Scanning Microscopy Vol. 11, 1997 (Pages 1-21) 0891-7035/97$5.00+.25
Scanning Microscopy International, Chicago (AMF O'Hare), IL 60666 USA

MULTIVARIATE IMAGE ANALYSIS AND SEGMENTATION IN MICROANALYSIS

N. Bonnet[1,2]*, M. Herbin[2] and P. Vautrot[2]

[1]INSERM Unit 514 (IFR 53: Biomolecules), [2] University of Reims, Reims, France

Abstract

Instruments for microanalysis are now able to provide several images of the same specimen area. In this paper, two groups of methods are described for handling these multivariate maps. One group concerns dimensionality reduction, i.e., the projection of N-dimensional data sets onto a M-dimensional parameter space (M<N). It is shown that, in addition to linear mapping which can be performed by Multivariate Statistical Analysis, nonlinear mapping can also be performed (Multi-dimensional Scaling, Sammon's mapping, Self-Organizing mapping). The other group concerns Automatic Correlation Partitioning (ACP). With these methods, pixels are grouped into several classes according to the different signals recorded. This can be done by classical clustering methods (K-means, fuzzy C-means) or by new methods which do not make hypotheses concerning the shape of clusters in the parameter space.

Key Words: Microanalysis, image segmentation, multivariate image analysis, dimensionality reduction, clustering.

*Address for correspondence:
N. Bonnet
INSERM Unit 514, University of Reims
45, rue Cognacq Jay
51092 Reims Cedex, France
Telephone number: 33-3-26-78-77-71
FAX number: 33-3-26-06-58-61
E-mail: noel.bonnet@univ-reims.fr

Introduction

With the advent of parallel detection procedures, analytical instruments have become able to record several signals of interest for a given position of a probe on a specimen. By scanning the probe, we are able to record several maps of the same specimen. This is true for many types of microanalysis such as X-ray (XR) microanalysis, electron energy-loss (EEL) microanalysis, secondary ion mass (SIM) analysis, Auger electron (AE) analysis, X-ray fluorescence (XRF) microanalysis and nuclear probe (NP) analysis. It is even possible to combine several microanalytical techniques in some instruments, in order to increase the number of elements which can be investigated, or to perform coincidence experiments.

Simultaneous maps of a specimen enable recognition of the different chemical species which constitute the specimen, and quantification of their concentration or their spatial distribution. However, this is possible only if software tools dedicated to this purpose are available.

At present, only a few tools have been used by microscopists in this context. More tools have been developed in the context of pattern recognition, but they have not been tested in the framework of multivariate imaging. Thus, the aim of this paper is to describe the limitations of the few tools already in use and to introduce some new ones for multivariate image analysis.

In the next section, we review the tools commonly used for analyzing two or three microanalytical maps: scatterplots and interactive correlation partitioning.

Then we discuss dimensionality reduction which is a possibility for dealing with more images than two or three. We show that dimensionality reduction can be accomplished by linear methods or by nonlinear methods. Since linear dimensionality reduction is more well-known by microscopists, we emphasize nonlinear dimensionality reduction.

Subsequently, we discuss the problem of going from interactive correlation partitioning (ICP) to automatic correlation partitioning (ACP). The aim is to perform an automatic segmentation of the multivariate set, either with or without preliminary dimensionality reduction. Several methods are already available and well-documented in the field of pattern recognition, but these methods (K-means method, fuzzy C-means method) assume that clusters of patterns can be represented by hyperspheres or hyperellipsoids in the parameter space. Thus, we suggest a

new clustering method which does not make such an assumption.

Finally, we illustrate the whole set of methods on two examples: one with synthesized images and one real example.

Conclusions and suggestions for future work are given in the last section.

Tools Already in Use for Analyzing Two or Three Images (Maps) of the Same Specimen

When two maps of the same specimen are available, one way to analyze the information content of the whole data set is to build a scatterplot. This very useful tool was introduced in the context of microscopy by Jeanguillaume (1985) and in the context of microanalytical techniques by El Gomati *et al*. (1987) and by Bright *et al*. (1988).

The scatterplot relates the information content (gray levels) of image 1 to that of image 2, and is an estimation of the joint probability distribution function (pdf) p(g1, g2), where g1 is the gray level in image I1 while g2 is the gray level in image I2:

$$p(g1,g1) = \frac{card\left[I1(x,y) = g1; I2(x,y) = g2\right]}{P} \qquad (1)$$

where P is the number of pixels in images I1 and I2 and card, the cardinality of the set, is the number of pixels with gray levels g1 and g2.

Clouds of points appear in regions of the scatterplot where the joint pdf is high, while few representative points appear where the joint pdf is low. Thus, the information content of the image pair can be interpreted in terms of number, size, shape and position of clouds in the scatterplot. The number of clouds is related to the number of phases, or regions of the analyzed specimen area with relatively homogeneous composition (provided the two maps are sufficient to give an indication of the overall composition). The population of the clouds (that is the number of representative points they contain) is proportional to the spatial extension of the corresponding region. Clouds around the first diagonal of the scatterplot represent groups of pixels for which the two recorded signals are correlated while clouds around the second diagonal of the scatterplot represent pixels for which the two signals are anti-correlated. Overlapping of the clouds means that there is no clear distinction between the two regions but a gradient of composition exists between them. A boundary between two phases can also be evidenced by streaks between otherwise well-separated clouds. It should be stressed that noise or beam-specimen interaction effects may also produce overlapping distributions and streaks.

Examples of the application of this kind of analysis can be found in, among others, Browning *et al*. (1987), Prutton *et al*. (1990), and Kenny *et al*. (1992).

The scatterplot can easily be generalized to three maps (Bright and Newbury, 1991; Kenny *et al*., 1994). The principle remains identical: the aim is to estimate the joint probability distribution for three images:

$$p(g1,g2,g3) = \frac{card\left[I1(x,y) = g1; I2(x,y) = g2; I3(x,y) = g3\right]}{P} \qquad (2)$$

The only difference is in the technical tricks which have to be used for visualizing the three-dimensional scatterplot.

In addition to displaying the estimated joint pdf, the scatterplot can also be used in order to obtain a partition of the data set into groups of pixels with a homogeneous composition. This is the aim of multivariate image segmentation: partitioning the analyzed area into several regions corresponding to homogeneous sub-areas. If this task can be accomplished, then other tools can be used in order to perform subsequent quantitative analysis, on the basis of region areas or spatial distribution analysis. Up to now, the partitioning of the data set has been done interactively, hence the terminology "interactive correlation partitioning" (ICP). Clouds in the scatterplot are identified by the user and selected by means of the computer mouse. Then it is possible to return to the real space (images) in order to identify pixels and regions (sets of connected pixels) which correspond to the selected clouds (Paque *et al*., 1990). This procedure is called the "back-mapping procedure" or the "traceback procedure" (Bright and Newbury, 1991).

Obviously, the two procedures described above are insufficient in two respects: first, they are limited to data sets composed of two maps, or at most three when dedicated infographic tools (for three-dimensional display and interactivity) are available to the user; second, they require interactivity from the user and are thus not being automatic. Methods for handling a larger number of maps in a more automatic way are described in the following sections.

Dimensionality Reduction

When more than two or three maps of the same specimen are recorded, it becomes difficult, even for the very proficient human visual system, to interpret the content of the whole data set. This is because the information related to each pixel is in fact contained in an N-dimensional space, where N is the number of maps. Each pixel can be represented (in this space) by a vector whose coordinates are the N gray values in the N different maps:

$$\mathbf{V}(x,y) = \{ I_1(x,y), I_2(x,y), I_3(x,y) ..... I_N(x,y) \} \qquad (3)$$

However, due to the correlations and anti-correlations between the different maps, the true dimensionality (also called the intrinsic dimensionality) of the data set is

often smaller than the number of recorded maps. This means that the data set can often be represented in a space of dimension M lower than N, without loosing much useful information. The process of representing an N-dimensional data set in a space of lower dimension is often called a "mapping" (Note that this mapping should not be confused with the process of recording the "map" of a given element inside a specimen. The latter is an experimental process while the former is a data processing activity).

Mapping can be performed by linear processes or nonlinear processes.

Linear mapping processes

Linear processes involve defining a new space for representing the data where the new coordinates of individual objects (pixel vectors in our case) are computed according to an axes rotation matrix. The new representation space is chosen in such a way that the whole data set can be represented with a smaller number of components (M<N) than in the original representation space. Belonging to this group are the methods pertaining to Multivariate Statistical Analysis (MSA). These methods, Principal Components Analysis (PCA), Karhunen-Loeve Analysis (KLA) and Correspondence Analysis (CA) are based on the variance-covariance matrix of the data set and defined as new axes of representation the orthogonal directions of the parameter space which represent a high percentage of the total variance in the data set. (The first eigenvector is the one which explains the largest variance; the second one, which is orthogonal to the previous one, explains most of the residual variance, and so on). The coordinates of pixels on the new representation axes (eigenvectors) can be used to create new images (called eigen-images). Since the number of significant eigen-images is generally smaller than the number of original images (M<N), dimensionality reduction usually results. Such techniques have been used for many years in high resolution electron microscopy of macromolecules, when one has to find different classes of images and to explain the differences between subsets (Van Heel and Frank, 1981; Frank and Van Heel, 1982).

The application to microanalytical images has also begun (Geladi and Esbensen, 1989; Bonnet *et al*., 1992; Bonnet and Trebbia, 1992; Van Espen *et al*., 1992; Swietlicki *et al*., 1993; Quintana and Bonnet, 1994a,b; Bonnet, 1995a; Trebbia *et al*., 1995). Thus, we will not describe it further in this paper, but we would rather concentrate on nonlinear mapping methods.

Let us only remark that when the N-dimensional data set can be mapped onto a two-dimensional space (M=2) or a three-dimensional data set (M=3) the scatterplot technique can be used (with the two or three eigen-images) in order to estimate the joint probability distribution function corresponding to these eigen-images, and then to analyze the data set in terms of number of clusters. But when the N-dimensional data set cannot be reduced so efficiently (M>3), the techniques described in the previous section become difficult to apply and other techniques must be used. In the following sections, we describe some techniques related to nonlinear mapping.

Nonlinear mapping processes

As early as the beginning of the 1960s, techniques were developed by American psychologists in order to visualize the proximities between "objects" (stimuli) described by N features, i.e., represented in an N-dimensional space. Their idea was to project these data onto a subspace of reduced dimension (generally, M=2), with as little distortion as possible. This requirement means that the separation between objects must be maintained as well as possible. This idea led to several algorithms, named Multi-Dimensional Scaling (MDS) (Shepard, 1962; Kruskal, 1964) or Sammon's mapping (Sammon, 1969). To our knowledge, the only application of these techniques in microscopy was performed by Radermacher and Frank (1985). In the next subsections, we describe how such techniques can be used for solving the problem we have in mind, that is the mapping of multivariate images onto a space of reduced dimensionality, in order to visualize the content of the whole data set and to deduce useful information. Other methods for mapping are neural networks methods, the Self-Organizing Map (SOM) (Kohonen, 1984) being a typical example.

An empirical method for nonlinear mapping. As a preliminary investigation of the problem, we have suggested the following approach (Bonnet *et al*., 1995): since the scatterplot is only efficient when two images are handled simultaneously, the first step of dimensionality reduction should consist of replacing the N experimental images by two new images, computed from the experimental ones. For this, we define the concept of "observers" of a data set: "observers" are positions of the N-dimensional parameter space from which the data set is looked at (see Figure 1 in Bonnet *et al*., 1995). By "looked at", we mean that some measure is defined for a comparison of the observer position and the data point position. For simplicity, assume that this measure is the Euclidean distance. Since the distance between the observer and each pixel of the maps can be computed, a new image can be created which stores the information "seen" by the observer. If two such observers can be defined, two images are produced and thus, a scatterplot can be built. Since the two synthesized images contain information related to the whole data set, it can be expected that the scatterplot displays also this information, in terms of clouds of points. The key point which we have not addressed yet is the choice of "good" observers (a "good" observer is one which "sees what it is interesting to see"). In Bonnet *et al*. (1995), we have described two possibilities for choosing observers (other possibilities could also be defined): the corners and the diagonals of the

N-dimensional hypercube in the parameter space. There are thus many possibilities to define pairs of observers. For example, in a five-dimensional space, there are 32 corners and thus 496 possible pairs of observers. Among these many possibilities, some of them are interesting, in the sense that the corresponding scatterplot displays a useful mapping of the data set, and some of them are not interesting, in the sense that they produce a degenerated (non optimal) map of the data set, with overlapping clusters. At this stage, the only possibility we have described is to compute all possible scatterplots, to display them, and to choose the ones we think are the most appropriate for a given purpose. For instance, the number of clouds observed in a scatterplot is necessarily a lower bound of the number of clusters present in the data set. Thus, if the purpose of the mapping is to know the number of regions with a significantly different composition, one has to choose the scatterplot with the highest number of separated clusters (this means that the observers are situated at positions where they can "see" the different clusters without significant disturbance by the other clusters).

Nonlinear mapping obtained by minimization of a distortion criterion. As stated previously, psychologists were the first to define criteria for obtaining an optimal nonlinear mapping. We define two objects (pixels in our case) i and j (i=1...P; j=1...P) and D_{ij} and d_{ij} the distances between them in the original N-dimensional space and in the M-dimensional destination space. We will restrict to M=2 in this discussion.

Examples of criteria which can be used to quantify the distortion of the data set after mapping onto a lower-dimensional space are:

- the stress criterion (Kruskal, 1964):

$$s = \sqrt{\frac{\sum_{j=1}^{P}\sum_{i=1}^{j-1}\left(D_{ij} - d_{ij}\right)^2}{\sum_{j=1}^{P}\sum_{i=1}^{j-1} D_{ij}^2}} \qquad (4)$$

- Sammon's criterion (Sammon, 1969):

$$E = \frac{\sum_{j=1}^{P}\sum_{i=1}^{j-1}\left(D_{ij} - d_{ij}\right)^2 / D_{ij}}{\sum_{j=1}^{P}\sum_{i=1}^{j-1} D_{ij}} \qquad (5)$$

As stated by Radermacher and Frank (1985), these two criteria can be generalized to:

$$C = \frac{\sum_{j=1}^{P}\sum_{i=1}^{j-1}\left(D_{ij} - d_{ij}\right)^2 / w_{ij}}{\sum_{j=1}^{P}\sum_{i=1}^{j-1} D_{ij}^2 / w_{ij}} \qquad (6)$$

where w_{ij} is a weighting factor which can be used to preserve short distances at the expense of long distances, or the reverse. Once a criterion is defined, nonlinear mapping reduces to an optimization (actually, a minimization) problem and can be solved by any of the methods available nowadays for this problem (steepest descent method, simulated annealing, genetic algorithms, etc.). Up to now, a Newton-like steepest descent method has mainly been used: coordinates of point i in the destination space are changed iteratively according to the formula:

$$x_k^i(t) = x_k^i(t-1) - \alpha \frac{\frac{\partial C(d_{ij})}{\partial x_k^i}}{\left|\frac{\partial^2 C(d_{ij})}{\partial \left(x_k^i\right)^2}\right|} \; ; k = 1..M \qquad (7)$$

The efficiency of the mapping, in terms of preservation of distances (or, equivalently, limitation of distortion) can be ascertained through several additional tools. First, a scatterplot can be built which relates the distances (d_{ij}) in the destination space to the distances (D_{ij}) in the original space. In this scatterplot, distortion manifests itself as points (or clouds) off the first diagonal. This deviation can be quantified by the entropy of the scatterplot.

One of the main drawbacks of this approach for nonlinear mapping is the computation time (remember that these methods were developed for analyzing data sets composed of tens of objects, while we are attempting to use them for hundreds of thousands of pixels, and that the simple computation of ($\Sigma_i \Sigma_j D_{ij}$) has a complexity P^2, where P is the number of pixels). Thus, we have to choose among several solutions when the size of images exceeds some thousands of pixels:

- algorithmic methods try to perform iterative mapping on a parallel basis (all objects are moved simultaneously, instead of one at a time) (Demartines, 1994),

- a subsampling is performed so that only a limited portion of pixels is submitted to the analysis (for a discussion of subsampling, in the context of linear mapping (Geladi, 1995),

- the "observers"-based empirical solution described in the previous section is used, and "good" solutions are selected on the basis of the criteria defined in this section.

Towards the Automatic Segmentation of Multivariate Microanalytical Maps

In the previous section, we have considered the problem of displaying N-dimensional data sets within a space of reduced dimensionality (generally, M=2). The main purpose was to help in the qualitative interpretation (through visualization) of the data sets, in terms of number of distinct clusters (corresponding to different compositions of the specimen), of cluster overlapping, etc.

Another aim is to generalize the interactive correlation partitioning (ICP) approach to a number of images greater than two or three: if the (linear or nonlinear) mapping onto a space of reduced dimension is successful, interactive techniques can be used to isolate a cloud of points and to go back to the original image space, in order to identify pixels and regions which constitute this cloud. Then quantification (e.g., surface, spatial distribution) becomes an easy task.

Another (more ambitious) task consists of trying to avoid any interaction by the user with the data set and to go towards a complete automation of the process of identifying regions with a homogeneous composition, that is the automatic segmentation of multivariate images. We conducted a preliminary investigation of this problem in (Bonnet, 1995b). More specifically, we investigated two methods: one starting from concepts in the field of automatic classification and the other one starting from the concept of monovariate image processing (the region growing approach). In this paper, we limit ourselves to the classification approach, but we present several extensions to our preliminary study.

Automatic classification is a general problem which has been studied for a long time, but continues to be of interest. Thus, we will discuss techniques well-documented in the field of pattern recognition, as well as a new technique we have proposed recently.

Automatic classification (as any technique relevant to pattern recognition) can be approached either as a supervised method or as an unsupervised one. In the former case, the user has to train the classifier with a set of prototypes (objects which belong to previously known classes) while in the latter case, the classification is made on the basis of the data themselves, without reference to any *a priori* knowledge. Although supervised classification may have some meaning for microanalytical studies, we limit ourselves to the discussion of unsupervised classification techniques, also called clustering techniques. The aim is to group data (pixels) into several clusters, on the basis of the different measurements. This replaces interactive correlation partitioning (ICP) with automatic correlation partitioning (ACP). This can be done either in the original (N-dimensional) parameter space or in the reduced (M-dimensional) space after linear or nonlinear dimensionality reduction. We will begin with a discussion of classical techniques, and of the associated problems. We will then describe the new technique we have suggested.

The K-means technique.

This technique is described in detail in many textbooks on pattern recognition (Duda and Hart, 1973; Fukunaga, 1990) and in our preliminary paper on the subject (Bonnet, 1995b). Thus, we simply recall its principle. The K-means technique performs an iteratively refined partitioning of the data set into a predefined number of classes (K). At each step of the iterative process, objects are associated to the class whose center is the closest, according to a chosen distance (e.g., Euclidean, city-block, Mahalanobis). Then the centers of the different classes are updated, taking into account the objects which now belong to these classes. The process is iterated until convergence.

This method is easy to implement, but unfortunately, it suffers from a certain number of drawbacks:

- it is very sensitive to the initialization of the class centers. We have described in (Bonnet, 1995b) a method which works often as well (and faster) than the repetition of the procedure with a large number of different random initializations,

- when the number of classes is unknown, the procedure must be repeated with a varying number of classes, and the optimal solution must be chosen. The question raised is thus: “what is an optimal solution for the clustering problem?” In Appendix 1, we give a non-exhaustive list of clustering criteria which can be used to characterize the quality of a partition. Unfortunately, these criteria do not always predict the same number of clusters for a given data set,

- the main problem concerns the shape and size of clusters (in the parameter space). The K-means technique is efficient in the case of hyperspherical (when the Euclidean distance is used) or hyperelliptical (when the Mahalanobis distance is used) clusters. But it is much less efficient when clusters of arbitrary shape are present (this is due to the use of the distance to a center). Moreover, these clusters must have approximately the same extension, especially when the Euclidean distance is used. Obviously, these requirements are not always satisfied.

When the process of clustering is performed, several graphical tools can be used to display and analyze the results. First, since each object is labeled (i.e., assigned to a class), a map of these labels can be built and displayed, which indicates the spatial distribution of classified objects (pixels). In addition, the experimental values obtained (for the different signals) can be averaged for all objects clustered in the same class, thus producing a new multi-dimensional image where the pixels are described by the vectors:

$$\mathbf{V'}(x,y) = \{I'_1(x,y), I'_2(x,y), I'_3(x,y) .... I'_N(x,y)\} \quad (8)$$

with $I'_n(x,y) = 1/\text{card}(C_i).\ \Sigma\, I_n(x,y)$, the summation being performed over all pixels belonging to the class C_i. From this averaging operation, new images (one for each recorded signal) are constructed, which generally result in an important signal-to-noise ratio improvement. Finally, a map of confidence can be built: one way to do this is to compute, for each pixel, the correlation coefficient ρ between the vector describing the experimental data (**V**) and the vector corresponding to the averaged data (**V'**):

$$\rho(x,y) = \frac{E[V,V'] - E[V].E[V']}{sd(V).sd(V')} \quad (9)$$

with E = expectation value and sd = standard deviation.

When the pixel is classified unambiguously, this correlation coefficient is close to one, but when the pixel is more or less arbitrarily classified (when the chosen number of classes is wrong, for instance), the correlation coefficient is diminished.

The fuzzy C-means (FCM) technique.

When applying the K-means technique, we assume that one object belongs to one class or another, depending on the relative closeness of their class centers. Although this assumption can be considered as being true (within the limits indicated previously) at the end of the iterative process, it seems better to avoid using it at the very beginning of the process, when class centers are still imperfectly defined. One theory which does not assume strict membership is fuzzy logic. Within this theory, an object may be a member of several classes simultaneously, with partial degrees of membership (the sum of these degrees equals one). For instance, degrees of membership of an object i to a class c may be defined as:

$$\mu_{ic} = \frac{\left[1/d_{ic}^2\right]^{1/(m-1)}}{\sum_{c=1}^{C}\left[1/d_{ic}^2\right]^{1/(m-1)}} \quad (10)$$

where C is the number of classes (we have kept this usual notation here, but it must be recognized that C=K, according to the notation of the previous section), d_{ic} is the distance of object i to the center of class c and m is a fuzzy parameter (m=2 is used commonly). The iterative process used to perform fuzzy C-means clustering is very similar to that of the K-means technique. For a set of class centers, the membership degrees are computed for all the objects to classify. Then the centers of the classes are updated, according to:

$$x_c = \frac{\sum_i \mu_{ic} \cdot x_i}{\sum_i \mu_{ic}} \quad (11)$$

and the process is repeated until convergence. Then the defuzzification step takes place: objects are assigned to the class for which the degree of membership is maximum.

The fuzzy C-means technique has been claimed to be superior to the K-means technique in a number of ways, including a better convergence and the avoidance of being trapped in local minima. For the few examples for which we have compared the two techniques so far, we found that FCM was not greatly superior to KM. (FCM works well when KM works well, but does not surpass it significantly when KM fails, i.e., in situations indicated at the end of the last section). However, one advantage of FCM is that one can compute the degree of confidence of the classification result for every pixel. One way to do it is to compute the entropy of the degrees of membership:

$$H(x,y) = -\sum_c \mu_{ic} \cdot \log(\mu_{ic}) \quad (12)$$

This quantity can be displayed as an image, visualizing the degree of certainty with which the classification of the pixels was performed. This can help determine the "true" number of classes, because when the chosen number is incorrect, many pixels are difficult to classify, and this can be observed in the confidence map.

Other criteria can also be used to determine the number of classes. A non-exhaustive list is given in Appendix 2.

The FCM technique is not supposed to solve the problem of non-spherical and non-elliptical clusters mentioned previously.

A technique for finding clusters of arbitrary shape.

We now describe the application of a general technique, which is not limited to microanalytical maps, not even to image segmentation (more details on this technique, together with relevant discussion, can be found in Herbin *et al*. (1996). The technique is non-iterative, and consists of two steps.

The first step consists of an estimation of the continuous probability density function (pdf). This can be achieved by the Parzen window technique: in the parameter space (either the original one or the reduced one, after linear or nonlinear mapping), one N- or M-dimensional smooth kernel h is assigned to any point $\mathbf{r}_i$ (representative of the pixel i) and these kernels are added. The result constitutes an estimation of the pdf (Duda and Hart, 1973):

$$pdf(r) = \sum_{i=1}^{P} h(r - r_i) \quad (13)$$

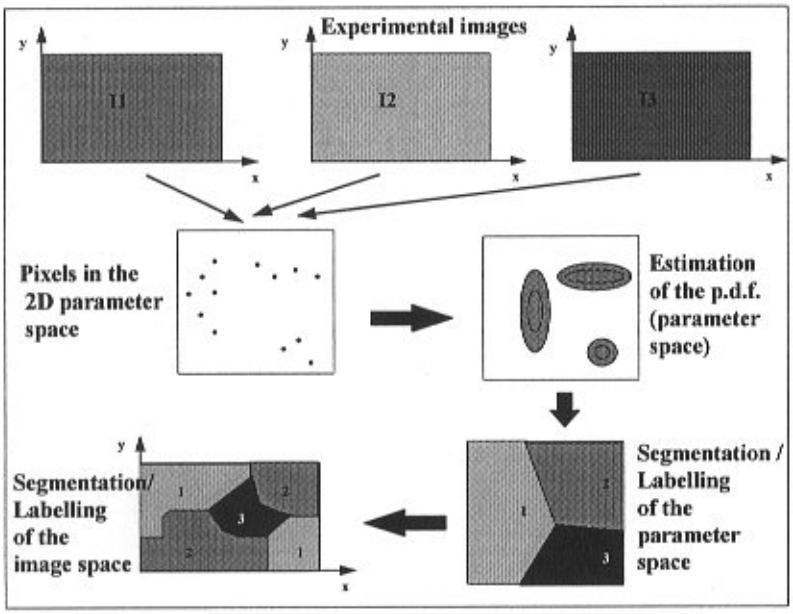

Figure 1. Illustration of the clustering procedure based on the estimation of the probability density function (pdf) followed by the clustering of the parameter space by the watershed technique.

where **r** represents any position in the parameter space and h is a kernel (a Gaussian kernel for instance). It is assumed that this estimated pdf is composed of several peaks (or modes), which represent the different classes present in the data set. Thus, we have to detect the zones of influence of these different peaks. Finally, points (i.e., pixels) within these different zones are attributed to the different classes. There are several ways to define the zones of influence of the different peaks, i.e., the subsets of the parameter space corresponding to a particular cluster (regardless of its shape). In the first description of the method (Herbin *et al*., 1996), we have reported a variant based on the iterative thresholding of the estimated pdf, starting from low values. Thresholding is then followed by a technique known as the skeleton by influence zones (SKIZ) (Serra, 1982). Other techniques based on mathematical morphology can also be used, for instance the concepts of catchment basins and watersheds (Beucher, 1992; Beucher and Meyer, 1992). For this technique there is no need to perform segmentation: the pdf is inverted; the local minima thus serve as holes from which a simulation of immersion by water is performed (Vincent, 1990) and in this way the boundaries between clusters can be located.

This technique (whatever the variant), does not involve the concept of distance to a center of a class. Thus, the shapes of the clusters do not influence the results of classification. In the same way, the extension and population of the clusters are also irrelevant (within limits).

One important point which remains to be discussed is that of the number of classes. The number of classes detected (as modes of the estimated pdf) depends largely on the width of the kernel h(**r**) used for estimating the pdf. For instance, when a Gaussian kernel is used, the number of distinct peaks depends on the standard deviation (sd) of the Gaussian kernel (the larger the sd, the smaller the number of peaks detected). The idea is thus to compute the estimation of the pdf for different values of the standard deviation and to plot the number of modes detected as a function of this parameter. What is observed frequently on this curve is a plateau, which corresponds to the number of classes present. Of course, it may happen that several plateaus appear, reflecting the fact that the data set may be composed of several classes, which can themselves be divided into several subclasses. In this situation, the user has to choose the resolution at which he wants to perform the classification task. The whole process is schematized in Figure 1.

The technique described above is not limited to a two-dimensional parameter space (i.e., to two images only). But when the number of maps is larger than three, it is better to perform dimensionality reduction first, because the computations, such as pdf estimation or the watershed technique, become prohibitively long in a multi-dimensional space.

Dimensionality Reduction and Automatic Correlation Partitioning: Illustration

In this section, we illustrate the methods described previously with two examples, one with synthesized images, the other with real images taken from X-ray fluorescence microanalysis. There is of course no restriction to this particular type of microanalysis.

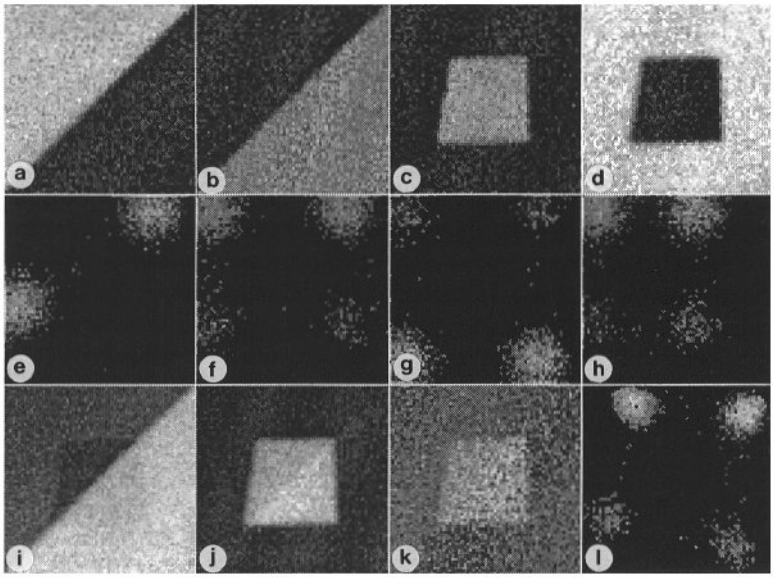

Figure 2. (**a-d**) Four simulated maps from which we expect to determine the number of regions with homogeneous composition. (**e-h**) Four of the six scatterplots which can be built with the four images. Some scatterplots display the four expected clusters, but not all of them. (**i-k**) First three eigen-images obtained after linear mapping by Correspondence Analysis. (**l**) Scatterplot built from the first two eigen-images (fig. 2i-j). As expected, four clusters are observed, corresponding to the four different regions.

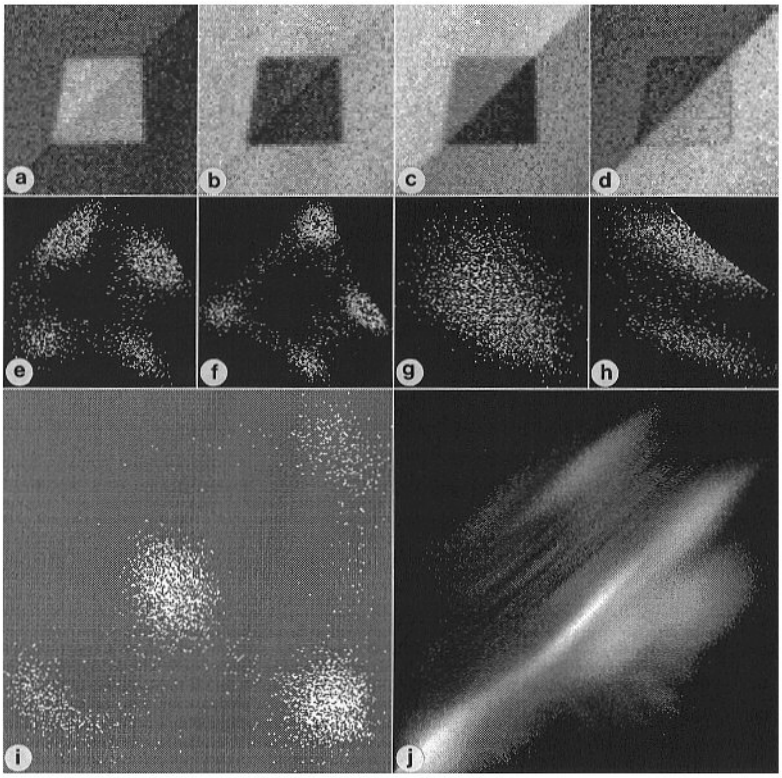

Figure 3. (**a-h**) Application of our heuristic method for nonlinear mapping. (**a-d**) Four images synthesized from the computation of distances of points in the four-dimensional space to "observers" placed at four corners of the hyperspace. (**e-h**) Four scatterplots (among the 120 scatterplots computed) built by combining pairs of synthesized images of distance. Several of them (**e**, **f**) display the four expected clusters, while others (**g**, **h**) do not. (**i-j**) Application of the nonlinear Sammon's mapping method: (**i**) Result of the automatic nonlinear mapping onto a two-dimensional mapping space. The four expected clusters can be observed. (**j**) Scatterplot built from the interpoint distances (D_{ij}) in the original four-dimensional space and the interpoint distances (d_{ij}) in the mapping space. This scatterplot serves only to check the quality of the mapping.

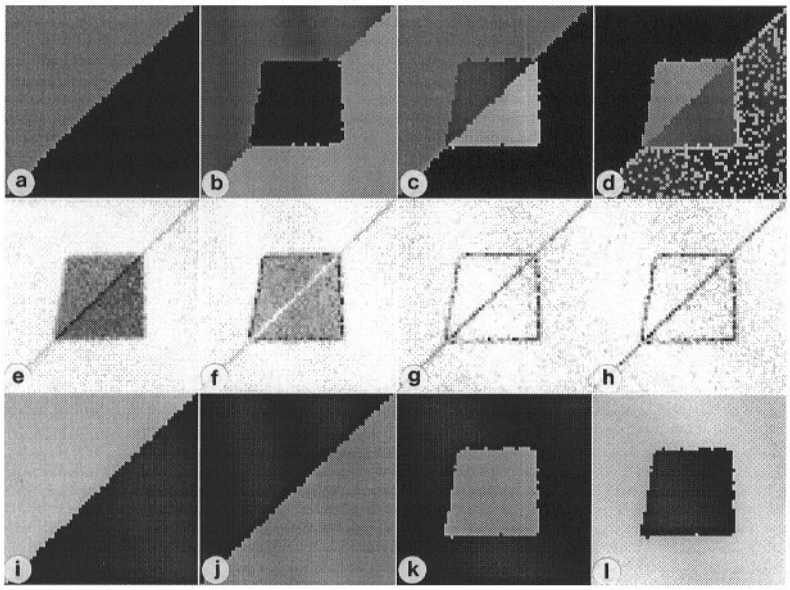

Figure 4. Results of segmentation using the K-means clustering (KMC) technique. a-d) Results of segmentation (pixels belonging to the different classes are represented by different gray levels) for a number of classes K equal to 2, 3, 4 and 5, respectively. (**e-h**) Display of the local (on a pixel basis) correlation coefficient between the experimental vector (four components) and the vector obtained after classification and averaging, for the four classification results displayed in Figure 4**a-d**. Pixels with a small correlation coefficient are represented in dark. For four classes (or more), only border pixels are badly correlated with their classified counterpart. (**i-l**) The four original images after averaging according to the classification into four classes.

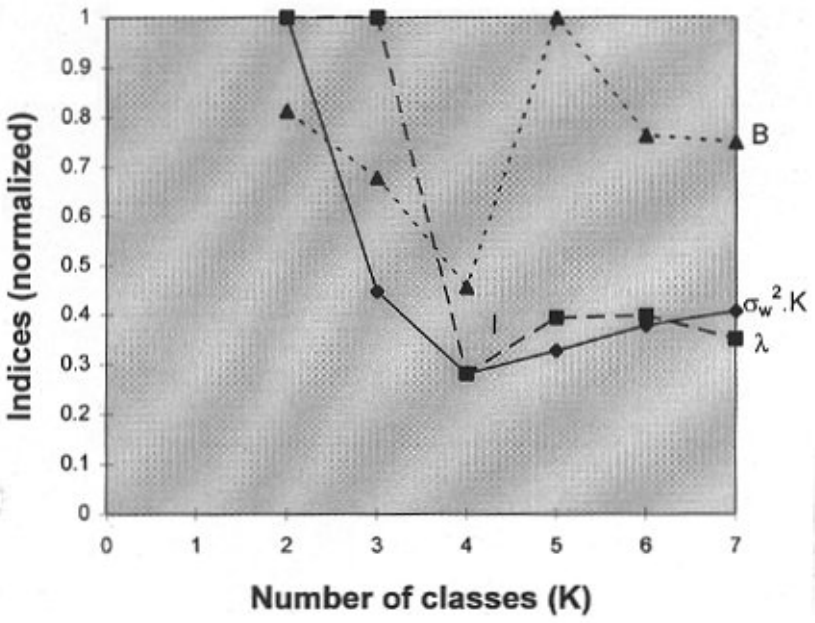

Figure 5. Some indices of the quality of the partition (see Appendix 1 for their meanings) plotted as a function of the number of classes K. (Classification with the K-means procedure). Minima of these criteria are obtained for K=4, indicating that this is the number of classes present in the data set.

Simulation example

We simulated four noisy images assumed to be representative of a situation where we want to depict the number of regions of the specimen with approximately similar composition (on the basis of four recorded signals: A, B, C, D) and to deduce the relative area of each region. This is a rather simple example (signals A and B are anti-correlated, so are signals C and D) and we can expect to find four regions with homogeneous composition (noted (1010); (0110), (1001) and (0101) for simplicity, where 1 stands for the presence of signal X and 0 stands for its absence).

The four images are displayed in Figures 2a-d. Four of the six scatterplots are displayed in Figures 2e-h. We can see that some of them give an indication about the four clusters of pixels, but some of them (Figure 2e for instance) do not, because only a limited part of the total information is taken into account. In more complicated situations, it may happen that no scatterplot displays the true number of classes.

One way to improve the situation is to perform dimensionality reduction. Linear dimensionality reduction can be performed through MSA. Figures 2i-k display the

Table 1. Results for the simulated data set.

Region	Area (pixels)	Averaged values for the four signals			
(1010)	1676(1686)	A: 197 (200)	B: 23 (20)	C: 22 (20)	D: 221 (225)
(1001)	361(366)	A: 197 (200)	B: 23 (20)	C: 167 (175)	D: 30 (20)
(0110)	1621(1613)	A: 23 (20)	B: 149 (150)	C: 23 (20)	D: 220 (225)
(0101)	438(431)	A: 26 (20)	B: 147 (150)	C: 169 (175)	D: 30 (20)

Values prior to addition of simulated noise are shown in brackets.

three eigen-images obtained using Correspondence Analysis [(CA); see Trebbia and Bonnet (1990) for details concerning the implementation]. The scatterplot built from the first two eigen-images is displayed in Figure 2l. For this simple example the four clusters are evident, but in more complex situations, the information can be spread onto more than two eigen-images and thus, the scatterplot approach may still be insufficient.

In this case, nonlinear dimensionality reduction may be attempted. Figure 3 displays some results obtained with our empirical method. The information (normalized distances) seen by "observers" placed at several corners of the four-dimensional parameter space is displayed in Figures 3a-d. We can observe that the existence of four regions with different composition is coded in these computed images. Figures 3e-h display some scatterplots built from pairs of these images. We can see that some of them (e,f) help to distinguish separated clouds of points (and can thus be used for interactive or automatic correlation partitioning) while some others (g,h) provide only a blurred view of the pixel clusters (and cannot be used for partitioning).

Better results can be obtained by optimized nonlinear mapping ("optimized" means that we force the mapping to preserve the distances between points in the original N-dimensional parameter space). Figure 3i represents the results of a mapping ($R^4 \rightarrow R^2$) obtained with Sammon's algorithm. Here, the four clusters are clearly evident. Figure 3j represents the scatterplot (D_{ij} - d_{ij}), which permits us to verify that the mapping onto the two-dimensional space preserves the original distances (weak distortion).

So, the linear and nonlinear mapping methods help the user to interpret the data set in terms of the number of independent clusters, shapes of clusters, etc. The next step is to depict the corresponding regions in the image space. This can be done starting from the original data set or from data mapped onto a reduced parameter space.

Figures 4a-d display the results of segmentation obtained using the K-means technique (applied to the original 4D data set), for two, three, four and five classes, respectively. Figure 5 shows the curves of quality criteria computed as a function of the number of classes (these criteria are described in Appendix 1). As expected, the optimal number of classes is found equal to four, but the situation is not always as favorable. In addition to the segmentation itself, several additional images can also be produced. For instance, the degree of confidence in the classification results (computed, for each pixel, as the correlation coefficient between the original vector and the vector representing the class center) can be displayed (see Figures 4e-h for the four previous classification trials). Also, when the optimum number of classes is selected, it becomes possible to average the experimental values for all those points which belong to the same class, which results in a smoothed image series (see Figures 4i-l, obtained for K=4).

The segmentation allows us to perform quantitative evaluations of the data set, in terms of relative area and of composition. For this simulated data set, the results obtained are given in Table 1.

Similar results are obtained using the fuzzy C-means technique. Some of them are displayed in Figures 6 and 7. Figure 6 shows different results of classification into two to five classes (a-d), the local entropy of the membership degrees for these different classifications (e-h), from which it can be deduced that the best classification is obtained for four classes, and the four averaged images obtained for a classification into four classes (i-l). Figure 7 shows different criteria (described in Appendix 2) obtained as a function of the number of classes, which (except HTmin) confirm that the best partition is obtained for four classes.

It should be stressed that these good results come from the fact that the four clusters were relatively well separated (with little overlapping) and of hyperspherical shape. When one of these conditions is not fulfilled, things become harder. This is why more sophisticated methods, like the one illustrated below, are necessary.

We start with Figure 8a, which is a reproduction of Figure 3i and represents the nonlinear mapping of the four-dimensional data set onto a two-dimensional parameter space (Note that Figure 2l, which is the result of a CA-

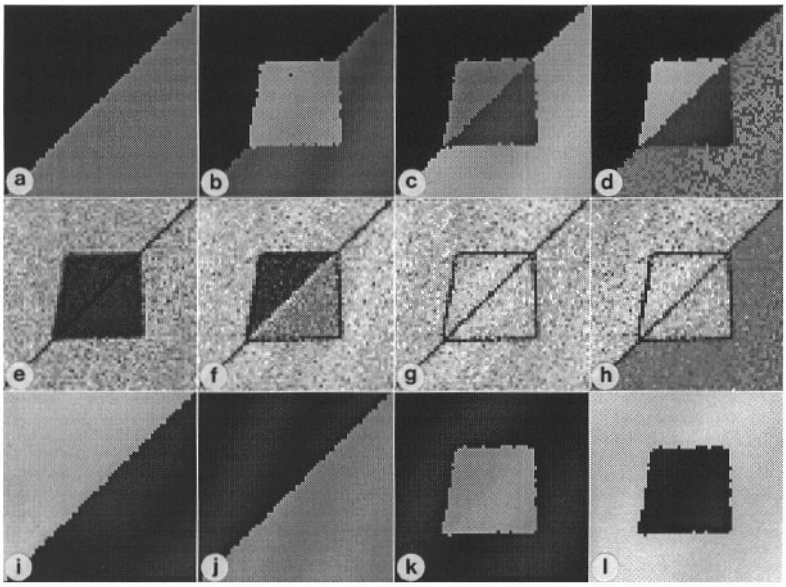

Figure 6. Results of segmentation using the Fuzzy C-means clustering (FCMC) technique. (**a**-**d**) Results of segmentation (pixels belonging to the different classes are represented by different gray levels) for a number of classes C equal to 2, 3, 4 and 5, respectively. (**e**-**h**) Display of the local (on a pixel basis) entropy of the degrees of membership to the different classes, for the four classification results displayed in Figure 6a-d. Poorly classified pixels are represented in dark. (**i**-**l**) The four original images after averaging according to the classification into four classes.

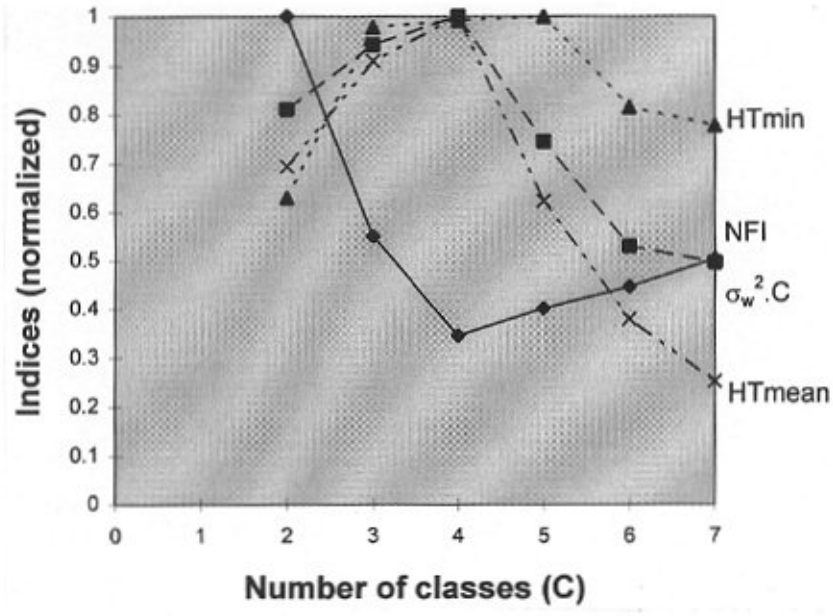

Figure 7. Some indices of the quality of the partition (see Appendix 2 for their meanings) plotted as a function of the number of classes C. (Classification according to the fuzzy C-means procedure). Extrema of these criteria (minimum of $\sigma_w^2.C$ and maxima of NFI, HTmin and HTmean) are obtained for C=4, indicating that this is the number of classes in the data set.

based linear mapping could be used as well, in this case). Figure 8b displays the result of the estimation of the probability density function with a Gaussian Parzen window (sd = 25 out of 256 quantization levels). Figure 8c displays the result of segmenting the parameter space automatically, by the watershed approach explained above (four clusters were found). Figure 8d shows the equivalent result of segmentation, but in the image space. We can observe that the difference from the result obtained by classical clustering techniques is rather slight, but this is not always the case. Figure 9 shows the number of depicted clusters (modes of the estimated pdf) as a function of the standard deviation of the Parzen smoothing kernel. A plateau (corresponding to four classes) is observed for sd ranging from 15 to 35 quantization levels. The value sd = 25 used for computing Figure 8b was chosen at the middle of this plateau.

Real example

Figures 10a-e display five micro X-ray fluorescence (μ-XRF) maps of a specimen of granite, recorded at the Department of Chemistry, University of Antwerp. The maps are those of K, Ca, Fe, and Sr. Their size is 39x40 pixels. The study of a similar specimen, including the combined use of Principal Components Analysis (PCA) and K-means

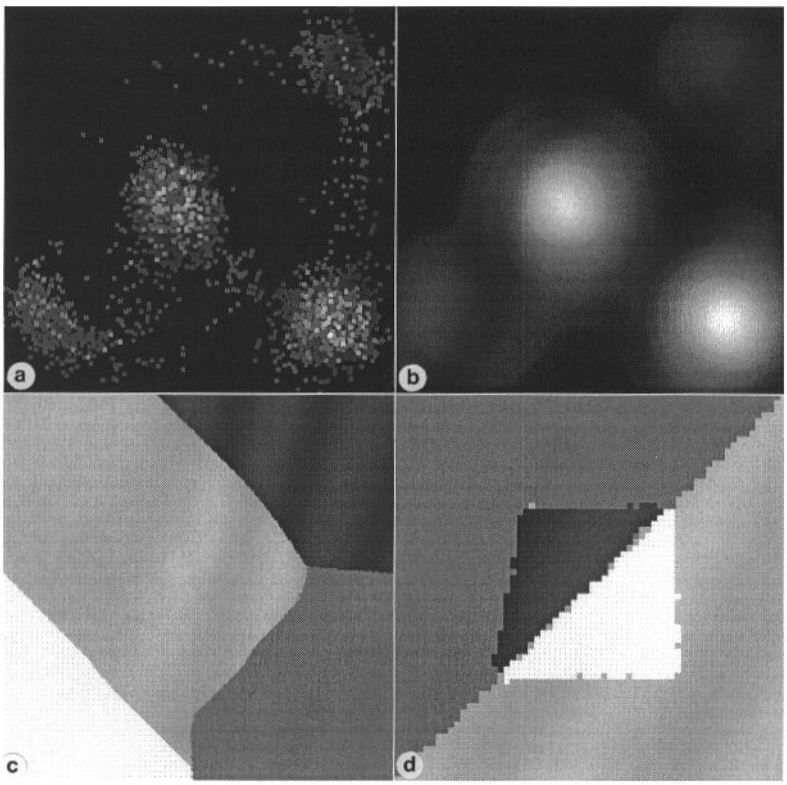

Figure 8. Application of a new method for clustering, which does not make assumptions concerning the shape of clusters. (**a**) Representation of pixels by points in a two-dimensional parameter space (this was obtained by nonlinear mapping, see fig. 3i). (**b**) Estimation of the probability density function with a Gaussian Parzen window (standard deviation = 25 quantization levels). (**c**) The parameter space is labeled according to the zones of influence of the four modes of the pdf, according to the watershed algorithm. (**d**) The image space is segmented according to the labels of the parameter space.

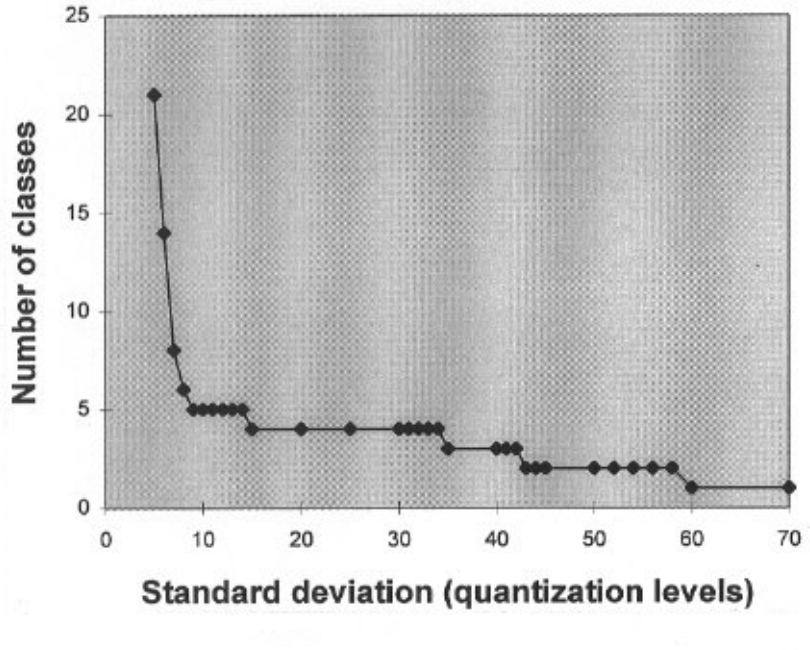

Figure 9. Curve of the number of modes of the probability density function as a function of the standard deviation (sd) of the Gaussian Parzen window. A plateau is observed for sd ranging from 15 to 35 quantization levels. Thus, a value of sd = 25 was chosen for computing Figure 8b.

clustering (KMC), was conducted by Vekemans *et al*. (1997). Readers interested in the interpretation of the classification results are referred to this paper.

Some of the methods described above are illustrated with this experimental data set.

Figures 10f-i display the four eigen-images obtained by Correspondence Analysis. Figure 10k displays the scatterplot obtained by combining the first two eigen-images. We can observe three main clusters. One of them seems to be composed of two smaller clusters. Figure 10m shows the result of Sammon's mapping (projection of the five-dimensional data set onto a two-dimensional parameter space). Again, we can make similar observations (three or four clusters are present). Figures 11a and 11b display the

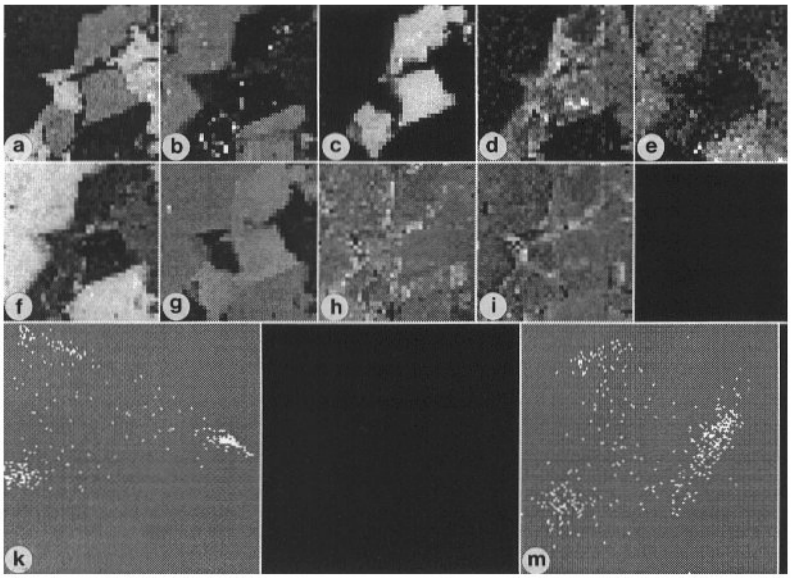

Figure 10 (**a**-**e**) Five micro X-ray fluorescence images (representing maps of K, Ca, Fe, Rb and Sr) of a specimen of granite (courtesy of K. Janssens, Department of Chemistry, University of Antwerp). (**f**-**i**) First four eigen-images obtained by Correspondence Analysis. (**k**) Scatterplot built from the first two eigen-images (f and g). Three clusters can be observed. (**m**) Result of Sammon's nonlinear mapping to a two-dimensional parameter space. Three clusters are also observed.

estimated pdf (the estimation was performed with sd = 25 and sd = 15 quantization levels, respectively, for a parameter space sampled with 256 quantification levels). In the former case, the number of modes is equal to three. Figures 11c and 11d represent another way of representing the probability density functions, as 3D plots in which density is coded as height. The corresponding classified parameter space (classification performed with the watershed technique) and classified image space are represented in Figures 11e and 11g, respectively. In the latter case, the number of modes of the pdf is equal to four. The corresponding labeled parameter space and image space are shown in figures 11f and 11h, respectively. Figure 12 displays the number of classes (modes of the estimated pdf) as a function of the standard deviation used for the estimation. The values sd = 25 and sd = 15 quantization levels (Figures 11a,b) were selected on the observed plateaus. Figure 13 displays the results obtained when the classical K-means clustering approach is applied to the original (5-dimensional) data set. Figure 15 displays the results obtained when using the fuzzy C-means technique. Figures 14 and 16 are curves of different classification quality criteria, computed for the K-means and fuzzy C-means algorithm, as a function of the number of classes of the partition. When comparing these different results of classification of the same data set, one can make the following observations. Up to three classes, the results produced by the different algorithms are very similar. The three classes were identified by Vekemans *et al*. (1997) as corresponding to the K-rich microcline phase, the Fe-rich mineral grains and the Ca-rich albite phase. When we try to split the data into four classes (as suggested by some of the criteria of a "good" partition), things become more complicated. The classical methods for clustering (K-means and fuzzy C-means techniques) tend to find the fourth class at the periphery of the three classes identified previously, which seems to indicate a more or less continuous transition between them. On the other hand, the new technique we have developed finds the fourth class by splitting the group of three regions of the K-rich phase into two groups, reflecting the fact that their composition could be slightly different (essentially, they differ by their content in Sr). We did not try to push the interpretation further, our purpose being just to point out the fact that one must be careful when interpreting results of automatic classification, because different algorithms may provide different results.

Conclusion

In this paper, we have described techniques for analysis of multi-dimensional maps which are being recorded more and more frequently by microanalysts. Two types of technique were investigated. The first one performs dimensionality reduction. When more than two or three images are recorded, it becomes difficult for the human visual system to infer even qualitative information. There is thus a need to project data onto a parameter space of two or three dimensions. This is then comparable to the scatterplots commonly used when two (or three) images are recorded. We have described several techniques able to perform this dimensionality reduction. Some of these techniques, like Multivariate Statistical Analysis, are linear. But we have put emphasis on nonlinear techniques, which are more general. We have described some algorithms for nonlinear mapping (heuristic mapping, Sammon's mapping). These techniques have to be investigated more deeply, for different

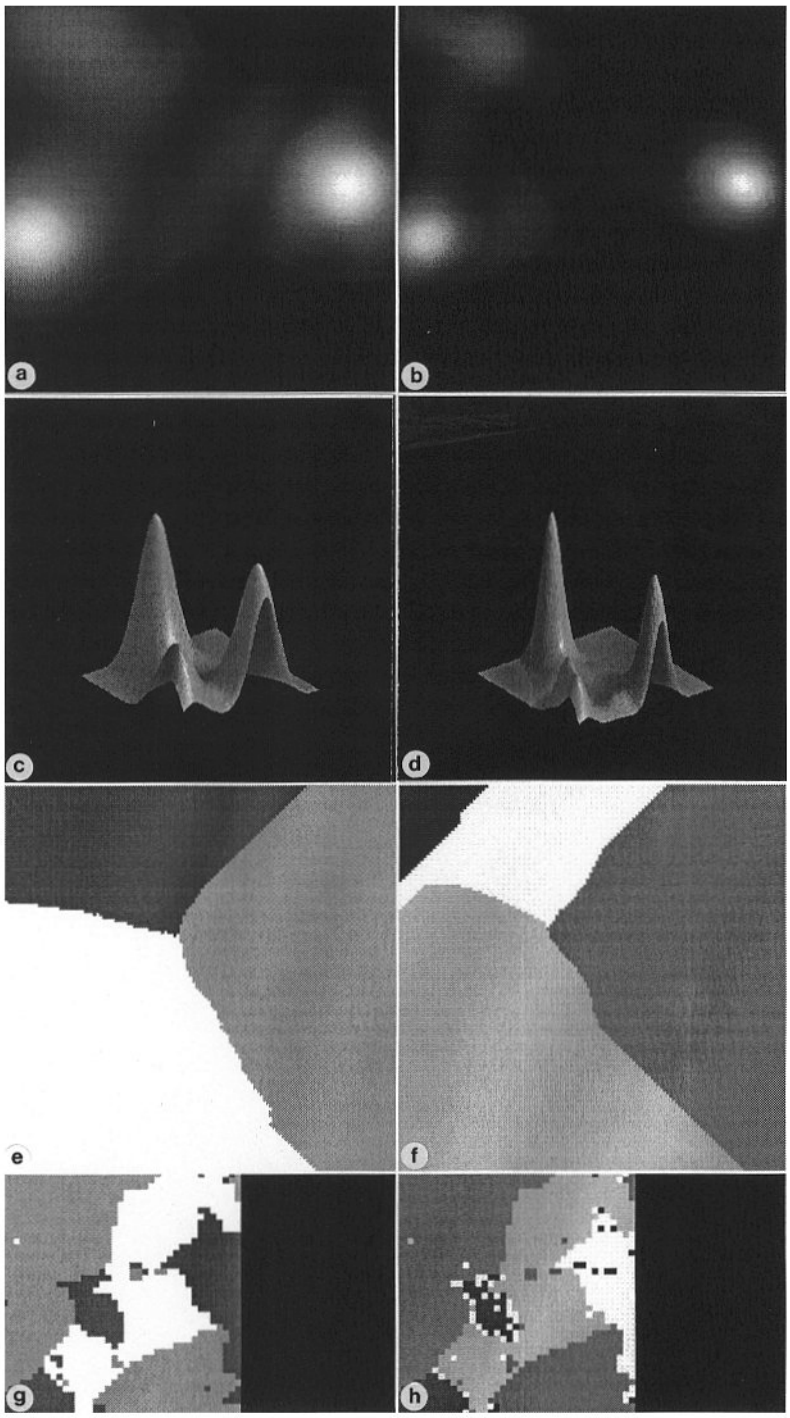

Figure 11. Results of segmentation obtained with the new clustering method. (**a**, **b**) Probability density function estimated with sd = 25 quantization levels and sd = 15, respectively. (**c**, **d**) Another way to represent the probability density functions. (**e**, **f**) Segmented parameter space (for the two values of sd given above). The number of classes found are three and four, respectively. In the latter case, the top-left peak in the pdf is split into two smaller peaks. (**g**, **h**) Segmented image space (for the two values of sd). In the latter case (fig. 11h), we can observe that the three particles represented by the darkest gray levels in figure 11g are now considered as belonging to two different classes, which means that their composition is slightly different, which can be confirmed by looking carefully at the mean values of the signals recorded for each of them.

microanalytical techniques, before a conclusion can be drawn concerning their relative efficiency. The consequences of the distortion introduced by the nonlinear dimensionality reduction have also to be evaluated.

The second group of techniques we have investigated is related to the problem of automatic segmentation of multivariate images. We have described several methods for unsupervised automatic classification or clustering of objects. We have shown that classical techniques (K-means, fuzzy C-means) can be relatively fast, but work only when restricted assumptions are fulfilled. When this is not the case, other techniques have to be used. We have described one of them, which is based on the estimation of the pdf followed by the partition of the parameter space using techniques inherited from mathematical morphology. This technique is rather efficient when the parameter space is

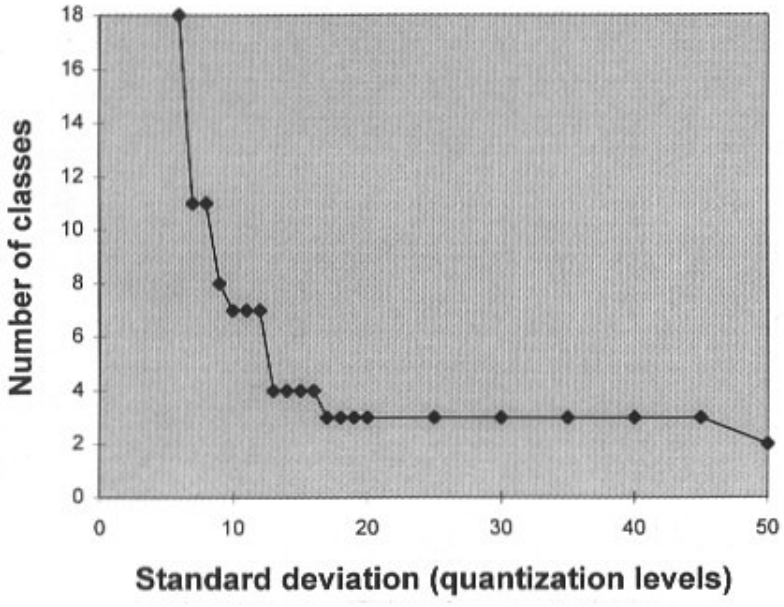

Figure 12. Curve of the number of modes of the probability density function as a function of the standard deviation (sd) of the Gaussian Parzen window. One large plateau is observed for three classes, but a small plateau is also observed for four classes.

two-dimensional. Otherwise, a reduction of dimensionality must be performed first. We are also investigating whether other techniques, including neuromimetic ones (networks based on the Adaptive Resonance Theory (ART) (Carpenter and Grossberg, 1987; Wienke *et al.*, 1994), for instance, can be used for the purpose of clustering). Again, more experiments have to be conducted before conclusions can be drawn concerning the relative figures of merit of the different clustering techniques.

With the development of such techniques, we hope we are able to predict that multivariate image analysis will attain a mature state in the very near future.

Appendix 1: Criteria Used for Evaluating the Quality of a Partition (Hard Clustering)

Due to the Huyghens theorem, the total variance σ_t^2 of a data set is the sum of the within-class variance σ_w^2 and of the between-class variance σ_b^2 (Duda and Hart, 1973). Most of the criteria are based on some combinations of these last two variances. Of course, the within-class variance σ_w^2 decreases as the number of classes increases (as a limit, σ_w^2 decreases to zero when the number of classes approaches the number of pixels) and cannot be used as a valid criterion. Different possibilities we have tested are:

- the within-class variance times the number of classes: a minimum of this criterion is often a good indicator,
- the ratio of the maximum within-class variance to the minimum between-class variance:

$$\lambda = \frac{\left(\sigma_w^2\right)_{\max}}{\left(\sigma_b^2\right)_{\min}} \qquad (14)$$

- the Bow estimator (Bow, 1992), expressed as:

$$B = \frac{1}{K}\sum_{k=1}^{K} \max_{k'} \left[\frac{\left(\sigma_w^2\right)_k + \left(\sigma_w^2\right)_{k'}}{d_{k,k'}} \right] \qquad (15)$$

where K is the number of classes, $(\sigma_w^2)_k$ is the within-class variance of class k and $d_{k,k}'$ is the distance between classes k and k′.

Minima of these criteria are expected for the true number of classes of the data set.

New criteria have been suggested recently in the literature (Xu *et al.*, 1993). We have not studied these criteria yet.

Appendix 2: Criteria Used for Evaluating the Quality of a Partition (Fuzzy Clustering)

In addition to the previous criteria, several other ones can be used, based on the fact that a "good" partition is one for which the fuzziness is at a minimum (before defuzzyfication). This means that, on average, each pixel must be assigned to a class with as little ambiguity as possible. Many criteria have been suggested for evaluating the amount of fuzziness of a partition (Windham, 1982). Some of these criteria are:

- the non-fuzziness index (Roubens, 1982):

$$NFI = \frac{\left[C \sum_{c=1}^{C} \sum_{i} \mu_{ic}^2 \right] - N}{N\left(C-1\right)} \qquad (16)$$

NFI goes toward zero for a uniform (i.e., completely fuzzy) partition and toward 1 for a hard (i.e., unambiguous) partition.

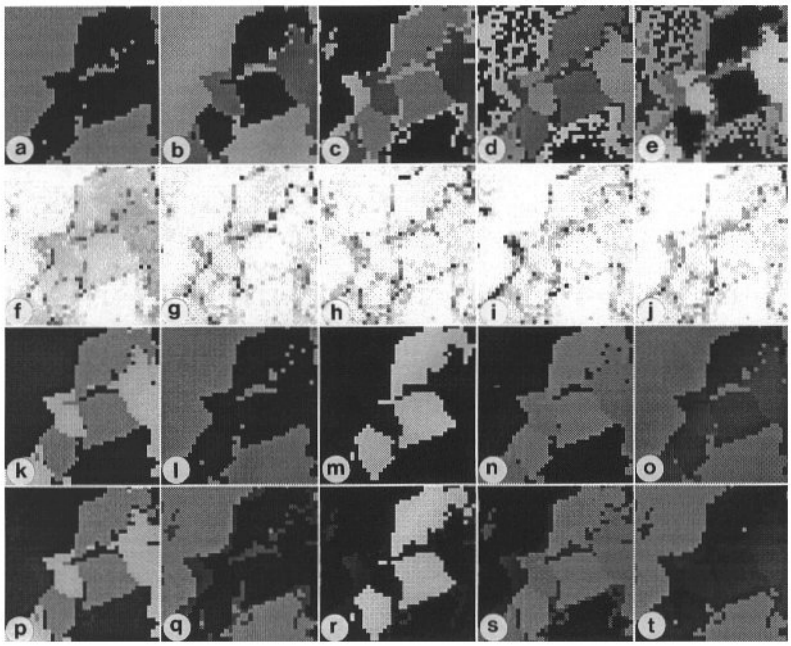

Figure 13. Results of classification using the K-means procedure. a-e) Results of classification for K=2, 3, 4, 5 and 6 classes, respectively. (**f-j**) Display of the local (on a pixel basis) correlation coefficient between the experimental vector (five components) and the vector obtained after classification and averaging, for the five classification results displayed in fig. 13a-e. (**k-o**) The five original images after averaging according to the classification into three classes. (**p-t**) The five original images after averaging according to the classification into four classes.

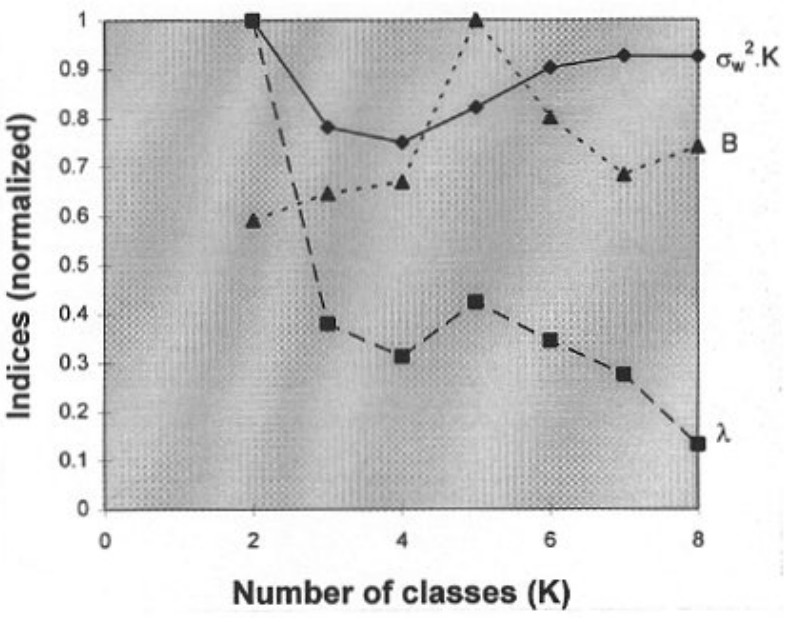

Figure 14. Some indices of the quality of the partition (see Appendix 1 for their meanings) plotted as a function of the number of classes K (classification was made according to the K-means procedure). Two of the criteria indicate an optimum number of classes equal to 4, but the Bow criterion does not show a clear minimum.

- the minimum hard tendency (HTmin) and mean hard tendency (HTmean) criteria defined by Carazo *et al.* (1990):

$$HTmin = \max_c \left(-\log_{10} T_c\right) \quad (17)$$

$$HTmean = \frac{1}{C}\sum_{c=1}^{C}\left(-\log_{10} T_c\right) \quad (18)$$

where $T_c = [\,\Sigma_i\, r_i\,]/\,\mathrm{card}\,(C_c)$, $r_i = \mu_{c2,i}/\mu_{c1,I}$, $\mu_{c1,i}$ is the highest membership degree of pixel i, $\mu_{c2,i}$ is the second membership degree and C_c is class c.

Maxima of NFI, HTmin and HTmean are expected for the true number of classes.

Acknowledgements

We are pleased to thank Dr. K. Janssens and his collaborators at the University of Antwerp for allowing us to use their experimental data set for illustrating this paper. We also thank S. Prevost for the programming of the watershed function. Thanks are also due to the four reviewers (D.S. Bright, E.S. Gelsema, P.G. Kenny, and N.K.

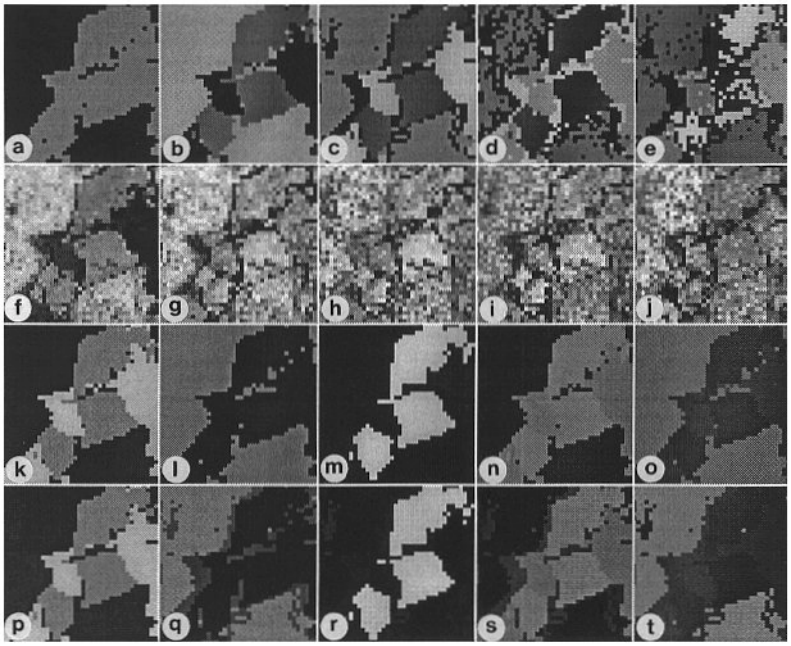

Figure 15. Results of classification using the fuzzy C-means procedure. (**a**-**e**) Results of classification for K=2, 3, 4, 5 and 6 classes, respectively. (**f**-**j**) Display of the local (on a pixel basis) entropy of the membership degrees, for the five classification results displayed in Figure 15a-e. (**k**-**o**) The five original images after averaging according to the classification into three classes. (**p**-**t**) The five original images after averaging according to the classification into four classes. These results are very similar to those obtained with the K-means technique (see Figure 13).

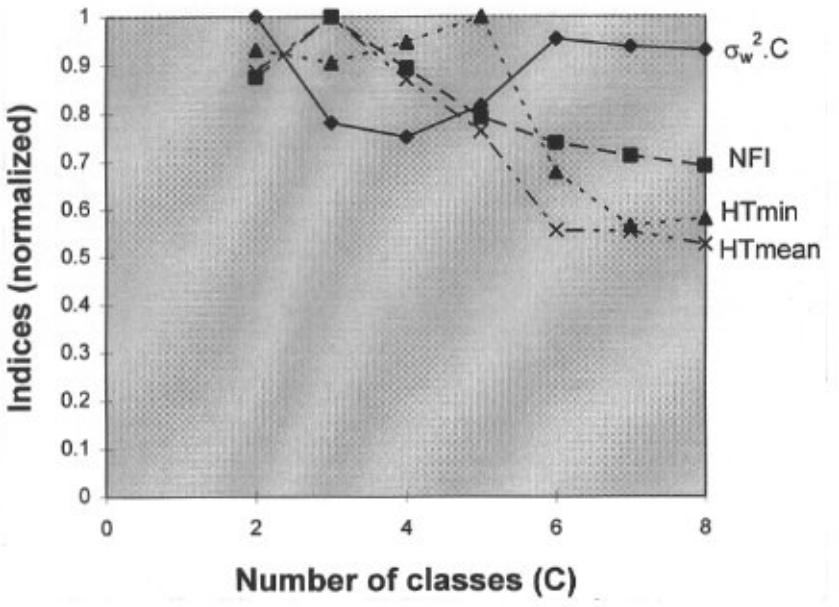

Figure 16. Some indices of the quality of the partition (see Appendix 2 for their meanings) plotted as a function of the number of classes C (classification was made according to the fuzzy C-means procedure). The weighted intra-class variance indicates an optimum number of classes equal to 4, while NFI and HTmean indicate 3 classes and HTmin 5 classes.

Tovey) and to S. Ricord for their helpful comments and their help in language improvement.

References

Beucher S (1992) The watershed transformation applied to image segmentation. Scanning Microsc **Suppl 6**: 299-314.

Beucher S, Meyer F (1992) The morphological approach to segmentation: the watershed transformation. In: Mathematical Morphology in Image Processing. Dougherty ER (ed.). Marcel Dekker, New York, pp 433-481.

Bonnet N (1995a) Processing of images and image series: a tutorial review for chemical microanalysis. Mikrochimica Acta **120**: 195-210.

Bonnet N (1995b) Preliminary investigation of two methods for the automatic handling of multivariate maps in microanalysis. Ultramicroscopy **57**: 17-27.

Bonnet N, Trebbia P (1992) Multi-dimensional data analysis and processing in electron-induced microanalysis. Scanning Microsc **Suppl 6**: 163-177.

Bonnet N, Simova E, Lebonvallet S, Kaplan H (1992) New applications of multivariate statistical analysis in spectroscopy and microscopy. Ultramicroscopy **40**: 1-11.

Bonnet N, Herbin M, Vautrot P (1995) Extension of the scatterplot technique to multiple images. Ultramicroscopy **60**: 349-355.

Bow S-T (1992) Pattern recognition and image preprocessing. Marcel Dekker, New York. pp. 87-141.

Bright DS, Newbury DE (1991) Concentration histogram imaging. Anal Chem **63**: 243A-250A.

Bright DS, Newbury DE, Marinenko RB (1988) Concentration-concentration histograms: scatter diagrams applied to quantitative compositional maps. Microbeam Analysis. Newbury DE (ed.). San Francisco Press, San Francisco, pp. 18-24.

Browning R, Smialek JL, Jacobson NS (1987) Multielement mapping of a-SiC by Scanning Auger Microscopy. Adv Ceramic Materials **2**: 773-779.

Carazo JM, Rivera FF, Zapata EL, Radermacher M, Frank J (1990) Fuzzy sets-based classification of electron microscopy images of biological macromolecules with an application to ribosomal particles. J Microsc **157**: 187-203.

Carpenter GA, Grossberg S (1987) A massively parallel architecture for a self-organizing neural pattern recognition machine. Comp Vision Graphics Image Proc **37**: 54-115.

Demartines P (1994) Analyse des Données par Reseaux de Neurones Auto-Organisés (Analysis of Data by Selforganizing Neural Networks). Doctoral Thesis, Institut National Polytechnique, Grenoble, France.

Duda RO, Hart PE (1973) Pattern classification and scene analysis. Wiley Interscience, New York. pp. 189-259.

El Gomati M, Peacock D, Prutton M, Walker C (1987) Scatter diagrams in energy analysed digital imaging: application to scanning Auger microscopy. J Microsc **147**: 149-158.

Frank J, Van Heel M (1982) Correspondence analysis of aligned images of biological particles. J Mol Biol **161**: 134-137.

Fukunaga K (1990) Introduction to Statistical Pattern Recognition. Academic Press, Boston. pp. 508-562.

Geladi P, Esbensen K (1989) Can image analysis provide information useful in chemistry? J Chemom **3**: 419-429.

Geladi P (1995) Sampling and local models for Multivariate Image Analysis. Mikrochimica Acta **120**: 211-230.

Herbin M, Bonnet N, Vautrot P (1996) A clustering method based on the estimation of the probability density function and on the skeleton by influence zones. Application to image processing. Patt Rec Lett **17**: 1141-1150.

Jeanguillaume C (1985) Multi-parameter statistical analysis of STEM micrographs. J Microsc Spectrosc Electron **10**: 409-415.

Kenny PG, Prutton M, Roberts RH, Barkshire IR, Greenwood JC, Hadley MJ, Tear SP (1992) The application of multispectral techniques to analytical electron microscopy. Scanning Microsc **Suppl 6**: 361-367.

Kenny PG, Barkshire IR, Prutton M (1994) Three-dimensional scatter diagrams: application to surface analytical microscopy. Ultramicroscopy **56**: 289-301.

Kohonen T (1984) Self-organization and Associative Memory. Springer, Berlin. pp. 119-157.

Kruskal JB (1964) Multidimensional scaling by optimising goodness of fit to a non metric hypothesis. Psychometrica **29**: 1-27.

Paque JM, Browning R, King PL, Pianetta P (1990) Quantitative information from images of geological materials. In: Microbeam Analysis. Michael JR, Ingram P (eds). San Francisco Press, San Francisco, 195-198.

Prutton M, El Gomati MM, Kenny PG (1990) Scatter diagrams and Hotelling transforms: application to surface analytical microscopy. J Electron Spectrosc Related Phenom **52**: 197-219.

Quintana C, Bonnet N (1994a) Multivariate statistical analysis (MSA) applied to X-ray spectra and X-ray mapping of liver cell nuclei. Scanning Microsc **8**: 563-586.

Quintana C, Bonnet N (1994b) Improvements of biological X-ray microanalysis: cryoembedding for specimen preparation and multivariate statistical analysis for data interpretation. Scanning Microsc **Suppl 8**: 83-99.

Radermacher M, Frank J (1985) Use of nonlinear mapping in multivariate image analysis of molecule projections. Ultramicroscopy **17**: 117-126.

Roubens M (1982) Fuzzy clustering algorithms and their cluster validity. Eur J Opt Res **10**: 294-301.

Sammon JW (1969) A nonlinear mapping for data structure analysis. IEEE Trans Comput **C18**: 401-409.

Serra J (1982) Image Analysis and Mathematical Morphology. Academic Press, New York. pp 424-478.

Shepard RN (1962) The analysis of proximities: multidimensional scaling with an unknown distance function. Psychometrika **27**, 125-139.

Swietlicki E, Larsson NP, Yang C (1993) Multivariate statistical processing of elemental maps for nuclear microprobes. Nucl Instrum Methods Phys Res **B77**: 195-202.

Trebbia P, Bonnet N (1990) EELS elemental mapping with unconventional methods. I. Theoretical basis: image analysis with multivariate and entropy concepts. Ultramicroscopy **34**: 165-178.

Trebbia P, Wulveryck JM, Bonnet N (1995) Progress in quantitative elemental mapping by X-ray imaging. Microbeam Analysis **4**: 85-102.

Van Espen P, Janssens G, Vanhoolst W, Geladi P (1992) Imaging and image processing in analytical chemistry. Analusis **20**: 81-90.

Van Heel M, Frank J (1981) Use of multivariate statistics in analysing the images of biological macro-

molecules. Ultramicroscopy **6**: 187-194.

Vekemans B, Janssens K, Vincze L, Aerts A, Adams F, Hertogen J (1997) Automated segmentation of μ-XRF image sets. X-ray Spectrosc **26**: 333-346.

Vincent L (1990) Algorithmes morphologiques à base de files d'attentes et de lacets. Extension aux graphes. (Morphological algorithms based on queues and loops, with extension to graphs). Doctoral Thesis. Ecole des Mines, Paris.

Wienke D, Xie Y, Hopke PK (1994) An adaptive resonance theory based artificial neural network (ART-2a) for rapid identification of airborne particle shapes from their scanning electron microscopy images. Chem Intell Lab Syst **25**: 367-387.

Windham MP (1982) Cluster validity for the fuzzy C-means clustering algorithm. IEEE Trans Patt Anal Machine Intell **4**: 357-363.

Xu S, Kamath M, Capson D (1993) Selection of partitions from a hierarchy. Patt Rec Lett **14**, 7-15.

Discussion with Reviewers

N.K. Tovey: One of the problems in supervised classification is determining the number of relevant classes. The same applies in supervised classification as this presupposes that the interactive user does truly know the correct number of classes to start with. Frequently in an x-ray microanalysis set of image, there are a few classes which are well defined and relatively extensive in area, while others are less obvious.

Would it be possible to proceed as follows? Identify perhaps the three most dominant classes in parameter space, classify the images with just these three, but only accepting the class if the confidence is above a given level, say 90 or 95%. Now copies of the source images are modified so that where classification has taken place, all pixels are set to zero in all images, and these are then used in a second pass of the analysis except that the existing classified areas will appear as one dominant class which is removed before extraction of the next three most dominant classes. Repeat the procedure until say 99+% of the image has been classified.

Authors: We agree that the method you suggest has to be tried. However, we do not think it would solve the problem of finding the "true" number of classes. The reason is that, if you only accept the classification (during the first run) when a high degree of confidence is reached, you still have the tail of the distribution present during the second run, and this tail will indicate a new class, which will add, erroneously, to the ones found during the first run.

We think we had better rely on criteria working on the whole data set, as the examples of criteria described in the paper. It is true that some classes with small populations can be hidden by classes with large population. In this case, it could be better to work with the logarithm of the probability density function.

N.K. Tovey: One interesting point in the paper is the difference between the K-means and the new technique when defining four classes on the real sample. Boundaries between regions will always be regions of transition, and the new method would seem to have advantages. However, have the authors considered using an edge detection routine to define the outlines of the features in the X-ray maps in Figures 13a-f, and exclude those regions from classification? This would tend to remove the confusion between the methods. Any over-segmentation may be removed by the method described by Tovey *et al*. (1992).

Authors: As stated below, we are not convinced that the superiority of the pdf-based method concerns its ability to process better the regions of transition. But, if one wants to avoid classes associated with these regions, we agree with you on the fact that these can be excluded from the main classification step, and then the corresponding pixels can be aggregated to the most similar class. We agree also with the fact that the method described in your paper can be used to reduce some kind of over-segmentation, which concerns the number of regions created in the image space. Another kind of over-segmentation concerns the number of different classes, in the parameter space.

N.K. Tovey: Some of the examples shown are relatively small images. There is an implication in the text that some methods could be very time consuming. Could the authors give an estimate of the time required to classify say 8 images each 512x512 with perhaps 8 classes in each?

Authors: The computations whose times are indicated below were performed with a Sparc10-SUN workstation. The data set corresponded to 8 images each 512x512 and 8 classes (it was assumed that the number of classes was known). The K-means clustering procedure took 5 minutes. The fuzzy C-means procedure took 11 minutes

Linear mapping (PCA or CA) took 1 minute. With the implementation described in this paper, the Sammon's mapping would take several hours. Therefore, we limit the application of this procedure to 64x64 pixels. We are currently investigating another implementation which, hopefully, will reduce the computation time by one or two orders of magnitude.

When the dimensionality of the problem can be reduced to two, the pdf-based clustering technique (Parzen + watersheds) takes 15 seconds (with 256 quantization levels). When the dimensionality of the problem can only be reduced to three, the pdf-based clustering technique takes 3 minutes and 30 seconds (with 64 quantization levels).

These computation times are obtained when the number of classes (8 in the exemple) is known. When the

number of classes is unknown, the computation time increases significantly, because the clustering procedure must be applied several times and the number of classes must be chosen according to some criterion.

P.G. Kenny: Figures 9 and 12 show the numbers of classes as a function of Parzen window standard deviation for the simulated and real images. You state the correct number of classes can be deduced by looking for plateaus in these graphs. However, the correct choice of plateau for the simulated data is much less obvious than for the real data, even though it comprises exactly four well-defined phases. Can you comment on why the graph for such a simple case is so difficult to interpret?
Authors: The number of classes in a data set is not a perfectly well defined concept. The classification may be described by a hierarchical tree: the whole data set can be split into a few large classes (in hierarchical classification, a binary tree is generally built). Then, large classes can be split into subclasses, which can themselves be split again. This hierarchical structure of the data set is also apparent in the framework of the pdf-based method that we suggest in this paper: when the standard deviation of the Parzen window is large, only the main classes are detected. When the standard deviation is decreased, a larger number of classes (in fact, subclasses of the previous ones) are detected. The different plateaus in the curves indicate several possible levels for classification. It is then the responsibility of the user to choose among these different possibilities, according to the purpose of the classification procedure.

In the simulated case, a classification can be performed into two classes (it can be remarked that the result of this classification corresponds approximately to the first factorial image obtained by MSA, see fig. 2i). Classification can also be made into four classes, which can be explained easily on the basis of the first two factorial images (fig. 2i and 2l). A refined classification (suggested by another plateau) can also be performed into five classes: one more class (corresponding to the boundaries between the previous regions) is then added. Thus, the reason why the graph is difficult to interpret is simply an indication that the classification problem is not so trivial, although it corresponds to a simulation. Of course, a simpler simulation (without noise nor boundary transitions) would led to a much simpler curve and interpretation.

In the real example, the situation is somewhat simpler, with three main classes (a very stable situation indicated by a large plateau) and the possibility to split one of them into two subclasses.

P.G. Kenny: The watershed segmentation technique applied to the real image data provides a more satisfactory partitioning for the case of four classes (fig. 11f) than is revealed by the K-means or fuzzy C-means segmentation methods (figs. 13c and 15c). Could the difference between the results be due entirely to the different versions of the scatterplot/pdf used in each case? (The watershed technique was applied to the smooth estimated pdf in fig. 11b, whereas the other methods were presumably applied to the sparse scatterplots in figs. 10k or 10m.) Have you attempted to apply the K-means or fuzzy C-means methods to a smooth estimated pdf?
Authors: We would not conclude that the segmentation result provided by the watershed technique is "better" than the other results, because there is no ground truth on the subject. Even the observation of the experimental images must be cautioned, because it can lead to very subjective conclusions.

From a technical point of view, we do not think the difference between the results can be attributed to smoothing (The K-means and fuzzy C-means techniques do not make use of the pdf explicitly. Thus we do not think that smoothing in the parameter space can be envisaged in these cases). Rather, our point of view in this paper is that the classical techniques assume hyperspherical or hyperelliptical pdf while the pdf-based technique does not (see Herbin et al. for a more detailed discussion of this problem). Although the pdf for the different classes of this example do not seem to be very far from isotropic distributions, the deviation may be sufficient for explaining the differences. But it is difficult to be sure because we are working in a five-dimensional space, where it is difficult to visualize pdf and their detected boundaries.

P.G. Kenny: The new techniques reported in this paper appear to offer a robust solution for the partitioning of data that can safely be reduced to two dimensions. Would it be feasible to extend the techniques to cope with data that have a higher intrinsic dimensionality? If so, can you comment on the computational complexity (i.e., the relationship between intrinsic dimensionality, execution time and memory requirements). Also, for the case of a higher intrinsic dimensionality (say M=5), would it be better to apply your sophisticated techniques to a reduced form of the data (M=2) or to apply a simple K-means technique to the full form (M=5)?
Authors: Until now, we have limited our implementation of the techniques described in this paper (Parzen window pdf estimation, watershed-based boundary detection) to two- or three-dimensional spaces. This is not an ultimate limit: we could probably extend these techniques to work in a four- or five- dimensional space, provided the number of quantization levels Q is reduced in parallel. (The complexity of the methods is of the order of Q^M, in terms of memory requirements and execution time).

The choice between applying a simple K-means

technique to the full form (say M=5) or applying a pdf-based technique to a reduced form (say M=2 or 3) is not mainly governed by computation considerations. Returning to your previous question, one must be aware that applying a technique which assumes hyperelliptical clusters in a situation where they are not present will produce highly erroneous results. The error has to be estimated and compared to the error caused by the distortion resulting from the mapping (M=5 to M=2 or 3). It should also be added that other techniques than the one described in this paper are available for performing clustering without making assumptions concerning the shape of clusters (Bonnet N, Vautrot P, "A comparative study of clustering methods which do not make assumptions on the shape of clusters", in preparation).

D.S. Bright: Could you comment a little bit on your equation (7) ?

Authors: This formula describes the way points (representing pixels) in a N-dimensional parameter space can be mapped onto a parameter space of reduced dimension ($R^N \rightarrow R^M$). At first, the coordinates of each point i are initialized as $x_k^i(0)$, k=1...M. This initialization can be random, or governed by heuristics, or be the result of a preliminary linear mapping. Then, each point is moved (in the M-dimensional parameter space) in such a way that its new position ($x^i(t)$) corresponds to a decrease of a criterion (as expressed in formula (4) to (6)) as compared to its previous position ($x^i(t-1)$). Equation (7) represents a classical way (Newton-like steepest descent) to ensure the decrease of the criterion C (the first derivative governs the direction of the displacement while the second derivative governs the amount of displacement). These derivatives must be specified according to the chosen criterion, which gives several variants (Sammon's mapping, MDS) for the approach.

From a computational point of view, the drawback of the method is that the distances (D_{ij} and d_{ij}) of a point i to all the other points j must be computed a large number of times.

Additional Reference

Tovey NK, Dent DL, Corbett WM, Krinsley DH (1992) Processing multi-spectral scanning electron microscopy images for quantitative microfabric analysis. Scanning Microsc **Suppl 6**: 269-282.