

95/0079

Recognition of overlapping particles in granular product images using statistics and neural networks

F. Ros, S. Guillaume, G. Rabatel and F. Sevila

Image analysis can be used to characterize granular populations in many processes in the food industry or in agricultural engineering. Either global or individual parameters can be extracted from the image. However, granular products may agglomerate on the image, bringing bias measurements of individual parameters: products which agglomerate have to be recognized. This is done by a combination of image analysis (to pre-process and extract features), statistical methods (to reduce information) and neural network techniques (to take decisions).

Keywords: granular populations, image analysis, classification, multilayer neural network

INTRODUCTION

Characterization of a population of granular objects is needed for many industrial processes, particularly the food industry. According to its speed, accuracy and non-destructive operation abilities, image analysis is a well-adapted tool for this purpose. For instance, various authors have shown its potential for the study of cereal product granulometry [1–3]. Various types of parameters are used to characterize granular product images. Some of them are global and textural ones, such as constant grey-level run lengths [4–6]. In other cases, individual parameters usually related to the shape, the symmetry, the colour or the size are needed. Individual parameters may be very useful to evaluate some quality criteria (e.g. number of broken or irrelevant particles). Image grabbing can be done on flight because the flow need not be mechanically intercepted.

However, when populations of large particle size range must be characterized, the main problem is of flow thickness that induces particle overlapping. In this case errors will be made in the parameter measurements, leading to a bias in the product characterization.

This problem may be avoided using a manual system of separation of the particles and image grabbing need not be done in flight. Sometimes, this can be quite difficult when on-line characterization is required as, for instance, in the food industry.

For this reason, an efficient processing of overlapping particles is needed during the image analysis. Searchers have presented global methods to separate base particles. These methods were based on morphological transformations of the image that are applied to all particles, including those which are not overlapping [7,8]. These methods are often efficient but they may need a long time process if hardware solutions have not been studied. They can also significantly modify the shape, even if particles are not overlapping.

For these reasons, we have looked for a general method for agglomerate detection which does not transform the original image, and therefore does not modify base particles. The first purpose is only to detect agglomerates which should not be taken into account in further procedures. However, they should be processed later with specific procedures in order to try to find original particles. Some problems have been studied related to the following aspects.

Feature extraction and image pre-processing

The choice of features on the objects to be extracted is

CEMAGREF, 361 rue J.F. Breton, BP 5095, 34033 Montpellier Cedex 1, France. Received 2 August 1993; revised 29 October 1993; accepted 1 February 1994

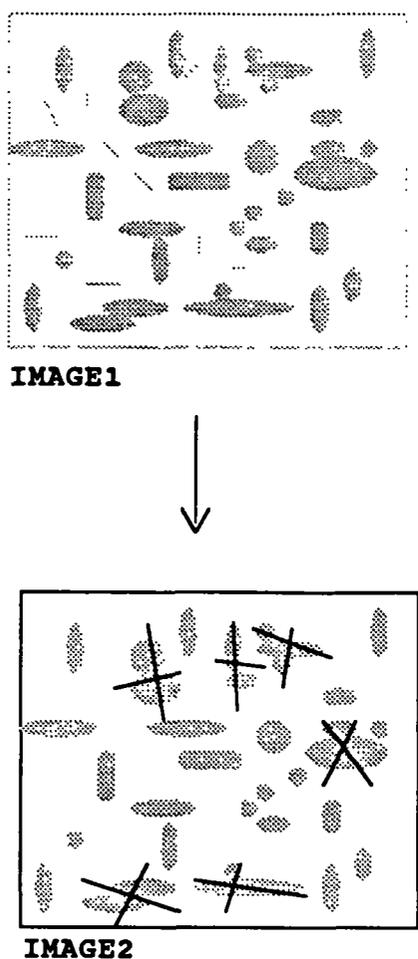


Figure 1 Example of agglomerate detection in granular population image

important. It influences the ability to find a relationship between them and a qualitative decision depending on objects. In a similar way, some complex procedures related to noise in the image can be avoided by pre-processing it.

Statistical considerations

Features extracted are often numerous and do not have the same pertinence. Therefore, this aspect of the study deals with the choice of techniques to transform features extracted in a new set of more pertinent virtual variables.

Artificial classifier design

This part of the study addresses the classification problem by multilayer neural networks. This classification tool has been entirely successful in many and varied applications [9-11]. Several interesting properties compared to more classical approaches are mentioned, as well as some difficulties in their practical use. Different classifiers which may or may not combine statistical considerations with multilayer neural networks are compared in order to point out the contribution of pre-processing.

OVERLAPPING PARTICLE DETECTION PRINCIPLE

The main purpose of this study is to develop a system capable of detecting whether an object in the scene is single or composed of particles which are overlapping. Overlapping particles can then be taken into account in further image analysis. The result wanted can be summarized in *Figure 1*. Image 1 is composed of single particles and overlapped particles. To avoid bias in the population characterization, it is necessary to consider only particles belonging to image 2. Particles with a small area (inferior to a fixed threshold depending to the optical grabbing system) are likely to be noise, therefore they are also rejected.

The problem of detecting agglomerated particles can be seen as a classical recognition problem. Therefore, in order to solve it, pertinent features must be extracted and be fed into the input of a decision system. These pertinent features will generally be obtained from a larger set of initial features using statistical techniques of selection and combination. Very often, feature vectors with a large number of components makes the classification algorithms slow and adds noise to them. Indeed, the training set elements are more sparsely distributed as the dimension of the feature vector increases and, consequently, the training set becomes less representative of the shape of the class conditional density function. Therefore, the reduced feature vector is likely to improve the classification efficiency. This leads to the scheme shown in *Figure 2*.

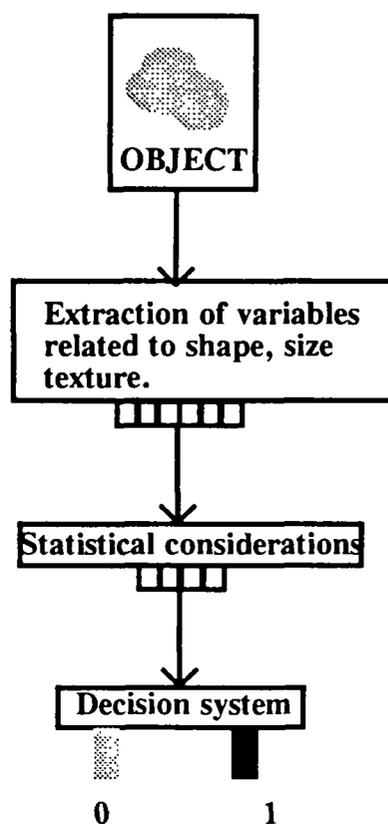


Figure 2 Summary of the principle

Image processing and feature extraction

Image pre-processing

At this stage, we suppose that all particles in the image can be isolated from the background: this can be achieved by appropriate grabbing conditions (e.g. lighting level) or by a previous segmentation step (e.g. grey level thresholding, colour selection) defined as a set of connected pixels (which does not belong to the background). In a lot of cases, some classical pre-processing in image analysis must then be applied in order to eliminate various types of noise: isolated outlines that would create bias in perimeter pixels, irrelevant contour evaluations etc. Some measurement evaluations of agglomerated objects may be very close to the ones for single objects if pre-processing has not been performed. Therefore, parameters which could be pertinent for the recognition problem lead to overlapped classes due to the presence of bias.

The enhancement process does not increase the inherent information content in the data, but it does increase the dynamic range of the chosen features. A large number of image enhancement techniques can be used, but often require interactive procedures to adapt them to the problem considered.

Feature extraction

The choice of the features extracted is largely dependent on the problem. Features related to size, shape, texture can be chosen. For the problem considered, shape parameters seem to be appropriate because overlapped shape objects are likely to be different from single objects. A binary image is sufficient to extract shape features.

Moreover, it seems useful to be independent from the optical characteristics of the grabbing system, therefore parameters without dimensions can be preferred. It may happen that an agglomerate is made of only two completely overlapped particles; in this case features related to the texture can be well adapted. For instance, if a method of constant grey level lengths is used, some discontinuities in the distribution of the constant grey level lengths of the object can be helpful to detect overlapped particles.

In any case, a subset of features has to be chosen thinking they are sufficient to build an agglomerate detection system. They are called 'possible parameters'. We propose 11 'possible variables without dimension' related to the shape of objects which can be extracted from them.

- (1) Object perimeter/convex object perimeter;
- (2) Object area/convex object area;
- (3) Object elongation (D_1 object/ D_0 object);
- (4) Convex object elongation (D_1 convex object/ D_0 convex object);
- (5) Distance (G_o, G_c)/squared (D_0 convex object)² + D_0 object)²;
- (6) Object area/(D_0 object)²;
- (7) Convex object area/(D_0 convex object)²;

- (8) Object area/(D_1 object)²;
- (9) Convex object area/(D_1 convex object)²;
- (10) D_0 object/ D_0 convex object;
- (11) D_1 object/ D_1 convex object.

where G_o is the gravity centre of the object, and G_c is the gravity centre of the convex object, D_0 and D_1 are the smaller and larger inertia axes of the object. They can be obtained by calculating the eigenvalues of the covariance matrix of the set of connected pixel coordinates defining the object.

Statistical considerations

It has been shown in various applications that it was worthwhile to pre-process the input features of a classification system (for instance multilayer neural network). Indeed, it improves the speed of the system: the fewer the features employed, the faster the algorithms will be and, often, the lower the cost.

It also eliminates redundancy. There is no point in extracting features that do not improve classification. Therefore, improvement and reduction of descriptive variables is needed.

Improvement of extracted variables

As mentioned above, this classification problem can involve a large number of features which are likely to be redundant and correlated. So, it seems necessary to rearrange them in a smaller set that are not correlated. PCA (principal component analysis) is an appropriate tool. It consists in calculating the eigenvectors and eigenvalues of the correlation matrix. The first eigenvectors determine the direction of maximum inertia of the objects represented in the feature space, which is equivalent to calculating the linear combination of features that contributes most of the variance. The second eigenvector determines the orthogonal axis to the first one, which associated to the first contributes the most of the inertia, and so on. The eigenvalues are a measurement of the importance of the contribution of each associated axis to the global variance.

PCA can be used to reduce the number of variables represented by the physical measures to a smaller number of components with little loss of information [12]. Therefore, 'possible parameters' are transformed into new 'possible parameters' which are not correlated and less numerous.

Selection of pertinent variables

The transformation by PCA provides a new reduced set of independent variables in respect of importance of total variance. However, all the variables obtained do not have equal pertinence. Especially, it seems important to note that variables with large variance may not be pertinent to the classification problem and may even be useless.

PCA addresses the aspect of reduction and improvement of input variables but not the one of the pertinent variable research for the problem considered. It is also

useful to design a tool which searches an optimum subset of pertinent variables and, especially, processes by considering a set of input patterns corresponding to decision classes. Thus, when there are a lot of quantitative independent variables, if evaluations of discrimination model are based on all these variables, good results are often obtained with training data, but poor results with test data. If a subset of these variables is used, results are likely to be more accurate and robust.

Methods of variable selection have been presented in the literature. SDA (stepwise discriminant analysis) [13] is one of the most famous. The selection of variables in the case of stepwise discriminant analysis is based on the Wilks Lambda criterion.

$$\text{Lambda } (r) = (\text{DET } (W)/(\text{DET } (V)))$$

where r is number of independent variables in use, V is the matrix of variances, W is the matrix of within-groups variances and DET is the determinant of V or W .

At each step, the algorithm decreases the Wilks Lambda. Two methods are defined: forward selection or backward selection. The first one begins by including in a model the single best discriminating feature, in terms of maximizing the mentioned variance ratio. This feature is coupled with each of the other features and a second is selected in the same manner. The third and subsequent features are chosen similarly. As new features are included, some of those previously selected can be removed from the model if the information they contained is available in some linear combination of the other included features. The process is stopped when the features not in the model do not significantly improve the variance ratio. The second one is similar, but it begins with all the variables and tries to eliminate at each step the less significant variable according to the variance ratio. It may be useful to combine the results obtained by the two approaches in order to keep variables only selected by the two. Consequently, the number of possible variables can be reduced according to this methodology.

Artificial classifier design

Many authors have studied the behaviour of neural networks and have shown their ability to solve classification problems [14]. Artificial neural networks or connectionist models are artificial intelligence systems which attempt to achieve good performance via dense interconnection of simple computational elements. The behaviour of an artificial network is given by its structure and the strength of the connections. The learning algorithm consists in determining connection strengths that are adapted to a given classification problem.

A MNN (multilayer neural network) is a feed-forward model with one or more layers of computing elements between the input and output layer. The elements of each layer are connected with elements of the previous layer(s). In a layer, there are no connec-

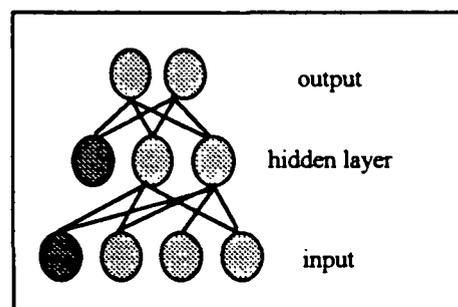


Figure 3 Architecture of one hidden layer neural network

tions between the elements (Figure 3).

A MNN functions as follows: the inputs are copied in the input layer. Each neurone computes its output value by an activation function $f(I)$, where I is given by:

$$I = \sum_{i=1}^N w_i x_i + \theta$$

where N is the number of inputs, w_i a weight connection, x_i the output of the connected element and θ 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 [15]. It is an iterative gradient algorithm designed to minimize the mean square error between the output signal of the network and the desired output. Detailed descriptions of this algorithm have been published.

Performance and criticism

MNN as classifiers provide several potential advantages compared to existing approaches [17,18]. First, they can provide the high computation rates required for problems using many simple processing elements operating in a parallel fashion. Second, they provide a good degree of robustness or fault tolerance. Third, they are non-parametric and make weaker assumptions concerning the shapes of underlying distributions than a traditional statistical classifier. However, they have also been criticized by several statisticians, e.g. Ripley [19], for their lack of theoretical advances.

Building a multilayer neural network

The design of architecture adapted to a real complex problem is not easy because the training phase cannot be made automatically. Parameters related to the back-propagation rule have to be adjusted according to each problem considered. Only the combination of the theoretical techniques concerning the MNN and human experience can lead to good training.

Two aspects have to be looked at. The first concerns the architecture of the MNN in terms of the number of neurones. Convergence of the back-propagation algorithm is widely dependent on the number of neurones. If the training base, as often occurs, is not balanced and if the number of neurones is numerous, the neural network is likely to stabilize itself quickly. This leads to results which are likely to be good on the training base

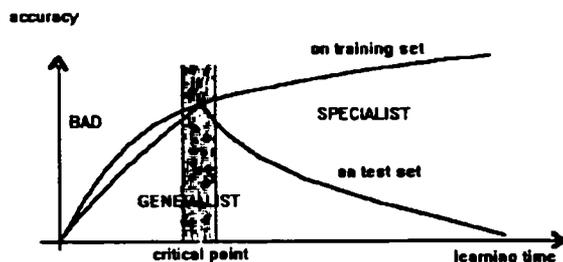


Figure 4 Accuracy of multilayer neural network in function of time training

and poor on the test base. Therefore, the smaller the neurone number the more the neural network will be likely to have good results when pattern not learned is presented as input. But if the number of neurones is too small, only an insufficient part of the relationships between inputs and outputs will be learned during the training. Thus, a compromise has to be found in order to obtain optimal architecture.

The second aspect concerns the number of iterations of training. During training, if the neurone number is sufficient, accuracy can be improved to provide a very good result. As a number of iterations increases, the neural network tends to specialize itself to the training data by capturing some details; therefore performances are likely to be good only for patterns belonging to the training base. In fact, the performance increases during the learning phase until a critical point is reached. Beyond this point, the performance on the test base decreases. The best results are obtained when training is stopped around this critical point (Figure 4). A method to avoid specialization consists in dividing the training base in two parts. With the first one, the network is trained, and with the second one the training is stopped.

APPLICATION TO FOOD PRODUCTS

This method has been applied to a classification problem. The purpose of the study was to find an artificial system that is able to evaluate the quality of grains as well as can be done by experts, in order to assist their decisions. Five classes of quality have been defined. They are mainly related to the food quality of the granular product and to various types of foreign bodies in it (e.g. sticks and stones). Several samples belonging to the classes defined by the experts are stored, and product of each sample was poured down in air in order to obtain a bank of images related to the different classes defined.

Individual parameters were convenient to evaluate the quality and, therefore, detection of agglomerates was needed. By an appropriate software and using human vision (to show agglomerates in the image), agglomerates and single granules belonging to 100 of these images are extracted and used to make the classifier. A similar procedure has been made to build the test base.

Agglomerates

Describing the shape of one particularity of agglomerates is difficult. In the application they are made of two or three different biological objects which are randomly connected. However, it is always possible to attribute different nature probabilities of agglomerates depending on the knowledge of food product.

RESULT AND DISCUSSION

Statistical considerations

Improvement of variables

The use of PCA makes it possible to obtain new non-correlated variables which are a combination of the previous ones. After the translation of the original data set into orthogonal components, only those components which have a contribution of >1% of the total variance present in the original database are selected. Therefore seven components (PCA variables) have been selected. Although each component is a linear combination of all the measured variables, we can distinguish some variables which contribute more than the others to the components. Among these, the first seven main components include:

- (1) 46% of the total variances;
- (2) 26% of the total variance;
- (3) 15% of the total variance;
- (4) 8% of the total variance;
- (5) 3% of the total variance;
- (6) 2% of the total variance;
- (7) 1% of the total variance.

Consequently, we can verify that the variables were correlated. In any case, principal component analysis is useful only if variables are correlated.

Selection of pertinent variables

The use of SDA allows selection of four variables (component 1, component 2, component 4, component 6) according to the criteria of stepwise discriminant analysis. This subset of variables has to be more robust than the previous one.

Neural network design

Architecture

The four variables resulting from statistical considerations are the input of a two-output multilayer neural network. The first output represents the score of agglomerates and the second output represents the score of single products. If outputs expected are 0 and 1, it will be useful to define a fuzzy area such as [0.45, 0.55] representing a zone where the detection system meets difficulty in taking decision.

Several experiments have been realized to show the contribution of statistical considerations, the contribution of training base division and the number of neurones. Table 1 represents the results obtained and a summary of techniques is shown in Figure 5.

Table 1 Best results obtained on the training and test base with different approaches

Approach	Hidden neurones	Training (%)	Test (%)
MNN	7	74	65
PCA + MNN	6	85	75
SDA + MNN	6	75	65
PCA + SDA + MNN	4	90	82

These results have been obtained by applying the method described above: the training was stopped around the critical point to avoid overtraining of the neural network. PCA promotes the process of the neural network. The number of iterations required to obtain best results decreases when input variables are not correlated. In any case, it was shown that the first hidden layer of a perceptron realizes transformations which can be compared to a PCA. In this problem, stepwise discriminant analysis makes it possible to improve the results but less than PCA.

This can be explained in two ways. First, variables introduced in the neural network are correlated and numerous enough. Accordingly, the neural network may need a long time to train the relationships between inputs and outputs. However, by increasing the time to train, the neural network tends to learn only the pattern belonging to the training base, and to become a bad classifier when unknown patterns are introduced to its input.

Second, neural networks can realize a natural selection if the number of inputs are not too large. If the number of variables is large and is reduced much more, SDA would be more efficient. Combination of the two statistical approaches makes it possible to obtain better results. Similar results have been obtained [20], where we combined a multiple linear regression and a MNN in order to simulate the behaviour of an expert in plant grading. The first approach improves the variables, and the second helps the neural network to find relationships between the two classes and only pertinent variables.

Experiments have been made by others in order to point out the influence of neurone number in the training of neural networks with back-propagation algorithm. The results are dependent on the number of neurones belonging to the hidden layer. When the number of neurones is less than four, the neural network has some difficulties in learning. It can only find a general relationship between inputs and outputs. When the number of neurones is more than four, the neural network tends to specialize itself to the training

Table 2 Best results obtained with various numbers of hidden neurones without taking into account the critical point

	Neurone number					
	2	4	5	9	15	20
Training base (%)	50	95	96	97	97	98
Test base (%)	40	75	75	72	72	67

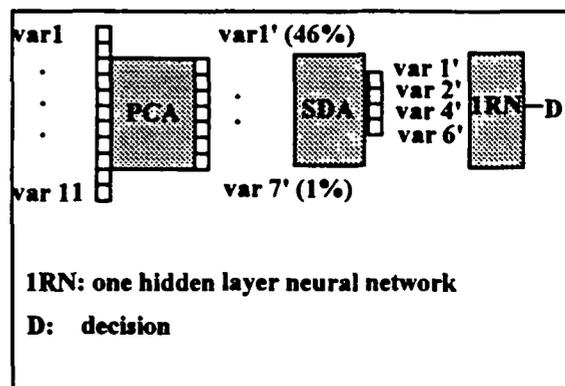


Figure 5 Summary of the techniques used in the application

data. This leads to poor classification results when unknown patterns are presented to the input (Table 2).

CONCLUSION

The purpose of this research has been to design a method for on-line agglomerate detection. Evaluation of the method has been made on a real case and has been efficient. It has been shown that pre-processing using statistical considerations make the training of MNN easier. It makes it possible to provide 'better' information to the input of the neural network (i.e. without redundancy). Although MNNs have shown their ability to solve some difficult classification problems, using them demands some experience and knowledge. It has been shown that performances of MNNs in generalization are dependent on their internal architecture and the number of iterations during training. A further step could be the rebuilding of the different objects detected as agglomerates in order to take them into account individually in further processes. So, no information from the images should be lost.

This method is general enough to be applied to other products. Food industries (such as coffee lyophilization, milk dehydration) which transform powders and granular products are mainly concerned with such techniques.

REFERENCES

1. Bertrand D., Courcoux P., Autran J.C., Meritan R. and Robert P. (1990) Stepwise canonical discriminant analysis of continuous digitalized signals: application to chromatograms of wheat proteins. *J. Chemometrics* 4, 413-427
2. Bertrand, D., Robert, P., Melcion, J.P. and Sire, A. (1991) Characterisation of powders by video image analysis. *Powder Technol.* 66, 171-176
3. Sinfort, N., Sevilla, F. and Bellon, V. (1992) Interest of global analysis methods for the study of images of granular products grabbed on-flight, *Agricultural Engineering International Conference*, Uppsala, Sweden, pp. 81-89
4. Galloway, M.M. (1975) Texture analysis using grey level run lengths *Computer Graphics Image Proc.* 4, 172-179

5. Loh, H., Leu, J.-G. and Luo, R.C. (1988) The analysis of natural textures using run length features. *IEEE Trans. Ind. Electron.* **35**, 323–328
6. Gottlieb, C.C. and Kreyszig, H.E. (1990) Texture descriptors based on co-occurrence matrices. *Computer Vision Graphics Image Proc.* **51**, 70–86
7. Serra, J. (1992) *Image Analysis and Mathematic Morphology* Vol. 2, *Theoretical Advances* Academic Press, San Diego, pp. 101–114
8. Serra, J. (1989) *Image Analysis and Mathematic Morphology*. Vol. 1, *Theoretical Advances*, Academic Press, San Diego, pp. 318–360
9. Le Cun, Y., Boser, B., Denker, J.S. *et al.* (1992) Handwritten digit recognition with a back propagation network. *Neural Information Processing Systems 2*, 396–404
10. Fukushima, K. and Imagawa, T. (1993) Recognition and segmentation of connected characters with selective attention. *Neural Networks* **6**, 33–41
11. Rodellar, V., Naharro, F. and Garcia, C. (1991) A neural network for the extraction and characterization of the phonetic features of speech. *Proc. NEURO NIMES '91*, pp. 203–212
12. Resurrection A.V.A. (1988) Applications of multivariate methods in food quality evaluation. *Food Technol.* **42**, 128–136
13. Tomassonne, R. (1983) *La Régression Linéaire: Nouveaux Regards sur une Ancienne Méthode Statistique*. Masson, Paris
14. Lippmann, R.L. (1987) An introduction to computing with neural nets. *IEEE ASSP Magazine* **4**, 4–22
15. Pao, Y. (1989) *Adaptive Pattern Recognition and Neural Networks* Addison Wesley, Reading, USA
16. Shekhar, S., Amin, M.B. and Khandelwal, P. (1992) Generalization performance of feed-forward neural networks. *Neural Networks Advances and Applications* **2**, 68–79
17. Ros, F., Brons, A., Sevilla, F., Rabatel, G. and Touzet, C. (1993) Combination of neural network and statistical methods for sensory evaluation of biological products: on line beauty selection of flowers. *Proc. IWANN '93 Spain*
18. Brons, A. (1992) Contribution des Techniques Connexionistes à l'Evaluation Qualitative des Produits Agro-Alimentaires par leurs Aspects Visuels. Doctoral thesis, Engref of Paris
19. Ripley, B.D. (1992) Statistical aspects of neural networks. *Proc. Sem. Stat. Denmark* **2**–68
20. Brons, A., Rabatel, G., Ros, F., Sevilla, F. and Touzet, C. (1993) Plant grading by vision using neural networks and statistics. *Computers Electronics Agric.* **9**, 25–39