JOURNAL OF CHEMOMETRICS, VOL. 11, 483–500 (1997)

COMBINING GLOBAL AND INDIVIDUAL IMAGE FEATURES TO CHARACTERIZE GRANULAR PRODUCT POPULATIONS

F. ROS,[1] S. GUILLAUME,[1]* G. RABATEL,[1] F. SEVILA[1] AND D. BERTRAND[2]

[1] *Centre National du Machinisme Agricole, du Genie Rural, des Eaux et des Forêts, Génie des Equipements, Avenue Jean-François Breton, BP 5095, F-34033 Montpellier Cedex 1, France*

[2] *Institut National de la Recherche Agronomique, Technologie Appliquée à la Nutrition, BP 527, Rue de la Géraudière, F-44026 Nantes Cedex 03, France*

SUMMARY

The characterization of granular product populations using image analysis is a difficult problem because it often requires the extraction and combination of many different features. We propose to study in a general way these problems of granular product classification, considering the image analysis phase, the processing of the information extracted and the decision making. In this paper we focus rather on the decision system development. It is based on a hierarchical approach to the problem, including a generalist system whose outputs are ambiguous (an approximative solution), connected to specialist systems trained to give non-ambiguous solutions. The inputs of the generalist system are the components of a vector containing the most important information for discriminating all the decision classes, while the inputs of the specialist systems are those which best distinguish a given class from another. This strategy enables us to overcome the multiclass aspect of the problem. It is independent of the choice of the techniques to select the pertinent information and to take the decision. This method is applied in the framework of a meal classification where three types of classifier (discriminant analysis, *k* nearest neighbours and multilayer neural networks) are compared.

J. Chemometrics, Vol. **11**, 483–500 (1997) (No. of Figures: 12 No. of Tables: 7 No. of References: 14)

KEY WORDS image analysis; granular product; selection; classification

INTRODUCTION

From the first milling of cereals to modern activity sectors such as building or pharmacology, humanity has been confronted with the production of granular products. The two main difficulties in the production of granular products are the improvement of qualtity and increasing productivity. This double objective is usually reached by automation. However, quality control has to take into account both objective and subjective criteria.[1] In many industries this quality control is ensured only by an expert who controls all the mechanisms of production. He is guided not only by chemical and physical analysis but also by subjective characterization (vision, touch). To help this expert and to improve the quality of granular products, many researchers have proposed artificial methods[2,3] using different sensors. Image analysis seems to be the best-adapted sensor because it is fast and non-destructive and because there is no restriction on the type of products that can be analysed.

Two types of invariant can be extracted from this type of images. The first concerns extraction of invariants on isolated particles and leads to a statistical distribution of parameters such as object area which we call an HF (histogram feature). The second corresponds to measurements of the whole images and is rather related to texture or shape, e.g. constant grey-level run lengths. These features are

Correspondence to: S. Guillaume, Centre National du Machinisme Agricole, du Genie Rural, des Eaux et des Forêts, Génie des Equipements, Avenue Jean-François Breton, BP 5095, F-34033 Montpellier Cedex 1, France

CCC 0886–9383/97/060483–18 $17.50

Received 4 March 1994
Accepted 20 October 1995

different in their nature (shape, texture, etc.) and in their dimension. Most researchers[4] addressing these subjects limit themselves to a single type of extracted feature to solve their classification problem. They argue that it is difficult to combine information from different sources and that using all the information slows down the computing time. However, in the framework of a classification problem as complex as a qualitative evaluation the information contained in one feature only may not be sufficient. The paper by Ros *et al.*[5] shows that even with a rigorous data preprocessing (histogram intervals generated automatically, size of the samples estimated, intra-parameter treatment) it is necessary to consider both types of features, because individual information is not complete enough to solve the problem. The difficulty is especially related to the way the parameters have to be combined knowing that they are non-homogenous, non-equally pertinent and redundant.

Another point to be considered is related to the multiplicity of the classes. If a decision is based on two classes, a classifier system only is necessary. Generally, granular image classification is based on more than two classes, say n classes (n=3, 5, 7 classes are typical situations). In this case the above-mentioned selection method may only lead to an ambiguous, non-accurate final decision. This is because the decision criterion is global and may not be satisfactory for all the decision classes, except if there is no overlapping between them (which is not frequent).

We propose a general method to overcome this type of pattern recognition problem. The phases, image analysis (evaluation of the invariant to be extracted) and data analysis (search of the sample size to be considered, automatic generation of histogram intervals, intra-parameter treatment) have been studied in detail.[5] We develop here the strategy chosen to combine the pertinent parameters in order to improve the results from a classification obtained when they are considered individually.

This work has been realised in the context of a real classification problem where the objective was to characterize the different populations of granular product obtained after a set of transformations corresponding to a particular spacing of the grinding surfaces.

OVERVIEW OF IMAGE ANALYSIS AND DATA ANALYSIS PHASES

These two phases are fundamental to pattern recognition because they prepare the development of the decision system. As the decision cannot be infringed, we propose to consider most of the possible pertinent parameters in order to increase the possibility of finding the relationships between each feature and the decision space. This approach has the merit to be general but makes the use of a specific methodology to combine all these image features necessary.

All the individual features and some global features are represented by a statistical distribution of measurements. The description of these features can be found in several references.[5–7] The HF representation has been considered because its information is more complete than that coming out of some statistical parameters such as average or deviation. However, the histograms can be constructed in many different ways by varying the class intervals. It is therefore necessary to study the class intervals which give the best results according to predefined criteria. An automated procedure devoted to the search for the most relevant classes has been developed. Details are given in Reference 5.

Since the purpose is to classify different populations of granules, the size of the samples considered to construct the training and test bases has to be evaluated. A method based on a compromise between the minimum size of the samples to distinguish the different classes and the minimum size to obtain a stable HF in each class has been used.[5] For each sample, containing approximately the same number of granules, the individual and global variables are grouped together in several vectors. The vectors obtained with all the images are then gathered in several matrices.

A factorial analysis[8] enables us to reduce the dimension on each matrix and synthesize the information of the feature vector by eliminating some possible redundancy in the components. It is useful to apply PCA (principal component analysis) on an NHF (non-histogram feature), while it is

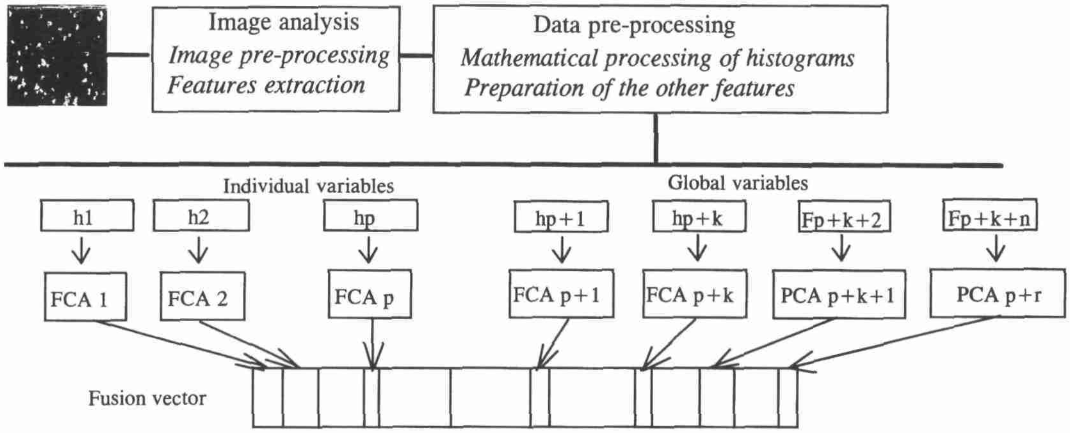


Figure 1. Schematic diagram of image and data analysis phases (h_i, histogram number i; p, number of individual variables; r, number of global variables; FCA, factorial correspondence analysis; PCA, principal component analysis)

more advantageous to use FCA (factorial correspondence analysis) on an HF. The condensed information is then kept in a fusion vector for each set of samples.

Figure 1 summarizes the procedures from the image capture phase to the production of non-correlated vectors representing each feature extracted and that can be used directly as input of a classification system.

DESCRIPTION OF CLASSIFICATION SYSTEM

The components obtained by factorial analysis on all the original variables were merged in a single vector called the 'fusion vector'. This vector was supposed to contain almost all the information extracted from the images. As the classifying system was unable to use all the elements of the fusion vector as input data and unable to distinguish all the classes, a specific method had to be employed.

The methodology adopted involves two steps. Firstly, a system designated the 'generalist system' has to guide the final decision indicating to which subset of possible classes the pattern tested is likely to belong. Its main purpose consists of approximating the solution.

The answer can be either 'definite', if the generalist system is able to neatly classify the sample in a single class, or 'ambiguous' in the contrary case. If the answer is ambiguous, an adapted 'specialist system' is used. Its role is to definitely classify the sample into a single group, knowing the subset of classes which was previously identified by the generalist system.

The specialist systems are devoted to the separation of classes which are generally 'overlapping' according to the generalist system. A specialist system is likely to produce a less ambiguous and better decision than the generalist system as it has been specialy developed to distinguish the probable classes.

For example, suppose that the sample belongs to one qualitative class among five (a, b, c, d and e). The generalist system is supposed to be able to separate a from the others classes, but confounds b with e and c with d. In this case, two specialist systems must be built which are specifically devoted to the separation of b from e and c from d respectively. The input data of the generalist and specialist systems are generally not the same. A module of selection is used to select the relevant input variables for each system.

The size of the subset to be considered depends of course on the complexity of the problem but also

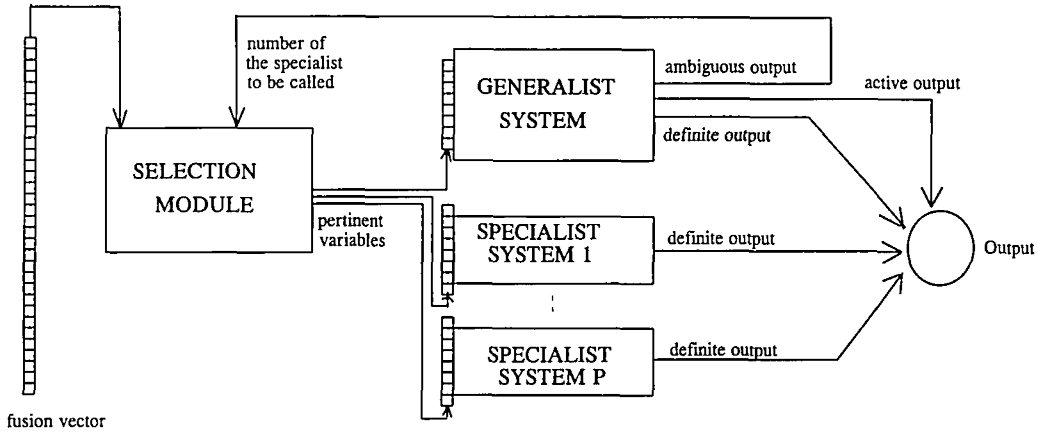


Figure 2. General schematic diagram of method

on the total number of classes. In a three- or five-class problem it is reasonable to consider an ambiguous decision between two classes, while in a ten-class problem a three-size subset will be more reasonable.

The inputs of the generalist and specialist systems are the components of a vector that yield the largest amount of information to distinguish the classes related to the different systems. The 'specialist system' enables the projection of the vector presenting the input pattern in the most appropriate subspace to segregate probable classes. The general scheme of the system is summarized in Figure 2.

The selection module is a black box which provides the most appropriate variables required according to the specific goal. In order to find the inputs of the generalist system, the selection module will select those which can separate one class from the others. The selection module is naturally linked to the fusion vector. The first output of the module gives the number of specialist systems to be called according to the ambiguous response of the generalist system. The second output gives the corresponding pertinent varibles which are the input of the specialist system. When the generalist system gives a definite output, the specialist systems are not called and the module of selection is only used to give the pertinent variables devoted to the separation of all classes.

There is no restriction either on the type of selection or on the classification procedures to be adopted, since the strategy gives free choice. However, it should be noted that the techniques employed have to be determined via the study and its objectives. Generally, some classification techniques are compared only by examining their performances in classification. Other criteria such as learning rapidity and complexity or a time test have also to be taken into account especially when a real-time decision or on-line training is required.

PRESENTATION AND DISCUSSION OF TECHNIQUES USED IN CLASSIFICATION

Information selection

For any classification problem involving a large number of measurements, it is worthwhile to find a method to reduce the dimension of the feature vector without compromising the classifer, to improve the speed of the system and eliminate possible redundancy which can disturb the training. We propose to use stepwise analysis to find pertinent variables, because although it is non-optimal, it enables us to obtain a good subset of pertinent variables rapidly. The selection of variables in the case of stepwise discriminant analysis is based on a criterion that follows a Fisher Snedecor distribution. The criterion is the Wilks lambda[8]

$$\text{lambda}(r)=(\det(W))/(\det(V)) \tag{1}$$

where r is the number of independent variables used, V is the matrix of variances and W is the matrix of within-group variances. It begins by including in the model the single best discriminating feature in terms of maximizing the mentioned criterion. This feature is paired with each of the other features and a second one is selected in the same manner. The other features are chosen similarly. As new features are included, some of the previously selected features can be removed from the model if the information they contain is available in some linear combination of the other included features. The process is stopped when the features not in the model do not improve the variance ratio significantly.

Classification methods

Three types of classifier have been studied. They have not been chosen randomly but rather as representative of three types of techniques: discriminant analysis, k nearest neighbours and multilayer neural networks.

Discriminant analysis (DA)

Discriminant analysis[9] represents the search for the boundaries which best discriminate the different decision classes. Under the assumption that the variables are normally distributed and have equal covariance matrices in each class, the technique consists of searching for the linear combination of initial variables which best separates the different classes of the problem by minimizing the intra-class variance and maximizing the inter-class variance. It enables us to determine an optimal subspace of $\mathbb{R}^n$. The data x_k projected on this new space are separated in compact clouds and isolated from each other. The main inconvenience of this method concerns the shape of the variable distributions to be distinguished.

k nearest neighbours (KNN)

The k-nearest-neighbour method[10] estimates the density of probability associated with each decision class. This method of classification consists of setting a number of neighbours and increasing the volume around the pattern to be tested until the defined region includes the given number. It should be noted that when the conditions

$$\lim_{n\to\infty} k_n=\infty \quad \text{and} \quad \lim_{n\to\infty} k_n/n=\infty \tag{2}$$

(n, total number of patterns; k_n, number of neighbours) are verified, the value of the density of probability of the pattern to be tested can be adjusted to the volume defined and this density estimator is free of bias. A usual value for k_n is $\sqrt{n}$. Variations on this method have been proposed especially to accelerate decision making, which requires the computing of all the distances with the patterns of the training base. These methods are available chiefly when the number of training patterns is very large. The main inconveniences of this technique are the storage of all n training patterns and the processing of n distances to identify a sample test. It is, however, useful as it is non-linear and independent of the population shapes.

Multilayer neural networks (MNN)

Multilayer neural networks[11] are examined because they are non-parametric and can make fewer assumptions concerning the shapes of underlying distributions than traditional statistical techniques. Numerous papers have shown the behaviour of a multilayer neural network which can be in some precise cases compared to discriminant analysis or principal component analysis. They are artificial intelligence systems which attempt to achieve good performance via dense interconnection of simple computational elements. In this respect, artificial neural networks are based on our present understanding of biological nervous systems. The behaviour of an artificial network is determined by its structure and the strength of the connections. The learning algorithm consists of some modifications of the connection strengths to improve performance. A multilayer neural network is a feedforward model with one or more layers of computing elements between the input and output layers (see Figure 3). The elements of each layer are connected with elements of the previous layer. Within a layer there are no connections between the elements. A multilayer network functions as follows. The inputs are copied in the input layer and propagated across the hidden layers to reach the output. Each neuron computes its output value by an activation function $f(I)$, with I given by

$$I=\sum_{i=1}^{N} w_{ij}*x_i+q \tag{3}$$

where N is the number of inputs, w_{ij} is a connection weight, x_i is the output of the connected element and q is a bias. The activation function is often a sigmoid function. The most widely used learning algorithm for this type of network is the back-propagation rule.[12] It is an iterative gradient descent algorithm designed to minimize the mean square error between the output signal of the network and the desired output. For detailed descriptions of this type of algorithm, see Reference 13. They have been criticized by several statisticians for their lack of theoretical advances and their empirical development. In fact, the major difficulties with neural networks come from aspects of generaliztion dependent on the topology of the network, the local minimum of the error function[12] and the time required for training. Many criticisms of neural networks stem from the fact that they have been used as a black box without taking into account the problems described above.

In conclusion about these techniques, it should be noted that the discriminant analysis approach is very useful because it is easy to develop and because the evaluation of a sample test is very fast. The

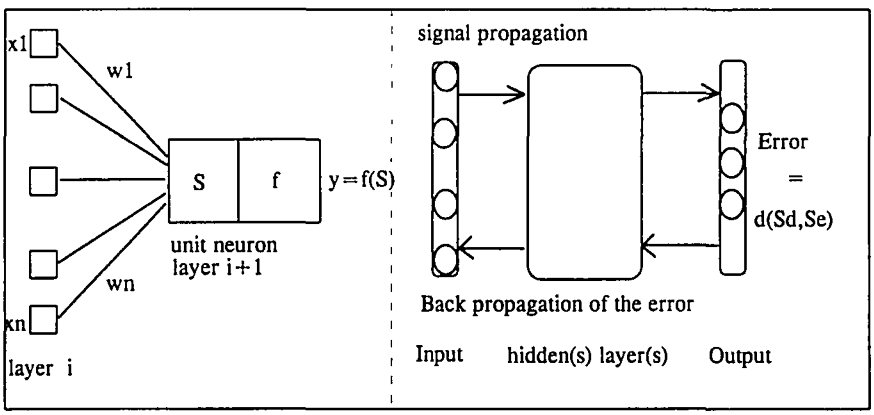


Figure 3. Schematic diagram of multilayer neural network

only problem is related to the shape of the populations to be recognized. The k-nearest-neighbour approach offers both intuitive appeal and proved effectiveness but suffers in its implementation from the often considerable computation needed to identify nearest neighbours. Several approaches to improve the efficiency of this method via a preprocessing of the prototype set have been presented. Today the loss in classification accuracy is not negligible, so the k-nearest-neighbour approach is not used. Therefore it can be satisfactory only in the case where the real-time aspect is not crucial. The MNN approach can solve complex problems but is quite difficult to develop for a non-specialist. It is a problem in some applications where it is necessary to introduce new patterns and therefore modify the training.

EXPERIMENTAL

Equipment and data

Cereal industry milling processes involve two distinct operation units, namely breaking and separation. The breaking of grains and mill products is carried out by machines fitted either with corrugated rolls (break and scratch rolls) or with smooth rolls (converters). The setting of these machines is of major importance from an economic point of view (milling yield) as well as for the commercial value of the finished products. In durum wheat milling, for instance, the most accurate control of the breaking and scratching processes is necessary, since the separation between the endosperm and the bran must yield particles with a granulometry higher than 200 μm, flour being considered as a by-product. The milling diagram includes three kinds of devices: roller mills, a plansifter, which consists of stirred sieves, and a purifier, which splits the granular product into semolinas and products to be processed again. The industrial diagrams are very complex, with many cycles of grinding and separating operations. The basic milling diagram for semolina production is shown in Figure 4. The granulometry of the product depends on the gap between the rollers of the breaker.

This control is only monitored by an expert, the miller. To adjust the variation between the rollers of the machine cylinders, he analyses the milling products. He guides his judgement by granulometry measurements and by observations which cannot be easily quantified (vision, touch). The first break roll (B1) is designed to shear the grain, while the second (B2) starts the process of separation between the endosperm and the bran. The mill was loaded with a durum wheat lot (commercial blend from the south of France, 1990 crop) which had been cleaned and tempered up to 17% for 3 h prior to milling. The grains were crushed under the conditions mentioned above. The feeding rate was the usual flow of the mill, 130 kg h^{-1}. The adjustment of the first break roll was kept constant (roll gap 0·70 mm), while the tightening of the second one was set successively at 0·30, 0·40 and 0·50 mm to define three breaking intensities leading to three classes of mill products called indifferently Ep or class i (p being the roll variations and i the index of the class). The variability of the real parameters has been limited, since the process is very close to the usual exploitation conditions. For each class, 350 images have been stored, each made up of about 120 elementary particles.

Image capture and analysis

Images have been collected with a 486 DX 33 computer connected to a CCD camera. All images were coded in 256 grey levels and had a resolution of 256×256 pixels. The different procedures have been processed with home-made software written in C++. The main procedures are the following.

(i) A mathematical operation[14] of erosion and dilatation is used to smooth the contours of the granules and to eliminate the noise of the image.

(ii) The examination of the image stops each time a set of active connected pixels (grey level comprised between two threshold values) is met, in order to store all its horizontal segments in a binary tree, where 16 individual variables related to the size, shape and luminance are extracted.

(iii) The image is examined again in order to extract the global features. The method of constant grey-level run length has been applied in three directions (0°, 45°, 90°), taking into account four grey-level classes and eleven length classes. The five parameters proposed by Galloway (grey-level uniformity, short- and long-line indicators and the ratio of the two) in each direction have been regrouped in a 15-dimensional vector. The normalized length histograms have also been considered and resulted in a 132-dimensional vector ($3\times 11\times 4$). Similar considerations have been made in order to apply the grey-level spatial interdependence. A 15-dimensional vector gathered the parameters extracted from the three-co-occurrence matrix (5×5). Moreover, all the matrix coefficients are grouped together to form a 72-dimensional fusion vector.

Var1: area object	Var2: perimeter object	Var3: convex hull area	Var4: convex hull perimeter
Var5: Var1/Var3	Var6: Var2/Var4	Var7: Var2^2/Var1	Var8: Var4^2/Var2
Var9: elongation	Var10: lighting moment 1	Var11: lighting moment 2	Var12: lighting moment 3
Var13: D1/$\sqrt{}$Var1	Var14: D0$\sqrt{}$Var1	Var15: D1/Var2	Var16: D0/Var2
Var17: Rlp	Var18: RlHF	Var19: Cop	Var20: CoHF

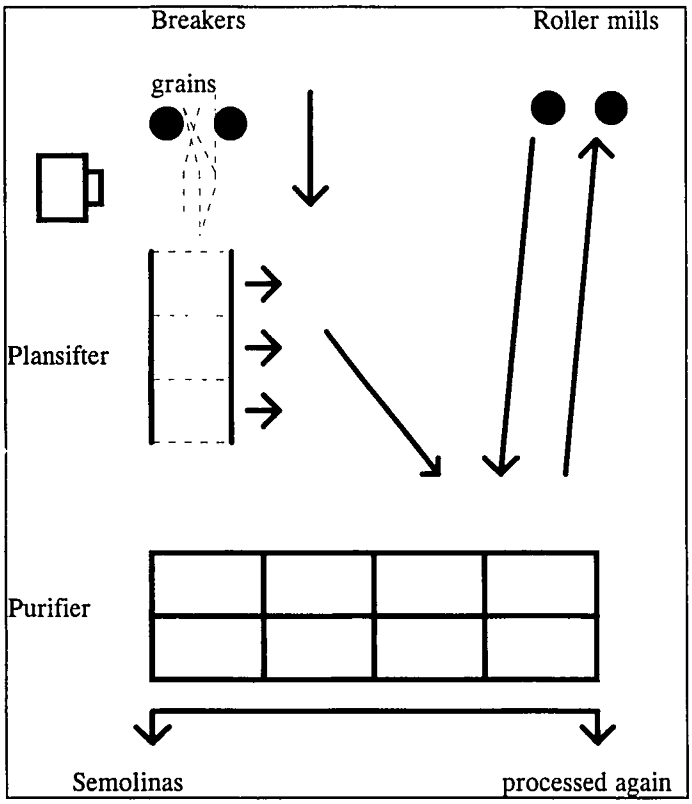
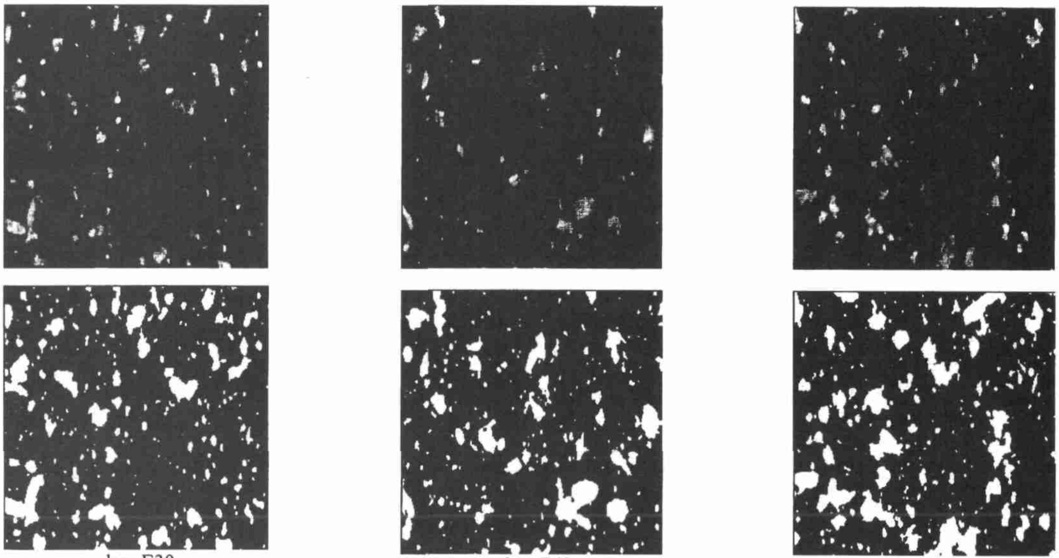


Figure 4. Basic milling diagram for semolina production



class E30 class E40 class E50

Figure 5. Examples of images of classes E30, E40 and E50 (grey-level and binary images)

D1: large-inertia axis
DO: small-inertia axis
R1p: run length parameters
R1HF: run length histogram feature
Cop: co-occurrence parameters
CoHF: co-occurrence histogram feature

Figure 5 shows images of each decision class to segregate. It should be noted that it is very difficult to judge and recognize the populations with our eyes, which points out the role of the expert in semolina. The first image in each column represents the image after capture and the second one the image after processing.

Preliminary statistical study

This part is very important because it provides information about the variables. It enables us to justify the decision to think about the set of histograms as features rather than as some standard parameters such as average or deviation. The purpose of this part is to point out the main trends of the individual features of the training base, considering a large number of particles of all the classes which can globally represent the population under study. The computing of the statistical parameters shows that the features related to the size evolve with the roll variations as expected. The larger the spacing of the grading surface, the larger is the average size of the particles. For the granules area this fact is verified:

$$A_{E30} < A_{E40} < A_{E50}, \quad \text{where } A \text{ is the average}$$

The deviation values show that the standard parameters are not sufficient to be considered as features, leading us to consider more complete information. The first two measures of the group define a region $[A-S, A+S]$ around the average for the size parameter and the others for the shape parameter (S is the deviation). It is interesting to note the significant overlapping of the three classes by observing a line which intercepts the three regions. Table 1 and Figures 6 and 7 summarize the numerical values of the average and deviation for all the individual parameters.

The correlation matrix of all the individual variables shows redundancy. The size parameters are

Table 1. Average (Ave) and standard deviation (Dev) per class

Var	Ave E30	Dev E30	Ave E40	Dev E40	Ave E50	Dev E50	Ave tot	Dev tot
1	84·7	148·84	109·16	109·16	122·94	253·96	109·5	194·43
2	27·22	27·91	32·0	36·39	33·91	44·58	37·16	32·5
3	99·27	198·41	133·9	278·611	54·92	383·23	135·2	278·97
4	29·49	24·12	33·15	29·42	34·35	33·83	29·54	33·65
5	1·02	0·21	1·009	0·21	1·02	90·23	0·20	1·01
6	0·84	0·11	0·866	0·12	0·87	0·136	0·127	0·86
7	10·95	5·14	11·77	6·25	11·9	57·39	6·58	11·83
8	14·33	1·23	14·3	1·17	14·43	1·3	1·24	14·33
9	0·47	0·21	0·49	0·209	0·49	0·21	0·48	0·48
10	0·93	0·48	0·95	0·48	1·0205	0·5	1·02	0·48
11	0·047	0·03	0·04	0·029	0·049	0·03	0·031	0·048
12	0·00385	0·004	0·0034	0·0035	0·0042	0·0049	0·004	0·00431
13	1·0356	1·038	1·182	1·248	1·234	1·55	1·36	1·22
14	0·414	0·352	0·465	0·385	0·484	0·455	0·41	0·468
15	0·286	0·195	0·307	0·22	0·313	0·249	0·23	0·316
16	0·12	0·072	0·129	0·072	0·131	0·078	0·07	0·129

correlated with a coefficient of more than 90%. This is not the case for the shape parameters whose correlations are diversified. Finally, the lighting moments of orders two and three are sufficiently correlated.

Principal component analysis (PCA) applied on training samples

Let $\lambda_0, \lambda_1, \ldots, \lambda_j, \ldots, \lambda_n$ be the eigenvalues of the covariance matrix obtained with the training data, n being the dimension of the input space. The number k of principal components retained is such that $(\lambda_0+\lambda_1+\cdots+\lambda_k)/(\lambda_0+\lambda_1+\cdots+\lambda_n)>90\%$. Using this definition, the dimension of the feature space (initially 16) can be reduced to five while conserving a large part of the total inertia (about 92%). Inertia of the principal components obtained with the matrix containing only the training samples from class E30 is shown (see Table 2). Similar responses were obtained with the other classes. The first component accounts for the size and a little of the shape, the second component accounts for the shape and a little of the size, while the third component accounts mainly for the lighting. The projection of

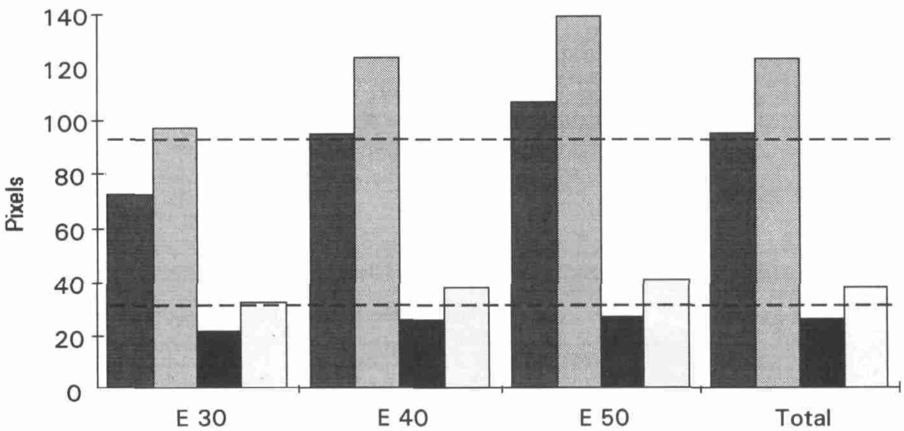


Figure 6. Statistics on area and perimeter

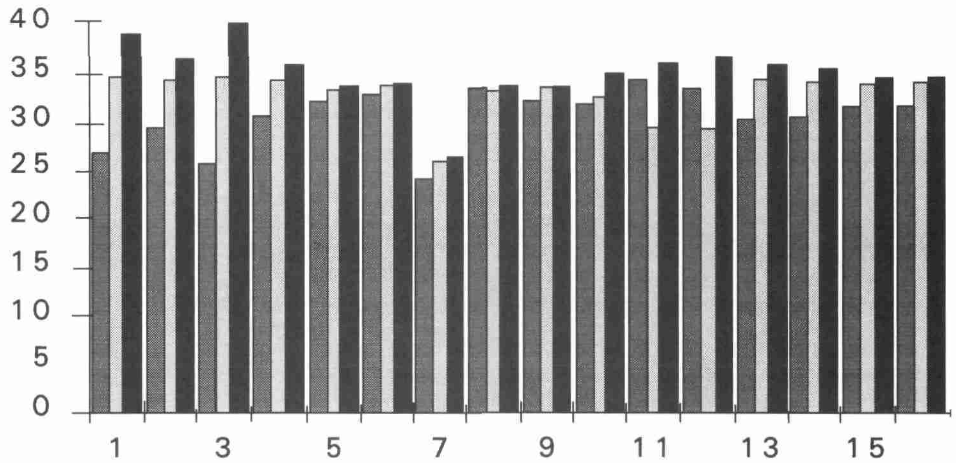


Figure 7. Average per class of individual parameters

the patterns on the factorial design (see Figure 8) points out that a given pattern cannot be assigned to a population class by considering only its position in the multidimensional space.

Statistical preprocessing

Among the 20 variables extracted, 18 were individual ones and therefore were represented as histograms. In order to determine the number of individual particles necessary to obtain stable histograms, a preliminary experiment was done on the histograms of size and shape taken as examples. Thirteen images were found to be sufficient to obtain a representative population of particles. In a second step the histograms of all the individual variables were built. The number of classes in each histogram was determined. For the 16 individual features we have begun the procedure with 100 classes for each histogram variable, decreasing this number according to predefined criteria.

Table 2. Percentage of variance explained (class E30)

Var	% Variance	Cumulated variance
1	59·30337	59·30337
2	16·22806	75·53144
3	8·54128	84·07271
4	6·26635	90·33906
5	3·21365	93·55271
6	2·24520	95·79791
7	1·99900	97·79691
8	1·28097	99·07788
9	0·57941	99·65729
10	0·15130	99·80859
11	0·11634	99·92494
12	0·03428	99·95922
13	0·02001	99·97923
14	0·00979	99·98902
15	0·00664	99·99566
16	0·00434	100·00000

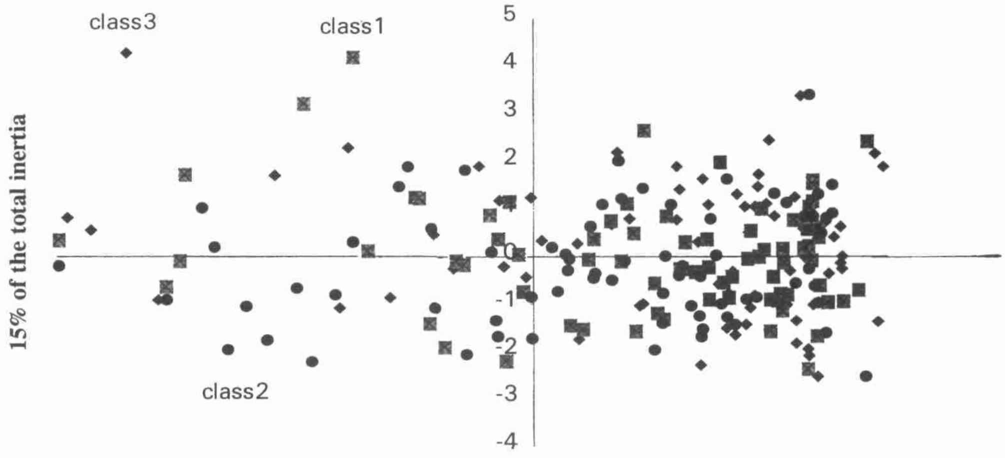


Figure 8. Projection of some examples of three classes on factorial design

Concerning the histograms obtained with global features, the intervals at the beginning are the ones given by the definition of the different classes adopted at the beginning (134 and 72). They have been reduced and optimized using the same procedure. Factorial analysis has been applied to all the extracted features. The procedure for the retention of the number of principal components consists of keeping those which contribute the most to the total inertia and provide a significant amount of data (see previous section). Table 3 shows the reduction of the feature space after these procedures.

Fusion of variables

We have proposed to build the fusion vector by merging the components obtained by factorial analysis on all the original variables. In order to emphasize the efficiency of the approach throughout this application, a more classical way to obtain the fusion vector was envisaged. Firstly, all the components coming from the original variables were merged in a single vector. The vectors obtained with all the training images were grouped together in a matrix (155 columns) reduced in size after being submitted to PCA (25 columns). The training base defined in this way is called base A, while the other base is B.

Whatever the approach, it is necessary to normalize its components in order to make them homogeneous. Let M be the training pattern centre-of-gravity vector and S the deviation vector:

$$M=(1/m)\sum_{i=1}^{m} X_{pi} \tag{4}$$

$$S_i^2=(1/m-1)\sum_{p=1}^{m} (X_{pi}-M_i)^2 \tag{5}$$

where X_{pi} is the component number i of the training pattern number p, m is the number of training examples, M is the estimation of the average vector and S_i^2 is the estimation of each component. The transformation is the following:

Table 3. Reduction of feature dimension ($n0$, base dimension; $n1$, dimension after automatic generation of HF; $n2$, dimension after factorial analysis)

Var	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$n0$	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	15	137	15	67
$n1$	9	6	5	10	7	7	7	5	6	7	6	7	6	10	7	6	15	14	15	18
$n2$	5	3	2	5	4	4	4	3	4	4	3	4	3	2	4	3	4	2	4	2

$$X_{pi}=(X_{pi}-M_i)/S_i \tag{6}$$

Each test vector is centred and reduced in the same way, using always the centre of gravity and deviations associated with the training vectors.

Selection of relevant variables

The selection of the relevant variables was guided by the minimum increase (5%) during the ascending step, the maximum decrease (3%) of the variance ratio during the descending step and the maximum number of variables to be selected in the model. The shape of the evolution function of the ratio is represented in Figure 9. The appropriate number of variables to be selected falls into the region associated with phase 2. The information kept using only the components associated with phase 1 is not sufficient to describe the complexity of the problem, while that corresponding to phase 3 retains too many details. It is likely to achieve good results for the training patterns but is too specialist for unknown examples. In fact, the components which should be kept are in the region corresponding to phase 2. In this region the results are almost as good as the ones obtained with the components of phase 3, but the model (whatever the classifier tool used) has the ability to generalize.

Four experiments have been carried out for each training base: one for the generalist system and three for the specialist systems. The number of selected variables for the different data sets is given in Table 4. It should be noted that the components selected are not automatically the same for the distinction of the different classes. This reveals the importance of the hierarchical approach chosen.

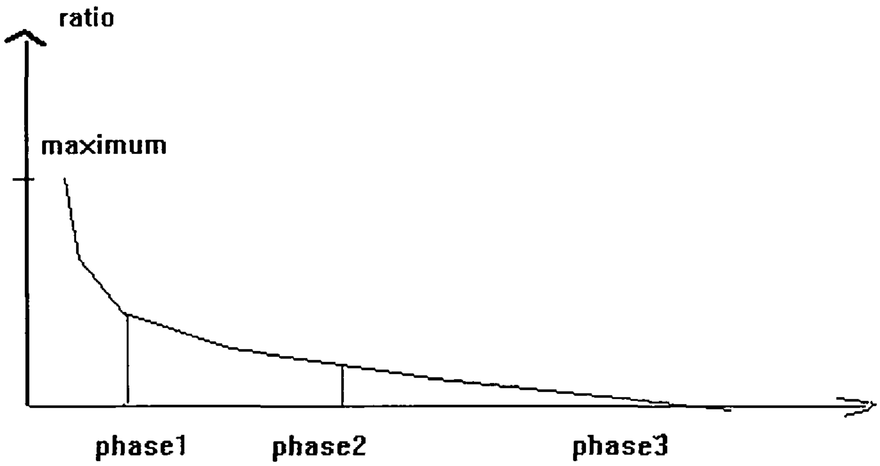


Figure 9. Shape of function ratio

Table 4. Number of selected variables for generalist (Gen) and specialist (Sp) systems

Base	Gen	Sp1	Sp2	Sp3
A	7	5	8	5
B	5	5	6	4

The generalist system has to approach the solution using the most appropriate components for this purpose, while the specialist systems need the pertinent variables related only to the classes they have to distinguish.

Design of system

Definition of ambiguous output

Whatever the classifier used, the responses of each class are transformed so that the sum is equal to one ($\sum R_i = 1$ and $i \in \{0, 2\}$). All the responses are considered ambiguous expect in the following cases:

$R_j = \max(R_i) > 0{\cdot}8, \quad \forall i \in \{0, 2\}$ and $i \neq j$
$0{\cdot}5 < R_j = \max(R_i) \leq 0{\cdot}8$ and $R_j > 2R_i, \quad \forall i \in \{0, 2\}$ and $i \neq j$
$R_j = \max(R_i) \leq 0{\cdot}5$ and $R_j > 2{\cdot}5R_i, \quad \forall i \in \{0, 2\}$ and $i \neq j$

For example, if the output of the generalist system training sample belonging to class 1 is $O = \{$score 1=0·4, score 2=0·5, score 3=0·1$\}$, the decision given by the generalist system will be designated as ambiguous and non-correct. For $O = \{$score 1=0·6, score 2=0·4, score 3=0$\}$ it will be designated as ambiguous and correct. In both cases the specialist system devoted to distinguish between classes 1 and 2 will be called. The output of this system will define the most probable class.

Parameters of classifiers

Seven ($\sim\sqrt{60}$) nearest neighbours are considered for the generalist system and four are considered for the specialist systems to identify the most probable class of samples. The Euclidean distance is used to assess the class of an example in the space of discriminant analysis. A one-hidden-layer neural network only is considered and a minimal architecture is sought.

RESULTS AND DISCUSSION

The first conclusion is that whatever the classifier used, the hierarchical approach produces better results than a single system. A significant percentage of the results obtained by the generalist system are ambiguous (both correct and non-correct). This is mainly the case between classes 2 and 3, less between classes 1 and 2 and little between classes 1 and 3. This can be explained by considering that the first class can be distinguished quite easily from the others, while the second and third classes are more difficult to separate. If we consider the system built using discriminant analysis, the results obtained using only a generalist system should yield about 60% of correct classification. By adding specialist systems, more than 80% of correct classification is obtained. All the ambiguous outputs given by the generalist system using the test samples of class 1 are transformed into good outputs using the appropriate specialist system, as are a large number of those in the other classes. Figures 10–12 show the projection of the training examples on different discriminant planes and illustrate the results obtained.

The second conclusion concerns the comparison between the classifiers. There are no significant differences between the performances obtained using multilayer neural networks, discriminant analysis and the k-nearest-neighbour method. It seems that the complexity of the problem is more thoroughly integrated by the boundaries found with the neural network. However, as the dimension of the training set is not very large, it may be thought that similar results using discriminant analysis could be obtained with a complete training set.

The third conclusion concerns the training base considered. Using factorial analysis on each

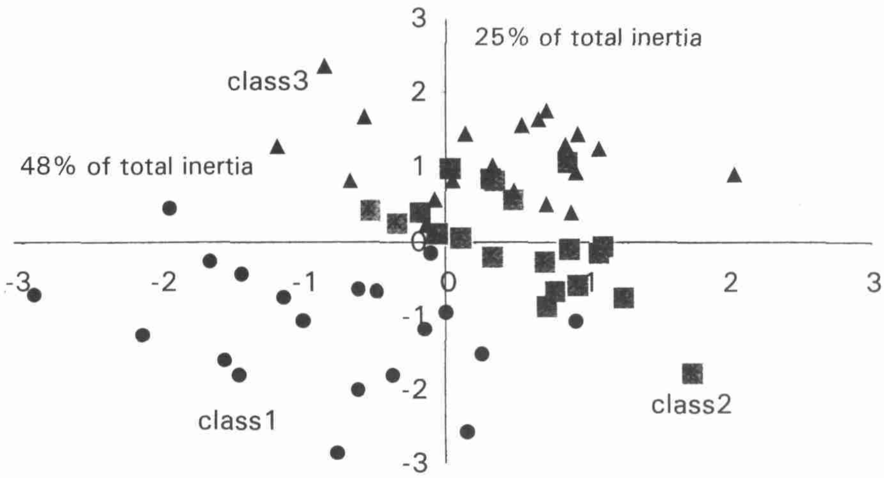


Figure 10. Projection on discriminant design 1–2 (generalist system B)

extracted feature provides virtual features whose components are non-correlated and the most significant (in terms of variability). Therefore a small subset of selected variables applied on them contains condensed information, while the same information could be obtained with a larger set of selected variables. However, the larger the input classifier dimension, the more difficult it is to design a robust classification system. On the other hand, it may be dangerous to apply PCA directly on all the variables extracted, because some pertinent variables will be transformed into virtual variables which will be less pertinent. This aspect can induce a decrease in the efficiency of the three classifiers. Tables 5–7 summarize their efficiencies (results on both the training and test sets). The differences between the results obtained with the training set and the test set using the multilayer neural network show that good results from the training set do not always imply similar results from the test set. It is often possible to obtain deceptively good training results by overparameterizing and overtraining the neural network. However, it does not always follow that the network will generalize correctly

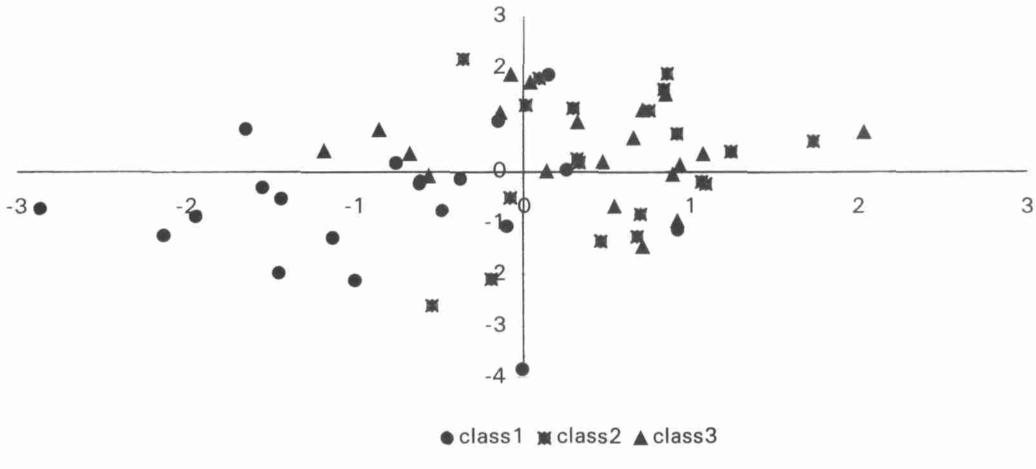


Figure 11. Projection on discriminant design 1–2 (generalist system A)

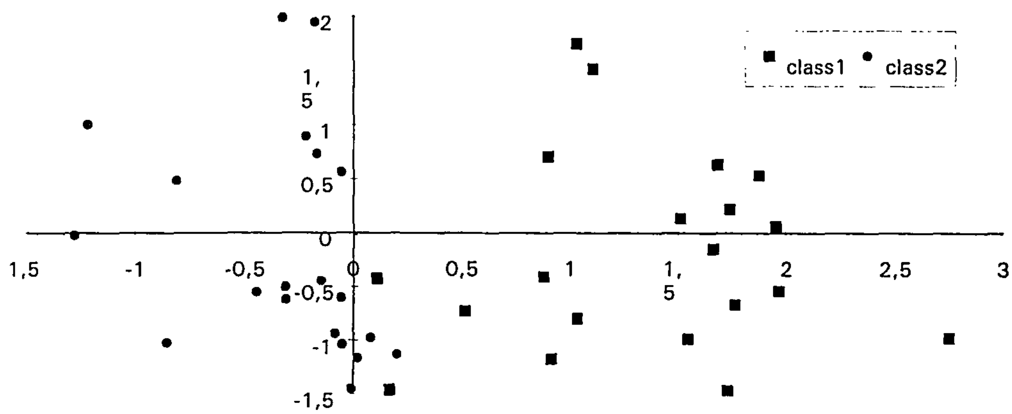


Figure 12. Projection of two classes 1 and 2 on discriminant plane 1–2 (specialist system B)

enough to obtain good results with unknown samples. Results obtained with the generalist system and with the global system (generalist with specialists) using the test samples are reported in Tables 6 and 7. The general system corresponds to the combination of the generalist and specialist systems.

Table 5. Results on training set using three classifiers (DA, KNN, MNN) for two bases (A, B)

	Correct	Correct ambiguous	Bad ambiguous	Bad
Per cent A (DA)	61·66	20	18·33	0
Per cent B (DA)	73·33	13·33	13·33	0
Per cent A (KNN)	61·66	20	18·33	0
Per cent B (KNN)	73·33	13·33	13·33	0
Per cent A (MNN)	61·7	30·07	8·3	0
Per cent B (MNN)	73·33	20	6·7	0

Table 6. Results on test set for two bases A and B obtained with only generalist system

	1A	1B	2A	2B	3A	3B
Correct (DA)	14·28	57·14	0	0	0	0
Correct ambiguous (DA)	57·14	28·57	45·45	27·27	52·94	29·41
Non-correct ambiguous (DA)	42·8	14·28	55·55	72·72	41·17	52·94
Non-correct (DA)	0	0	0	0	5·88	17·64
Correct (KNN)	21	57·14	0	0	0	0
Correct ambiguous (KNN)	50	21·42	55·55	36·36	58·88	64·7
Non-correct ambiguous (KNN)	28·5	21·57	45·45	54·54	41·17	24·23
Non-correct (KNN)	0	0	0	9·09	0	11·76
Correct (MNN)	42·8	57·1	0	0	0	0
Correct ambiguous (MNN)	42·8	28·6	55·4	36·36	52·94	41·17
Non-correct ambiguous (MNN)	14·4	14·3	45·5	63·63	41·17	41·82
Non-correct (MNN)	0	0	0	0	5·88	17·64

Table 7. Final results on test using three classifiers (generalist with specialists)

	A: c1	B: c1	A: c2	B: c2	A: c3	B: c3
Correct (DA)	100	100	72.72	81.81	70.5	70.5
Non-correct (DA)	0	0	27.27	18.18	29.4	29.41
Correct (KNN)	100	100	72.27	72.27	70.45	70.45
Non-correct (KNN)	0	0	27.27	18.18	29.4	29.41
Correct (MNN)	100	100	81.18	81.1	70.5	76.47
Non-correct (MNN)	0	0	18.18	18.18	29.4	23.53

CONCLUSIONS

The purpose of this work was to test the efficiency of our multivariable approaches in the framework of a real classification problem of granular products. It has been shown that with a rigorous data preprocessing, a classification problem where the features extracted are heterogeneous can be transformed into a classical pattern recognition problem. The hierarchical approach adopted enables us to divide the problem into more easily solved subproblems. In the framework of our application it has been concluded that the different classes could be distinguished with correct percentages using only hyperplanes as boundaries between the different classes (discriminant analysis technique). This is important in an on-line characterization problem where the computing time and the system design facility are the most important aspects. Neural networks require some human knowledge owing to the absence of a systematic method for training them, while the k-nearest-neighbours approach requires both the storage of all the training examples and a long computation time before decision making. Therefore, in future works, discriminant analysis will be the most appropriate tool to use. The problem will be extended to five quality classes and more variability in the parameters will be introduced.

REFERENCES

1. F. Ros, A. Brons, F. Sevila, G. Rabatel and C. Touzet, *Proc. IWANN'93*, pp. 726–731, Springer (1993).
2. D. Bertrand, P. Courcoux, J. C. Autran, R. Meritan and P. Robert, *J. Chemometrics*, **4**, 413–427 (1990).
3. F. Ros, S. Guillaume and F. Sevila, *proc. SPIE Conf. on Vision, Sensors, and Control for Automated Manufacturing Systems*, pp. 120–128, SPIE, Philadelphia, PA (1993).
4. N. Sinfort, 'Caractérisation des produits granulaires par analyse d'image', *Doctor Thesis*, Université de Montpellier (1993).
5. F. Ros, S. Guillaume, G. Rabatel, D. Bertrand and F. Sevila, *J. Chemometrics*, in press.
6. A. Khotanzad and R. L. Kashyap, *IEEE Trans. Syst., Man, Cyber.* **SMC-17**, 1087–1095 (1987).
7. H. H. Loh and R. C. Luo, *IEEE Trans. Ind. Electron.* **IE-35**, 323–328 (1988).
8. T. Foucart, *Analyse Factorielle sur Microordinateurs*, Masson, Paris (1982).
9. L. Lebart, A. Morineau and J. P. Fénelon, *Traitement des Données Statistiques*, Dunod, Paris (1982).
10. T. Cover and P. Hart, *IEEE Trans. Info. Theory*, **IT-13**, 21–27 (1967).
11. R. L. Lippmann, *IEEE ASSP Mag.* **4**, 4–22 (1987).
12. Y. Le Cun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, *Neural Info. Process. Syst.* **2**, 396–404 (1992).
13. Y. Pao, *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, Reading, MA (1989).
14. J. Serra, *Image Analysis and Mathematical Morphology*, Academic, London (1982).

