A KNOWLEDGE-BASED SYSTEM FOR UNSUPERVISED CLASSIFICATION OF HIGH RESOLUTION MULTISPECTRAL SATELLITE IMAGES.

A. Baraldi

IMGA-CNR, Modena (Italy)

Abstract

A knowledge-based, hierarchical, unsupervised classification scheme for high resolution multispectral satellite images is proposed. This scheme, which finds its conceptual bases in the work of Nagao and Matsuyama (1980) for structural analysis of aerial photographs, introduces a new filtering algorithm which is able to preserve fine linear structures of the image.

The classification products are: 1) a raster image where detected output classes are characterized by some specific information content and/or by spectral separability (by cluster analysis); 2) property tables describing each elementary region in terms of descriptive and geometric attributes; 3) geometry of output elementary regions converted to vector format.

An example of the application of this classification scheme to a Landsat TM multispectral image is presented.

Key Words: satellite image classification, knowledge-based classification, image understanding.

1 Introduction

Starting from the work of Nagao and Matsuyama (1980) (hereafter referred as N-M), our aim was to develop a general system for unsupervised classification of high resolution multispectral satellite (HRMS) images. This approach results in a knowledge-based, hierarchical, classification scheme. The scheme is characterized by simple user interaction and high processing speed and should be used whenever: 1) thematic mapping must be performed without reference data available; 2) an extremely fine classification precision is not required. For these reasons this classification scheme could also be used to identify more specific training samples to be used in further supervised classification.

2 Evaluating the feasibility of the project

Some basic strategies can be chosen in order to obtain, from a classification scheme, output product of cartographic quality.

1. Any existing thematic map where land use/land cover is presented shows minimum map unit size around 2.5 mm by 2.5 mm (Lillesand and Kiefer, 1979). This is due to human expectation about thematic uniformity and also to readibility problems for the map reader (e.g. precision affected by graphicism error)(Swain and Davis,1978). Because of these two elementary considerations, we want to reduce the small region thematic scattering of the original picture on the output product. This goal can be pursued providing the pre-processing stage with an edge preserving and blurred edge sharpening spatial filter.

2. It is well known from the literature (Swain and Davis, 1978) that pixel-by-pixel classifiers often produce an even more detailed classified image than is actually needed by the resource management. Besides, these classifiers deal with tonal information only, ignoring spatial information. Thus, pixel-by-pixel classifiers are acceptable only when spectral rules can guarantee separability of target cover types. Homogeneous regions classifiers (sample classifiers) (Ketting and Landgrebe, 1976) are more suitable for resource management. Their first step is an image partition algorithm that splits the scene into objects that are spectrally homogeneous (segmentation process) and then classifies these elementary areas as whole units. Thus, the classification of each pixel in a sample is the result of the spectral properties of the sample as a whole, i.e. the contextual information is used. Furthermore, the explication of spatial attributes (e.g. position, shape, etc.) of homogeneous objects allows exploitation of spatial information (by means of spatial rules). It is also possible to consider a straightforward conversion of such classified data on a vector-based GIS.

3. Finally, absence of training fields is among the requirements of our classification scheme. In supervised classification, the user is required to explicate his *a priori* knowledge about the scene under investigation regarding which classes are present and where training areas are localized. The training areas must individually refer to a single cover type and be scattered through the image. As a consequence, it is often impracticable to locate "pure" (and sufficient in number) supervised training areas to characterize ground covers over a large scene due to lack of ground truth information (Swain and Davis, 1978). On the other hand, unsupervised analysis, based entirely on unsupervised classification, reveals statistical patterns that are inherent in the data. This method is expected to extract relatively few, large spectral classes that do not necessarily coincide with information classes. Therefore, unsupervised analysis can be applied successfully only when the information classes are easily discriminated in the feature space. To provide optimal analysis of large and heterogeneous images, hybrid procedures (unsupervised and supervised) have been adopted (Swain and Davis, 1978; Lillesand and Kiefer, 1979). In a hybrid classification scheme, supervised training areas, each one containing a single cover type, are used altogether with unsupervised training areas. Each unsupervised training area must be selected to contain multiple cover types and is clustered independently. This means that signatures (i.e. sets of statistical data) are derived from training samples stemming from a supervised training method and/or from clusters obtained in an unsupervised training procedure. Commercial tool boxes, like ERDAS (1990), provide a set of software tools to create signature files of statistical data from training samples as well as from clusters. These signatures can be tested, deleted and merged with any other signature file. Thus, in order to improve unsupervised analysis we have to move toward the use of unsupervised training areas. Each one of these is a portion of the image where cluster analysis is independently run to extract a subset of the overall image spectral classes. This idea can be applied to hierarchical classification: our scheme should first extract some knowledge subset of the original image and then perform cluster analysis in each subset separately. In this way, an increment in spectral discrimination of the cluster analysis is expected; besides, depending on the spatial/spectral pattern used to create his knowledge subset, the user benefits from matching information classes with output clusters. Our classification scheme was at first described as being unsupervised, meaning that it needs no supervised training procedure. However, it was also stated that this approach is based on domain-knowledge. In other words, the analyst exploits his *a priori* knowledge about the image regarding spectral and spatial patterns in order to obtain specific subsets of the original image. On the other hand, the operator is not required to know where these patterns are located on the image itself as in supervised analysis. This means that, in terms of domain-knowledge, this classification scheme cannot be considered as purely unsupervised nor supervised but rather as a hybrid classification procedure.

2.1 The N-M classification procedure

As reminded in the Introduction, a classical example of multispectral classification procedure, based on segmented scenes, was developed by Nagao and Matsuyama (1980) for aerial photographs. This classification scheme is: 1) knowledge-based (some procedures are implemented by the analyst based on his *a priori* knowledge of the image domain); 2) highly structured (black boxes organization); 3) hierarchical (a study area is described at several levels of details).

What follows is a conceptual description of the N-M procedure; the description stresses the philosophy underlying the operations performed by the different blocks.

The classification system uses as input aerial multispectral images and is organized in three major modules:

1. *N-M image pre-processing.* It consists of three steps: filtering, image segmentation and image description.

2. *Feature extraction.* The idea is that any possible image data information level must be exploited in order to extract a separate kind of characteristic regions (cue regions). Any cue region

category represents a subset of reliable knowledge (kernel information) (Ton et al., 1991). For reliable knowledge we mean any information that can be extracted from the image with a general purpose, image independent, standard method. This stage is based on image domain knowledge that can be considered as: a) spectral dependent (presence of spectral domain rules); b) spatial knowledge dependent in terms of size/shape features; c) contextual/topological rule-free; d) general-purpose (i.e. task-independent: any cue region may be used as input to decision rules of the classification box). This means that each cue type either refers to: i) a particular spatial pattern; ii) a particular spectral pattern. In the N-M scheme the different image information levels, each of them characterizing a separate cue category, are: a) color (spectral signatures); b) brightness; c) shape; d) size; e) texture. The selected cue region categories are respectively: a) vegetation; b) shadow and shadow making; c) elongated; d) large; e) high contrast.

3. *Pattern recognition stage*: each separate land cover category is detected by a specific module, independent from the others, whose input are cue regions. This stage is: a) task-dependent: each module is specialized in recognizing a particular output category; b) spatial-rule dependent: exploiting contextual/topological rules.

3 The N-M approach to HRMS imagery

In order to evaluate the applicability of the N-M classification approach to HRMS imagery, the information content of HRMS images before and after N-M pre-processing must be compared. Visual interpretation technique and N-M segmentation scheme do operate in the same manner. In fact, both processes divide the scene into elementary regions characterized by: 1) homogeneous (low variance) spectral response (color); 2) size and shape features enhanced by contrast in spectral response between adjacent objects.

As a consequence, it is possible to evaluate the potential applicability of N-M classification scheme to HRMS images by visual interpretation of N-M preprocessed images.

Our assessment starts with some general considerations of pattern recognition.

Each homogeneous object stemming from segmentation is characterized by two items of information: 1) spatial/contextual (pictorial); 2) spectral (numerical). Within-region tonal information is strongly spatial correlated. Thus, this spectral information can be summarized in a set of statistics that refers to the region as a whole. If different cover types are spectrally separable, no spatial investigation is necessary. Whenever spectral separability between cover types is not guaranteed, some spatial patterns may be used as differentiating factors. In this case, however, discrimination between these cover types is possible only in those parts of the image where they do not belong to the same surrounding. In other words, spatial (size/shape) investigation is neither a necessary nor sufficient condition for identification of objects. On the contrary, spectral identity is necessary and sufficient condition for identification. In fact, any spatial (size, shape, contextual) investigation criteria have the following characteristics:

1. *not always possible*: context-dependent, i.e. relying on contrast in context; this means that spatial investigation is possible only where contrast does exist (maintaining the geometric identity of the object under investigation);
2. *not precise, whenever possible*: if contrast does exist between the object and its surrounding its size/shape identity is decayed by edge effect problems that reduce the focusing of the boundaries (as happens with HRMS imagery).

This blurring edge problem affects: 1) 'Large' homogeneous areas: their spatial identity (size/shape) is maintained but is affected by unfocused-edge problems; their within-region variability is increased because of the "melting" contributions of small objects of different nature; 2) 'Small' regions: they are merged into their dominant background, losing their spatial identity ('large' and 'small' terms are intended in comparison with dimensions of ground resolution cell).

The general effect is a widespread loss in spatial details even when a natural pattern is conserved through sensor investigation. Thus, tonal information is still predominant in HRMS imagery with respect to spatial one.

As a consequence of the relatively low spatial resolution of HRMS images, also texture information level is lacking on such images (e.g. forests).

Sometimes texture patterns are conceptually confused with land cover heterogeneity, rather due to noise-like phenomena stemming from edge effect problems (see above). In fact, texture information is noiseless only when spatial edges between different land cover elements are respected, that is when within-region heterogeneity (texture) is actually produced by changes in reflectance due to a single cover type.

These considerations must be kept in mind any time some spatial rules are exploited as discriminating factors. A brief description of the information content of raw HRMS and N-M preprocessed images, as it results from visual interpretation, is given.

3.1 Spectral and spatial patterns

HRMS images pre-processed by the N-M method can be visually assessed in order to evaluate their informational utility in digital processing. The N-M filter is first applied to the image of Fig.1. In order to emphasize its effects, the N-M filter was iterated until convergence (i.e. until no pixel changed its value)and the result is shown in Fig.2.

As already said, due to functional analogies between analog and digital segmentation, it is possible to switch between these two segmentation processes. In particular, visual assessment may regard: 1) Typical spectral patterns as they are ouput on the N-M filtered image (see Fig.2); 2) Typical spatial patterns as they are emphasized by the automatic segmentation process applied on the N-M filtered image.

An overall assessment of Fig.2 shows that the filtered image has maintained main spectral information from the original image (only eliminating scene heterogeneity) as well as high spatial frequencies (edges). However, HRMS filtered image leads to segmentation results radically different from those derived from raw data.

For example, what looked originally like a vegetation crop field, that is a large thematically homogeneous area, appears on the filtered image as a set of smaller fields each one characterized by its own quite uniform radiance value. These elementary fields are detected by the automatic segmentation algorithm. Thus, the single elements of the overall field have lost on the filtered image what was the crop field specific spatial pattern (straight borders) on the original image.

A further example regards urban areas, that is fine textured areas on the original image. Their texture pattern becomes much coarser on filtered image. The final effect on segmented filtered image cannot be forecast. Sometimes urban areas generate high contrast areas, that is areas where many small elementary regions are gathered. At other times smoothed urban areas generate a unique oversized region (Baraldi and Parmiggiani, 1989).

These different results in segmentation are due to some inherent properties of the filtered image with respect to the original one, that is to some specific behaviors of the N-M spatial filter.

These behaviors are: 1) Blurred edge sharpening; 2) Small details removal.

The first feature causes spectral differentiability between subregions originally belonging to a thematically homogeneous area. This effect increases with the number of iterations of the filter. It tends to separate large areas of the original image into smaller regions each one characterized by a quite uniform color information.

The second feature is an implicit rule in the behavior of any spatial domain filter. This rule says that the local area affected by the spatial filter investigation has to be smaller than the smallest detail to be conserved on the original image.

Our filter implementations involve local masks whose area goes from 5 to 7 pixels. This means that elementary regions smaller than 5 pixels are absent on filtered image in Fig.3. Thus, any textured area (i.e. any clustering of small elementary regions) on the filtered image is composed of textons (Julesz, 1981) as large as 5 pixels or larger, i.e., only coarse texture areas can be conserved.

Generally speaking, spatial information (in terms of texture and in terms of size/shape properties) in the filtered image with respect to the original one, is either spoiled (in terms of discriminating ability) or lost. More precisely it can be stated: 1) There is no correlation between thematic content and spatial patterns in filtered image. In fact, in the partitioned image several regions belonging to the same class of geographic entities may appear very different in terms of shape and size. Generally speaking, this is true for any thematic class. From this observation it is derived that any time a size/shape knowledge subset on any specialized classification module is introduced, it is likely that some thematically homogeneous elementary regions are directed to different output classes. This means that the number of output classes produced by the classifier will be bigger than the number of different cover categories actually present on the image. Generally speaking, output classes present an "n to 1" relationship with land cover types.

2) Final discrimination leading to selection of an output class (starting from cue region categories) must always rely on spectral information: for example, if no spectral signature has been used in selecting a specific cue category the classification scheme must perform a cluster-oriented segmentation in order to produce the related output class.

3.2 Spatial relationships

Generally speaking geometric, contextual and topological relationships that forecast the organization of a scene can be modeled for those areas where there is strong human knowledge of the objects' spatial organization. This means that many spatial rules involve geographic entities related to human activities in the outside world.

Unfortunately, objects of the real world in which the human contribution to regularity, in terms of shape (e.g. houses, streets), texture pattern (e.g. tree plantations), etc., comes to be relevant are those whose size are comparable or smaller than ground resolution pixel in HRMS images.

Contextual and topological models for some of these real world objects can be conceived for aerial (Nagao and Matsuyama,1980) or HRMS imagery (Ton et al., 1991). However, since these objects are hardly restored by the HRMS sensors (typical map scaling from HRMS images is: 1:100,000), general (i.e. image independent) spatial rules are hardly detectable on the image.

As for original image, general spatial rules (contextual and topological) cannot be detected on N-M pre-processed images (this also depends on the filter's fidelity in restoring linear elements).

4 System implementation

In order to write the software for the designed operations, some practical objectives must be kept in mind: 1) perform efficient data processing in terms of computation time with respect to output data quality; 2) reduce the number of parameters needed by the algorithm, supporting the operator with experimentally set default values and with interactively data-driven information; 3) exploit operator's ability in evaluating pictorial data quality (spatial pattern) creating output images whenever necessary.

4.1 Input data

Landsat 5 TM images from the satellite pass of 27th July 1990 (Path / Row = 192/29) were used as input data. Image windows 512× 512 in band 2,3 and 4, centered onto the town of Modena, were selected (see Fig.1 for band 4).

The choice was based on the consideration that VisRed and NearIR channels are extremely important for vegetation regions extraction. Any additional band is not necessary for operating the system. Regarding this, any feature selection criterion may be adopted asking for minor adaptation in the presented modules (e.g. brightness equation).

4.2 Hardware and software platforms

A SUN SPARCstation 4/330, with 16 MB RAM memory and 1.8 GB disk capacity, was used for this study. The software package SUNvision, v.1.1, was utilized for basic image processing operations (display, enhance, zoom, color, etc.)

4.3 Smoothing stage

It was verified that the original version of the N-M spatial filter (Nagao and Matsuyama, 1979), involving 9 and 7 pixels per investigating mask (taken from the SPIDER Subroutine Library, 1983), destroys thin regions of the image (e.g. roads, canals; see Fig.2).

Two modified and optimized versions of the N-M filter have been tested for linear element conservation: 5-pixel-long linear masks have been added to some 5 and 6 pixel polygonal masks (obtained by specular division of masks implemented in the original N-M filter version).

The optimization aspect is pursued through: 1) reduction of partial sums during media value calculation using already calculated pixels' sum values; 2) investigation of polygonal masks first (monodimensional elementary regions are less likely to occur on the image with respect to bidimensional elements); 3) zero variance condition detection after each mask investigation.

The two new filter versions present respectively a centric and an eccentric disposition of linear masks with respect to pixel being filtered. The total number of masks is 80 (60 linear and 20 polygonal) in the first case, 140 (120 plus 20) in the second case.

The results of filtering operations are presented by reporting the percentage of modified pixels for each filter iteration.

Experimental observations suggest that one iteration of the filter is able to reduce heterogeneity of the scene while saving small dimensions pictorial details and reducing computation time: the result of the eccentric filter applied with only one iteration to the image of Fig.1 is shown in Fig.3.

4.4 Segmentation

4.4.1 Region-growing

This operation follows the N-M pre-processing and was already implemented for HRMS images in (see Baraldi and Parmiggiani, 1989). The size of the subwindow under differential investigation is initialized to 64×64. The H parameter used by the valley detection algorithm (VDA) in differential histograms is set to 1.

An output segmented image is first produced with these two parameters set to their default value. After image evaluation the operator may decide to repeat the processing step by changing the H parameter or interactively setting the threshold parameters.

When the segmented output image is satisfactory the program computes a basic property table of the regions. The computed attributes are: 1)perimeter extraction (8-connected, stored in a raster image); 2) area value; 3) mean radiances; 4) MER (minimum enclosing rectangle) coordinates.

From this point on, any search on the image will be conducted in terms of region elements, reading the property table first and then moving to the image area of interest.

4.4.2 Region merging

Because of imperfect overlapping of homogeneous regions through multispectral image many small regions appear along large region borders. These small regions have intermediate radiance values between their large neighbors average reflectance due to sensor averaging on the edges. In this regard these small regions are thematically uncertain.

Merging of small regions with adjacent large regions presenting most similar multispectral mean values can reduce the total number of regions by some 10% up to 50%. Besides, pixel dimension is about 1 mm; thus, a 1-3 pixel area will result smaller than the graphicism error of human eye (2.5 mm).

The default area threshold ($MinArea$) below whose merging is performed is set to 1.

4.5 Classification

4.5.1 Feature extraction

At this stage, kernel information from general-purpose image analysis is extracted: each kernel information produces an ouput labelled picture.

The following general purpose cue regions are generated: a) low brightness regions; b) water regions; c) vegetation regions; d) large homogeneous regions; e) thin and elongated regions; f) very-thin and elongated regions.

In order to define the meaning of every cue region, the specialized actions of each module are now described.

a) b) Extraction of low brightness and water regions

Low brightness regions will be further investigated by means of an unsupervised approach, that is, their spectral separability is going to be evaluated. Water regions may be further divided by some geometric features in rivers, ponds, lakes or sea.

Brightness value for region j is first computed, by the expression:

$$B(j) = 0.2*(Ave.GreenVIS(j)+2*Ave.RedVIS(j)+2*Ave.NearIR(j))$$

This expression stresses the difference in radiometric values for those bands in which atmospheric (Rayleigh) scattering is less relevant (i.e. for longer wavelengths).

An adaptive threshold in brightness is evaluated through these steps: 1) Histogram computation of brightness values for the region population; 2) Average brightness value (AB) calculation for the region population; 3) Transformation of the histogram into a probability diagram in the [zero, AB] domain; the obtained plot is bimodal: the two modes should be quite clearly separated and the peak at the left side is due to low brightness regions; 4) Brightness threshold (BT) computation that makes the between-class variance maximum (Otsu, 1979).

Any region j for which:

$$B(j) \leq BT$$

is tested for water presence (in this manner the selection of fresh snow regions that are bright in visible wavelength but characterized by a ratio $(Ave.NearIR(j)) / (Ave.RedVIS(j)) < 1$ is avoided). In particular if:

$$V(j) < 1 \quad .and. \quad V2(j) < 1$$

where:

$$V(j) = (Ave.NearIR(j))/(Ave.RedVIS(j))$$
$$V2(j) = (Ave.NearIR(j))/(Ave.GreenVIS(j))$$

then region j is considered as water region, otherwise it is associated with the low brightness cue category. A similar water pixel selection criterion is used in Ton et al.(1991).

A correspondent rule could be applied to high brightness regions to detect fresh snow areas. A partial image is output, presenting both low brightness and water regions.

c) Extraction of vegetation regions

These regions can lead to further discrimination between vegetation agricultural land (vegetation crop fields) and grassland or forest.

Many vegetation indices have been compared in literature (Perry and Lautenschlager, 1984). The measure of the likelihood of vegetation that is adopted is the ratio:

$$V(j) = (Ave.NearIR(j))/(Ave.RedVIS(j))$$

This ratio is independent from multiplicative noise phenomena; in particular, it can reduce the effect due to changes in illumination conditions. Thus, it appears to be quite independent from the image under analysis but related to the kind of sensor.

For SPOT HRV images, the Vegetation Index Threshold (VIT) was empirically set to 2.2 (Baraldi and Parmiggiani, 1989), while for Landsat 5 images (band 4 and 3) default VIT is set to 2.5. A similar vegetation pixel selection criterion is used in Ton et al., (1991). In this scheme, any region j for which:

$$V(j) >= VIT$$

is considered as vegetation region.

After the first extraction of vegetation regions by means of the default VIT, the program produces statistics of region-overlapping between low brightness and vegetation regions. The partial result image is also created. The operator can evaluate the results and eventually decide to repeat the step by setting a user-defined VIT value.

d) Extraction of large homogeneous regions

These regions can be a cue for recognizing urban-areas, large bare soil areas, lakes and sea.

A statistical (adaptive) criterion for extracting image-driven information is adopted. Its steps are: 1) Creation of a histogram of the area values for the region population; 2) Applicaton of the VDA. In this case the H parameter default value has been set equal to 20: in fact, a single mode histogram is expected, having a long "cue" on the right side (low population for large area values). Thus, also this parameter is quite image-independent (i.e., changes in H parameter do not affect valley detection): in this way, an adaptive area threshold value (AT) is obtained. Any region j is considered a large one if:

$$\text{Area}(j) \geq AT$$

As well as any other stage, this step may be repeated by setting user-defined H parameter, until the desired pictorial result is reached.

e) f) Extraction of: i) thin and elongated regions; and ii) very-thin and elongated regions

These regions can represent a cue for roads, railroads and rivers detection. This module is, by far, the most influenced by the operator choice in adopting one of the possible N-M filter implementations for the input image. In fact, the change in linear element fidelity for different N-M implementations is evident in input images (see Fig.2 and Fig.3) leading to a loss in thin-elongated details.

A thin region can be defined as a region with a high percentage of boundary points with respect to total area. An elongated region can be intuitively defined as one whose "long side" is much bigger than its "short side".

Incidentally, thin regions may also be elongated as well as not elongated (when they are small and rounded). On the other hand, elongated regions may or may not be thin.

The geometric characteristic of roads and railways in HRMS images can be described as being very thin and elongated (see below), while rivers are long but not necessarily very thin.

The elongatedness factor evaluation algorithm is based on the following steps: 1) Fill in the holes of connected regions in order to obtain simply connected regions; 2) Application of a thinning algorithm (Hildtch, 1969); 3) Evaluation of the longest path on thinned regions; 4) Evaluation of average region width across the longest path direction; 5) Evaluation of the elongatedness factor as the ratio between the length of the longest path and the average width value.

For more details refer to Nagao and Matsuyama (1980). In order to reduce the set of regions to be investigated by the elongatedness evaluation subroutine some simple geometric controls based on general-purpose (image independent) parameters must be considered.

Up to 4 parameters are used through this module. However, only two of them (ET1, ET2) are sometime changed by the user from their default value. Two parameters are oriented to the selection of elongated regions (e.g. rivers), while the other two are oriented to the selection of thin regions.

Parameters relating to thin and elongated region selection are called:

ET1 (Elong.Thres.1) and IT1 (Inner Point Thres.1)

Parameters relating to very-thin and elongated region selection are called:

ET2 (Elong.Thres.2) and IT2 (Inner Point Thres.2)

ET parameters type refers to the expression:

$$(Length\ of\ the\ longest\ path\ in\ region\ j)/$$
$$(Ave.width\ along\ the\ longest\ path\ for\ region\ j)$$

In other words, for any region, ET refers to the ratio:

$$(long\ side)/(short\ side)$$

while IT parameters relate to the expression (for any region j):

$$100 \star [(area\ j) - (perimeter\ length\ j)]/(area\ j)$$

where perimeter is 8-connected.

This relationship represents the percentage of internal (i.e. not boundary) points for region j with respect to size of region j. Default parameter values are:

$$ET1 = 6, IT1 = 25$$

$$ET2 = 5, IT2 = 5$$

These values are determined by evaluating surface feature dimensions and HRMS spatial resolution (20÷30 meters).

Typical thin-elongated surface elements of interest are roads and railroads; elongated but not thin elements are rivers. One would expect that thinner objects result partitioned into shorter regions, due to filter unproper smoothing effect over small details.

Roads are less than 10 to up to 45 meters wide (Ton et al., 1989). This means that their major radiant contribution may involve 1 to 3 pixels width for SPOT HRV, 1 to 2 pixels width for Landsat 5 TM. Refering to a Landsat image, a thin-elongated region such as a road will present internal points percentage close to 0% (IT2=5) and ET2 $\geq$ 5. Thus, its area cannot be any smaller than ET2.

In general, some disequalities must be verified among the program's parameters:

$$MinArea < ET2; \quad ET2 \leq ET1; \quad IT2 \leq IT1$$

From these considerations the general implementation of the module is as follows: 1) If Area(j) < ET2, exit; 2) If Inner Point (j) > IT1, (region j is not thin) exit; 3) Elong (j) evaluation occurs;

4) If Inner Point (j) $\leq$ IT2 then (region j is very thin):
   if Elong (j) $\geq$ ET2 then
      region j is very-thin and elongated
   end if
else (region j is not very-thin but it is thin)
   if ELONG (j)$\geq$ ET1 then
      region j is thin and elongated
   end if
end if

Some implicit logical equations can be found among cue region categories. In particular: a .and. b = 0; b .and. c = 0; e .and. f = 0.

4.5.2 Pattern recognition

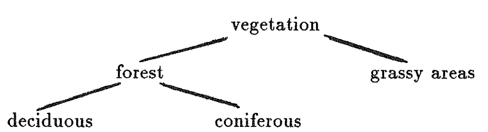
A consistent hierarchical knowledge organization was developed whose roots are the N-M pre-processed image, intermediate levels are the cue region categories (knowledge subsets) and whose leaves are all possible output classes.

As already said, ouput classes cannot be 1 to 1 related to geographic features categories whenever size/shape selection criteria are adopted. Thus, for all output classes some spectral discrimination has to be scheduled. This means that whenever a spectral rule is absent or not sufficient in characterizing the informational content of the output class, a cluster analysis is highly recommended. In this way the elementary regions within each output class are subdivided depending on their inherent spectral/textural properties.

Any output class should be conceived in order to refer (ideally) to a single land cover type.

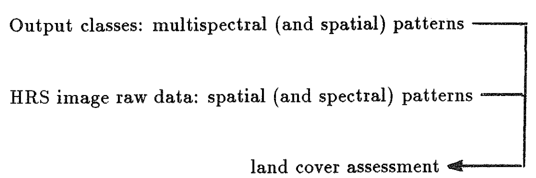
For HRMS imagery processed by a N-M filter-segmentation approach (NMHRMS) our informational hierarchy is described in Tab.1, where (a) to (f) are cue regions, (1) to (8) are output classes (each one presented in a separate labelled image). All output classes support clustering subclassification.

Some relationships between output classes and geographic elements that are likely to occur are: 1) Depending on the number of clusters and on the choice of sensors, vegetation class can be spectrally segmented in these major categories:



2)Rivers, canals, roads; 3)Built-up areas; 4) Rivers, canals; 5) Lakes, seas, ponds, rivers (whenever their elementary regions are not elongated); 6) Roads, railroads; 7) Large bare soil areas (agricultural or not agricultural), urban areas; 8) Urban areas, small bare soil areas.

These output classes should be used by the photo-interpreter as follows:



4.5.3 Unsupervised analysis

Unsupervised analysis is applied to a single output class at a time.

The analysis follows these steps: 1) Definition and initialization of the measurement space; 2) Cluster analysis by means of the ISODATA (Duda and Hart, 1973) procedure; first cluster results, statistical separability and sum-of-squared-errors (SSE) value are printed; an output image is created; 3) Clusters statistics are input to a maximum-likelihood decision rule; second cluster results, statistical separability and SSE values are printed; an output image is created; 4) Each region whose squared Mahalanobis distance from the closest cluster's mean is bigger than the 5% point of a χ^2 distribution is associated with class "unknown"; third cluster results, statistical separability and SSE values are printed; an output image is created; 5) A synoptic table of results is printed. The operator accepts either ISODATA or maximum-likelihood decision rule results.

Some of these steps are explained in detail.

1. *Measurement space*

   Standard feature domain for the clustering procedure is represented by the multispectral mean radiance values. In this case computation is very efficient and output products look satisfactory.

   Texture information, related to each segmented region, may also be evaluated (either in the original HRMS image or in the NMHRMS image bands).

   Up to seven textural parameters are extracted from gray-level co-occurence matrix (GLCM)(Haralick et al.,1976). Three parameters seem to be more effective: energy, entropy and variance (Baraldi and Parmiggiani, 1989). In particular, each region is investigated by a displacement vector that generates the co-occurence matrix correspondent to that region. In this way the main problem in GLCM computation, i.e. window size setting, is overcome.

   In texture analysis, window size selection is a compromise between contrasting needs: reducing window size causes greater sensitivity to scene noise and heavy computation time; on the other hand, window size should be smaller than the smallest object to be detected. In our case window size is as big as each region of interest. This region should be homogeneous in terms of surface feature category, that is in terms of spectral/textural information content. Noise contribution to GLCM evaluation should be reduced on NMHRMS image. Texture patterns should be more relevant on HRMS raw data.

   A mixed spectral-textural feature space, based on normalized pattern vectors, was also tested.

   Computation time for textural investigation is high in comparison with all previous computation activities while classification improvement does not seem to be relevant.

2. *Unsupervised training*

   Unsupervised training is based on natural grouping of feature vectors in the feature space. The result of training is a set of signatures, which are statistical parameters describing each training sample (supervised training) or cluster (unsupervised training). These signatures can be deleted or merged with other signatures. Thus, all feature vectors are sorted into classes, based on their signatures, by means of a classification decision rule. Depending on the training method, one or several decision rules can be used (EDAS, 1990).

   In order to discover statistical patterns that are inherent in the data, an ISODATA algorithm was adopted. This algorithm is appropriate when the clusters form essentially compact clouds that are quite well separated from one another and when there are not great differences in the number of samples in different natural groups (Duda and Hart, 1973). Main characteristics of the ISODATA procedure are: 1) It is applied in parallel: it waits until all samples are clustered before updating cluster values; 2) it is applied iteratively; 3) it optimizes the SSE.

   The SSE criterion function uses Euclidean distance as the point-to-point distance measure that is required by the clustering procedure. Clusters defined by Euclidean distance are not invariant to linear transformations. Thus, ISODATA requires a data normalization before clustering. Besides, iterative optimization guarantees local but not global optimization: this means that different starting cluster centers can lead to different solutions. Thus, an important aspect of the procedure is the choice of the starting centers.

   The method selected for the starting points was already described (Cossu, 1988; Baraldi and Parmiggiani, 1989) while a stepwise-optimal hierarchical clustering (Duda and Hart, 1973) can also be applied.

   Once clustering is completed, an output classification image is created, the SSE and the statistical separability between all pairs of clusters is evaluated.

The adopted intercluster distance measure is the Jeffries-Matusita (J-M) distance. The minimum J-M distance value and the average pairwise separability of all pairs of clusters are then given as output. These numbers can be compared to other separability listings in order to determine the best classification sorting (Swain and Davis, 1978).

To compute the J-M distances, the classes are assumed normally distributed which results in a J-M relationship involving means and covariances but no integrals. If the clusters do not present a hyperellipsoidal shape, these statistics can give a very misleading description of the data and the J-M values become meaningless. This means that the operator must rely basically on the output image, rather than on the statistical separability measures, in order to select the best classification result. Besides, ISODATA clustering is not as parametric as other clustering algorithms, meaning that it does not heavily rely on a normal distribution of input data. Then, it is more likely to produce better results with data that are not normally distributed. However, although the ISODATA algorithm is the most similar to the minimum (Euclidean) distance decision rule, the signatures obtained by this method can produce good results with any type of classification decision rule. Therefore, no particular decision rule has to be recommended over others when ISODATA clustering is applied.

These observations lead to the following classification strategy: 1) A preliminary non-parametric classification is generated by the ISODATA procedure. This result is equivalent to using a minimum distance decision rule on the same signatures that are created by ISODATA. This cluster analysis is expected to produce good results for data that are not normally distributed; 2) A second parametric classification procedure is then applied, which is more likely to produce good results for normally distributed data. Because of ISODATA compatibility to any classification rule, signatures derived from ISODATA clustering are exploited in the parametric classifier. In particular, a maximum-likelihood decision rule is adopted.

3. *Classification decision rule*

   The maximum-likelihood classifier is very parametric. It tends to overclassify signatures with relatively large values in the covariance matrix. This occurs whenever a cluster (or a training sample) presents a large dispersion in the feature space.

   During classification all feature vectors whose Mahalanobis distance from closest cluster's mean is above the 5% point of the χ^2 distribution are marked as "unclassified".

   An output image is created and statistical separability measurements are printed when unclassified feature vectors are both considered and ignored.

   The output result of cluster analysis is shown in Fig.4.

5 Conclusions

An unsupervised classification system for HRMS imagery, working on information hierarchy, was developed.

A homogeneous region segmentation process was applied. The region-based classification procedure can be easily interfaced to traditional map systems (in terms of information details and vector data format).

Spectral rules and geometric properties were used to characterize spectrally homogeneous elementary regions.

Cluster-oriented investigation, in combination with the hierarchical organization, was utilized. Through this classification scheme the operator has more information than spectral separability alone to define informational utility of the classification output classes.

Some of the unsupervised algorithm's objectives are: 1) reduction in computation time; 2) machine-aided interactive setting of a few general-purpose parameters; 3) highly structured design, that allows software blocks to be independently created, updated, and removed.

Related to this general approach to HRMS image classification some general aspects of future developments can be pointed out:

1. In the feature extraction stage, any reliable information subset (kernel data) is extracted. Some effective cue category extraction techniques (in terms of computation time or information detail level) are pixel-oriented and/or are applicable to HRMS raw data imagery. In the paper by Ton et al. (1991) this is done for water pixels, vegetation pixels and road objects (identified and labelled as major, minor and local road instances). Pixels or regions detected in this way may be excluded from any further investigation starting on N-M preprocessed images. This approach could be particularly interesting for road detection, because the conservation of linear elements in the N-M filtered image demands a large increase in pre-processing computation time.
2. Any spectral rule of interest may be introduced in the feature extraction stage. Each rule should be quite image-independent and mainly refer to VisRed and NearIr bands. In particular, a snow and cloud detection algorithm is under testing.
3. A straight boundary detection and a small region gathering detection procedure should also be implemented.

Bibliography

Baraldi,A., and Parmiggiani,F., 1989, Urban Area Classification by Multispectral SPOT Images, IEEE Trans. Geosc. Remote Sensing, vol.28, pp.674-680.

Cossu,R., 1988, Segmentazione mediante analisi della tessitura, Pixel, nos. 1/2, pp. 21-24.

Duda,R., and Hart,P.E., 1973, Pattern classification and scene analysis, J. Wiley, New York,

ERDAS Field Guide 1990

Haralick,R.M., Shanmugam,K., and Dinstein,I., 1976, Textural features for image classification, IEEE Trans. on Systems, Man and Cybernetics, vol.SMC-3, no.6,

Hilditch,C.J., 1969, Linear skeleton from square cupboards, Machine Intelligence 6, Edimburgh Univ. Press, pp. 403-420.

Julesz,B., 1981, Textons, the elements of texture perception and their interactions, Nature 290, pp. 91-91.

Ketting,R.L., and Landgrebe,D.A., 1976, Classification of multispectral image data by extraction and classification of homogeneous objects", IEEE Trans. on Geoscience Electronics, vol. GE-14, pp. 19-26.

Lillesand,T.M., and Kiefer,R.W., 1979, Remote Sensing and Image Interpretation, John Wiley & Sons.

Nagao,M., and Matsuyama,T., 1979, Edge Preserving Smoothing, Computer Graphics and Image Processing, **9**, pp.394-407.

Nagao,M., and Matsuyama,T., 1980, A Structural Analysis of Complex Aerial Photographs, Plenum Press.

Otsu,N., 1979, A threshold selection method from gray-level histograms, Trans. of IEEE on Systems, Man and Cybernetics, vol.9, no.1, pp. 63-66.

Perry,C.R., and Lautenschlager,L.F., 1984, Functional equivalence of spectral vegetation indices, Remote Sensing Environ. vol.14, pp. 169-182.

SPIDER User's Manual, 1983, Jap. Ag. Ind. Sci. Technol.

Swain,P.H., and Davis,S.M., (Eds.), 1978, Remote sensing: the quantitative approach, Mc Graw Hill

Ton,J., Jain,A.K., Enslin,W.R., and Hudson, W.D., 1989, Automatic road identification and labeling in Landsat 4 TM images, Photogrammetria, vol. 43, pp. 257-276.

Ton,J., Sticklen,J., and Jain,A., 1991, Knowledge Based Segmentation of Landsat Images, IEEE Trans. Geosc. Remote Sensing, vol.29, pp.222-231.

```
-------------------------------------------------------------------------
|     (c)      |                           NO                           |
-------------------------------------------------------------------------
      |        |     (a)      |                  NO                     |
      |        ----------------------------------------------------------
      |        | (f)  |  (e) |NO |       (b)         |        NO        |
      V        ----------------------------------------------------------
     (1)          |_____|    |  | (f) | (e) |  NO  | (f) | (e) |  NO   |
 Vegetation       |          |  --------------------------------------
                  V          |       |___|    |     |___| | (d) | NO |
                 (2)         |       |        |     |     ------------
                Roads        |       V        |     V        |      |
                             V      (4)       V    (6)       |      |
                            (3)   Rivers     (5)  Roads      |      |
                     Built_up_areas         Lakes            V      |
                                                            (7)     |
                                             Large_bare_soil        |
                                                                    V
                                                                   (8)
                                                            High_contrast
```

```
(a): Low_bright
(b): Water
(c): Vegetation
(d): Large
(e): Thin and elongated
(f): Very thin and elongated
```

Tab.1

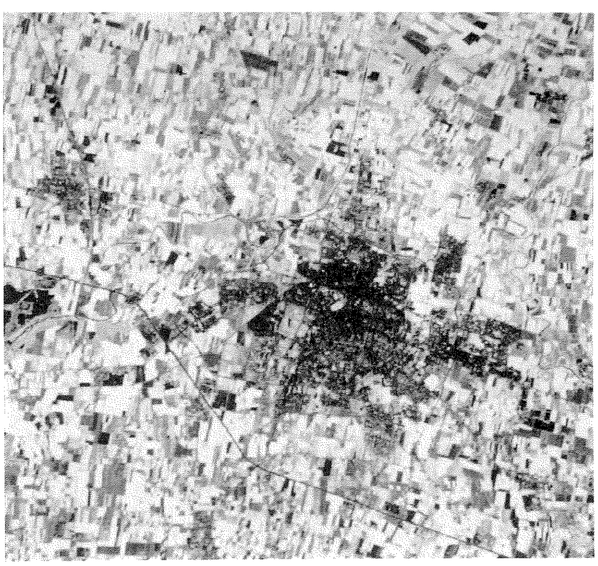


Fig.1

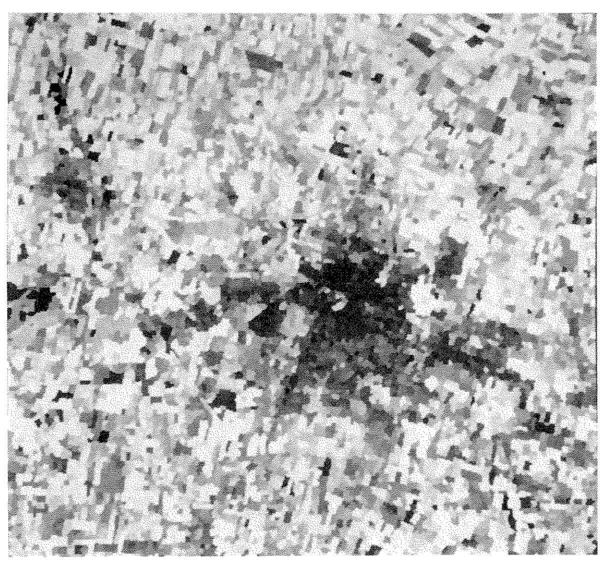


Fig.2

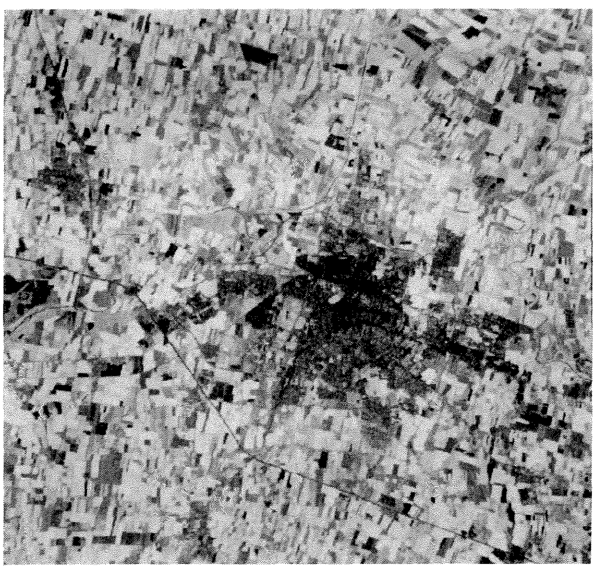


Fig.3

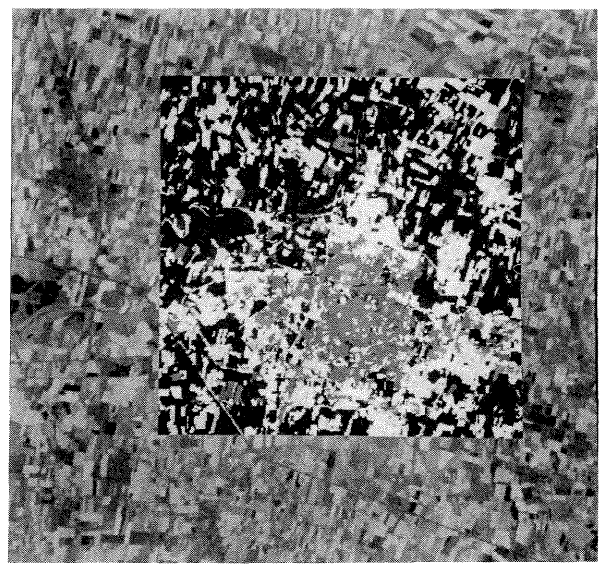


Fig.4