been corrected by looking at their neighbourhood. This effect is clear,



(a)



(b)

Figure 7. The classification results for (a) spatial analysis by ART2 after spectral analysis by ART1 or (b) both spectral and spatial analysis by ART2 (even in this case the final FCM clustering step is considered).

for instance, in the larger uniformity of the major roads in the new map: in figure 6(a) they contained many grey and white pixels, i.e. pixels classified as building or vegetation, while in figure 7(a) they are almost completely classified as light grey (bare soil).

Table 3. Spectral plus spatial classification with ART1 and ART2. Overall class performance (12 905/19 144) = 67.4% .

Class name	Omission accuracy (%)	Number of samples	Vegetation			Bare soil		Buildings		Unclassified	
Vegetation	71.2	5654	4023	1052	326	253					
Bare soil	66.6	9804	1152	6533	1430	689					
Buildings	63.7	3686	177	848	2349	267					
Total		19 144	5352	8433	4105	1209					
Commission accuracy (%)			75.2	77.5	57.2						

Finally, the best classification results we can present in this work were obtained by considering the ART2 spectral and spatial analysis, whose results are presented in figure 7(b). The corresponding confusion matrix (table 4) shows an increase in the overall class performance with respect both to the simple ART2 spectral analysis and the ART1 spectral+ART2 spatial analysis. Indeed, this index has increased from 71.9% (ART2 spectral analysis) to 79.6%, and the number of unclassified pixels has reduced to a mere 1%. The higher overall performance depends on larger values of the omission accuracy in all the classes. This result is different from what we have seen performing an ART2 spatial analysis after an ART1 spectral one, where not all the classes performed better. The reason is that the initial spectral analysis by ART2 already provides a better boundary definition than the same analysis by ART1, and therefore the successive spatial classification is devoted basically to 'fill in the gaps' more than to refine these boundaries.

### 3.3. Comparisons with other supervised and unsupervised approaches

We discussed in §2.2 the relative importance of the neural network and fuzzy clustering approaches. However, it is interesting to complete the discussion with the analysis of the performance of the FCM algorithm applied alone directly to our urban dataset. We stress that this result should be compared with those of a purely spectral classification (for instance, see figure 6(a) and table 1), since no spatial interaction is considered in the standard FCM approach. The results of spectral FCM classification are presented in figure 8 and table 5.

We observe that the overall class performance is around 58% (less than both the ART1 and ART2 spectral analyses) and that many more pixels are unclassified (they are more than 20% of the total pixel number). The low performance is due to a

Table 4. Spectral plus spatial classification with ART2. Overall class performance (15 242/19 144) = 79.6% .

Class name	Omission accuracy (%)	Number of samples	Vegetation			Bare soil		Buildings		Unclassified	
Vegetation	78.4	5654	4432	1010	195	17					
Bare soil	80.3	9804	1619	7877	244	64					
Buildings	79.6	3686	215	530	2933	8					
Total		19 144	6266	9417	3372	89					
Commission accuracy (%)			70.7	83.6	87.0						



(a)



(b)

Figure 8. The classification results obtained applying directly the FCM (a) and ISODATA (b) algorithms to the optical and SAR data.

generalized worse behaviour of the classifier for any of the three target classes. In particular, we have a completely unsatisfying result for the vegetation class, whose omission accuracy is 48.6%. Indeed, since the FCM is forced to determine only three classes at its output, the large variance of the vegetation pixels, already mentioned

Table 5. Spectral classification with FCM only. Overall class performance (10 988/19 144)= 57.4% .

Class name	Omission accuracy (%)	Number of samples	Vegetation	Bare soil	Buildings	Unclassified
Vegetation	48.6	5654	2746 438	323	2147	
Bare soil	62.1	9804	1873	6086	604	1241
Buildings	58.5	3686	155	833	2156	542
Total		19 144	4774	7357	3083	3930
Commission accuracy (%)			57.5	82.7	69.9	

above, is the reason for so many unclassified pixels. Anyway, the unclassified pixel rate is very high for each class, so that the same considerations hold for the complete dataset. As a result, we could say that the FCM approach is definitely less useful than our neural analysis for this kind of data.

Another unsupervised approach, based on standard and not fuzzy logic, is the ISODATA algorithm: we applied this procedure in the unsupervised mode to our dataset, looking for three classes, and finding the results shown in table 6. The overall class performance is now 68%, and, even if all the pixels are now classified (like in any *hard logic* classifier), we could observe that there are many problems in the detection of the buildings and vegetation classes. In particular, nearly 30% of the vegetation pixels are classified as buildings, and also 18% of the bare soil data is considered as a part of this class (as it is clear from the low value of the corresponding commission accuracy). Even the CPU time required for the classification is much higher than the one needed for the neural analysis.

Finally, it is instructive to compare the ART-FCM approach with other neural classification methods, even if supervised. To this aim, we used a standard three-layer feed-forward network trained by means of the back-propagation algorithm. The choice of this simple yet powerful neural network structure was due to the fact that this kind of approach has been used successfully in general multispectral data analysis. Unfortunately, a direct application of this strategy to our urban dataset does not provide satisfying results. For instance, using as a training set a small subset of the ground data set in figure 5, it is very difficult to train in a reliable way the network. So, the results on the remaining pixels of the same ground data are even worse than those obtained using the FCM approach alone. An explanation of this behaviour may lie in the large variance of the values of the pixels belonging to each

Table 6. Spectral classification with ISODATA. Overall class performance (13 013/19 144)= 68.0% .

Class name	Omission accuracy (%)	Number of samples	Vegetation	Bare soil	Buildings	Unclassified
Vegetation	68.6	5654	3876	194	1584	0
Bare soil	64.3	9804	1455	6300	2049	0
Buildings	79.7	3686	216	633	2837	0
Total		19 144	5547	7127	6470	0
Commission accuracy (%)			69.9	88.4	43.8	

ground data class, that prevent the network from learning a suitable way to classify the input into the target classes. However, further studies in this direction are needed.

### 3.4. Timing considerations

Finally, we present timing considerations along all the processing chain reporting in table 7 the CPU time (in millions of floating point operations) for each processing step of our implementation. We stress that what we are interested in is not the absolute value (depending on the computer and the implementation details) but the relative computational weight of each step of the proposed procedure. From these values it is clear that our method can be correctly defined as ‘a neural approach corrected by fuzzy clustering’: the CPU time required for the ART classification is more than three orders of magnitude larger than the one dedicated to the FCM algorithm. In the same table it is reported, for comparison, the operations required by a purely FCM approach to the classification: the value is larger than the time required for all the four steps of the neural classification chain considered together. The reason is in the number of iterations required for the FCM to reach the convergence: after the neural classification only two or three iterations are sufficient, because we need to aggregate pre-classified data in less classes. Instead, starting from the original data, the iteration number is quite larger (more than 20), because it is more difficult to compute the prototypes of the target classes. Moreover, after the neural analysis the data to be clustered by FCM are no longer all the original data, but only the prototypes of the output classes of the previous step.

## 4. Conclusions

In this paper we investigated the problem of the classification of urban environments by means of optical and SAR images. The classification scheme proposed in this paper aims to exploit the advantages of ART networks augmented by fuzzy clustering. The use of the fast ART2-A algorithm in both spectral and spatial data analysis leads to the best results.

A strategy to choose the parameters ruling the ART2 (only four in the ART2-A algorithm) to achieve a desired network behaviour has been presented. Moreover, the interaction between the neural and fuzzy clustering algorithm has been discussed.

The results obtained over an area in Santa Monica, Los Angeles can be considered satisfying for this preliminary work in the area of urban analysis using neural classifiers, even if the overall class performance is still too low for practical applications, even in our best results. A comparison with the direct application of the ISODATA and FCM algorithms to the same dataset encourages us to continue in the same direction with further analyses, since ART results are better. Currently, other methods using more complex ART networks already enclosing fuzzy concepts (fuzzy ARTMAP, for instance), or classification chains based on different features,

Table 7. Typical classification CPU times (expressed in Mflops per second).

Spectral ART2	97
Spectral FCM	0.01
Spatial ART2	114
Spatial FCM	0.02
Total	211
Only FCM	870

like texture or principal component analysis, are under investigation. Moreover, for a better characterization of urban environments, the integration of other sources of data (hyperspectral measurements and Digital Elevation Models extracted from interferometric measurements, for instance) will also be considered.

### Acknowledgements

The authors would like to thank M. Trenchi for performing the simulations and for the useful discussions about neural classification approaches. The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

### References

- BARALDI, A., and PARMIGGIANI, F., 1995, A neural network for unsupervised categorization of multivalued input patterns: an application to satellite image clustering. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 305–316.
- BENEDIKTSSON, J. A., SWAIN, P. H., and ERSOY, O. K., 1990, Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 540–551.
- BISCHOF, H., SCHNEIDER, W., and PINZ, A. J., 1992, Multispectral classification of Landsat images using neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 482–490.
- CARPENTER, G. A., GROSSBERG, S., and ROSEN, D. B., 1991, ART 2-A: an adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks*, **4**, 493–504.
- CARPENTER, G. A., and GROSSBERG, S., 1987a, A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision Graphics Image Processing*, **37**, 54–115.
- CARPENTER, G. A., and GROSSBERG, S., 1987b, ART2: self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, **26**, 4919–4930.
- CHEN, K. S., HUANG, W. P., TSAY, D. H., and AMAR, F., 1996, Classification of multifrequency polarimetric SAR imagery using a dynamic learning neural network. *IEEE Transactions on Geoscience and Remote Sensing*, **34**, 814–820.
- DAVÉ, R. N., and KRISHNAPURAM, R., 1997, Robust clustering methods: a unified view. *IEEE Transactions on Fuzzy Systems*, **5**, 270–293.
- DOBSON, C., LELAND, F. P., and ULABY, T., 1996, Knowledge-based land-cover classification using ERS-1/JERS-1 SAR composites. *IEEE Transactions on Geoscience and Remote Sensing*, **34**, 83–99.
- DOUSSET, B., 1995, Synthetic Aperture Radar imaging of urban surfaces: a case study. *Proceedings of IGARSS'95, Florence, Italy, August 1995*, pp. 2092–2096.
- DREYER, P., 1993, Classification of land cover using optimized neural nets on SPOT data. *Photogrammetric Engineering and Remote Sensing*, **5**, 617–621.
- GONG, P., and HOWARTH P. J., 1992, Land-use classification of SPOT HRV data using a cover-frequency method. *International Journal of Remote Sensing*, **13**, 1459–1471.
- GRUEN, A., 1998, TOBAGO—a semi-automated approach for the generation of 3-D building models. *Photogrammetry and Remote Sensing*, **53**, Special Issue on extraction of man-made objects from aerial images, pp. 108–118.
- GRUEN, A., KMUEBER, O., and AGOURIS, P. (eds), 1995, Automated extraction of man-made objects from aerial and space images. *Proceedings of the Monte Verita Workshop*, (Birkhauser Verlag).
- HARA, Y., ATKINS, R. G., YUEH, S. H., SHIN, R. T., and KONG, J. A., 1994, Application of neural networks to radar image classification. *IEEE Transactions on Geoscience and Remote Sensing*, **32**, 100–109.
- HENDERSON, F. M., and XIA, Z. G., 1997, SAR applications in human settlement detection, population estimation and urban land use pattern analysis: a status report. *IEEE Transactions on Geoscience and Remote Sensing*, **35**, 79–85.

- HEPNER, G. F., HOUSHMAND, B., KULIKOV, I., and BRYANT, N., 1998, Investigation of the potential for the integration of AVIRIS and IFSAR for Urban Analysis. *Photogrammetric Engineering and Remote Sensing*, **64**, 512–520.
- HERMANN, P. D., and KHAZENIE, N., 1992, Classification of multispectral remote sensing data using a backpropagation neural network. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 81–88.
- HRIKKONEN, J., KANELLOPOULOS, I., VARFIS, A., STEEL, A., and FULLERTON, K., 1997, Urban land-use mapping with multi-spectral and SAR satellite data using neural networks. *Proceedings of IGARSS'97, Singapore, August 1997*, pp. 1660–1662.
- LEVINE, D. S., and PENZ, P. A., 1990, ART 1.5—a simplified adaptive resonance network for classifying low-dimensional analog data. *Proceedings of the International Joint Conference on Neural Networks, Washington, DC, January 1990*, pp. 639–642.
- LIU, Z. K., and XIAO, J. Y., 1991, Classification of remotely-sensed image data using artificial neural networks. *International Journal of Remote Sensing*, **12**, 2433–2438.
- PAOLA, J. D., and SCOWENGERDT, R. A., 1995, A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 981–996.
- RITTER, N. D., and HEPNER, G. F., 1990, Application of artificial neural network to land-cover classification of thematic mapper imagery. *Computers and Geoscience*, **16**, 873–880.
- SILVA, S., and CAETANO, M., 1997, Using artificial recurrent neural nets to identify spectral and spatial patterns for satellite imagery classification of urban areas. In *Neurocomputation in Remote Sensing Data Analysis*, edited by I. Kanellopoulos, G. G. Wilkinson, F. Roli and J. Austin (Berlin: Springer), pp. 151–159.
- TZENG, Y. C., and CHEN, K. S., 1998, A fuzzy neural network to SAR image classification. *IEEE Transactions on Geoscience and Remote Sensing*, **36**, 301–307.
- XIA, Z. G., 1996, Applications of multi-frequency, multi-polarization and multi-incident angle SAR systems in urban land use and land cover mapping. *Proceedings of IGARSS'96, Lincoln, Nebraska, July 1996*, pp. 2310–2314.
- WEYDAHL, D. J., BECQUEY, X., and TOLLEFSEN, T., 1995, Combining ERS-1 SAR with optical satellite data over urban areas. *Proceedings of IGARSS'95, Florence, Italy, Sept. 1995*, pp. 2161–2163.