LE MATEMATICHE
Vol. XLVIII (1993) – Fasc. II, pp. 273-285

A NEURAL NETWORK APPROACH TO DISCRIMINATION BETWEEN DEFECTS AND CALYCES IN ORANGES

SALVATORE INGRASSIA - ENRICO COMMIS

The problem of automatic discrimination among pictures concerning either defects or calyces in oranges is approached. The method here proposed is based on a statistical analysis of the grey-levels and the shape of calyces in the pictures. Some suitable statistical indices are considered and the discriminant function is designed by means of a neural network on the basis of a suitable vector representation of the images. Numerical experiments give 5 misclassifications in a set of 52 images, where only three defects have been classified as calyces.

1. Introduction.

Research in data classification, one of the most classical topics in statistical analysis, in these years has undergone a great increase; several new problems have been recently proposed and many new techniques have been developed. In this direction, for example, studies in pattern recognition give a great contribution; in fact many important applications of pattern recognition can be characterized as either waveform classification or classification of geometric figures (Fukunaga, 1990). Another area of research that has given to new interest

Entrato in Redazione il 20 gennaio 1994.

Research in collaboration with Consorzio per la Ricerca in Agricoltura nel Mezzogiorno within the Galatea Esprit II Project 5293.

AMS 1991 Subject Classification: Primary: 62H30; Secondary: 68T10, 62H51

in data classification concerns neural networks – new algorithms for cognitive tasks, such as learning and optimization – which have now many applications in a wide range of statistical areas, first of all in classification and pattern recognition, but also in forecasting and non-linear regression (Ripley, 1992).

In this paper a new problem in data classification is considered: we approach the problem of the automatic discrimination of a set Ω of digital images concerning either defects or calyces in oranges into the corresponding classes, say Ω_D and Ω_C respectively. This problem is part of a larger project concerning defect detection in oranges: while an orange is carried on a belt, a camera takes digital pictures of it and then, if some defect is detected, the orange will be discarded by a suitable robot. The same problem for defect detection in potatoes has been recently approached by Grenander and Manbeck (1992) with techniques involving stochastic difference equations.

Given an image $\omega \in \Omega$, two main aspects are relevant to the problem stated above. First we should have a *numerical feature* of ω, say $x = x(\omega) \in \mathcal{X} \subseteq \mathbb{R}^k$, for some k, which contains essential information on ω; secondly we should have a *discriminant function* $\psi : \mathcal{X} \to \mathcal{Y}$, with $\mathcal{Y} \subseteq \mathbb{R}$. The set $\mathcal{Y}$ is then partitioned in two parts, say $\mathcal{Y}_C$ and $\mathcal{Y}_D$: if $\omega \in \Omega_D$ (Ω_C), then $y = y(x)$ should belong to $\mathcal{Y}_D$ ($\mathcal{Y}_C$). Following Hand (1981), Ω is the *measurement space* and $\mathcal{X}$ is the *feature space*; x is also called *vector representation* of ω (Fukunaga, 1991). Here the discriminant function ψ is designed by a feed-forward neural network which is trained in a subset of Ω.

Undoubtedly the choice of the feature space is the most delicate part of the problem, in fact the network must discriminate on the basis of the numerical feature (the input pattern) of the image ω. It seems quite obvious that any vector representation of an image should be based on two kinds of information: the grey levels of the image and the shape of the "object" represented in the image. The choice follows from the aims of the project. In our case we need only one bit of information: defect, no-defect. The method here presented gives good results in discriminating (the misclassifications were less than 10%) and does not require much computational effort. Obviously the two kinds of misclassification (calyces classified as defects and vice versa) do not have the same cost error: we prefer to discard a good orange rather than to accept a bad one.

The paper is organized as follows. In Section 2 we give a statistical analysis of the data; in Section 3 we describe the neural networks used in our case; in Section 4 we describe the input patterns used for the discrimination and give the results of the automatic discrimination; finally comments and conclusion are made in Section 5.

2. Data analysis.

In this section we summarize some statistical features of the set of digital images Ω here considered (we recall that a digital image is a matrix of positive integers giving the patterns of radiant energy emitted by the objects represented in the picture). The set Ω consists of 52 grey-levels images of 64×64 pixels concerning either mechanical defects (they are quite dark spots due to bruising) or calyces in oranges and it is partitioned in three classes:

- 18 images concerning mechanical defects (Ω_{md});
- 15 images concerning calyces without stem (Ω_{cns});
- 19 images concerning calyces with stem (Ω_{cws}).

We can consider two specific aspects of the pictures: the grey levels and the shape of the "object" represented.

Figures 1 shows three histograms, one for each type of images: with very few exceptions, they represent the trends of each class. In particular, the histograms of grey-levels concerning the images of the mechanical defects, suggested the idea of a segmentation into a binary image which extracts the subject of the picture from its background: in fact the presence of two peaks in a histogram demonstrates the existence of two distinct brightness regions in the image, one corresponding to the object and the other to its background. In this case, by means of thresholding, we can obtain a white object with a black background (see Figure 2).

Picture thresholding methods can be considered also for images of calyces, even if the corresponding histograms look different. The reason is the following: mechanical defects in general have a quite different brightness from the surrounding, while in calyces the two brightness regions are quite near, in part they coincide. In fact if we look at an orange, we can see that near the calyx there are some swellings and depressions (often some shadows increase this phenomenon). Picture thresholding highlights these features for our aims. The problem will be, how to make this evident by means of a numerical pattern.

In the rest of the Section, we shall consider the thresholding problem. Given a digital image $\omega \in \Omega$, we suppose that it contains two populations, say $\omega^{(0)}$ and $\omega^{(1)}$ (respectively background and object). Denote with ω_{hk}, the grey level of the pixel (h, k). Let us assume that if $\omega_{hk} \in \omega^{(0)}$ then $\omega_{hk} \sim N(\mu_0, \sigma_0^2)$ otherwise $\omega_{hk} \in \omega^{(1)}$ and then $\omega_{ij} \sim N(\mu_1, \sigma_1^2)$, where $N(\mu_i, \sigma_i^2)$, $i = 0, 1$ denotes gaussian distributions with mean and variance respectively μ_i and σ_i^2. Finally we denote with $b = b(\omega)$ the binary image of ω as the $n \times n$ matrix such that $b_{hk} = b(\omega_{hk}) = 0$ if $\omega_{hk} \in \omega^{(0)}$ and $b_{hk} = b(\omega_{hk}) = 1$ if $\omega_{hk} \in \omega^{(1)}$.

Given any value ω_{hk}, $h, k = 1, \ldots, n$, the problem is to decide if either $\omega_{hk} \in \omega^{(0)}$ or $\omega_{hk} \in \omega^{(1)}$. Let p_i, $i = 0, 1$, be the prior probability associated with

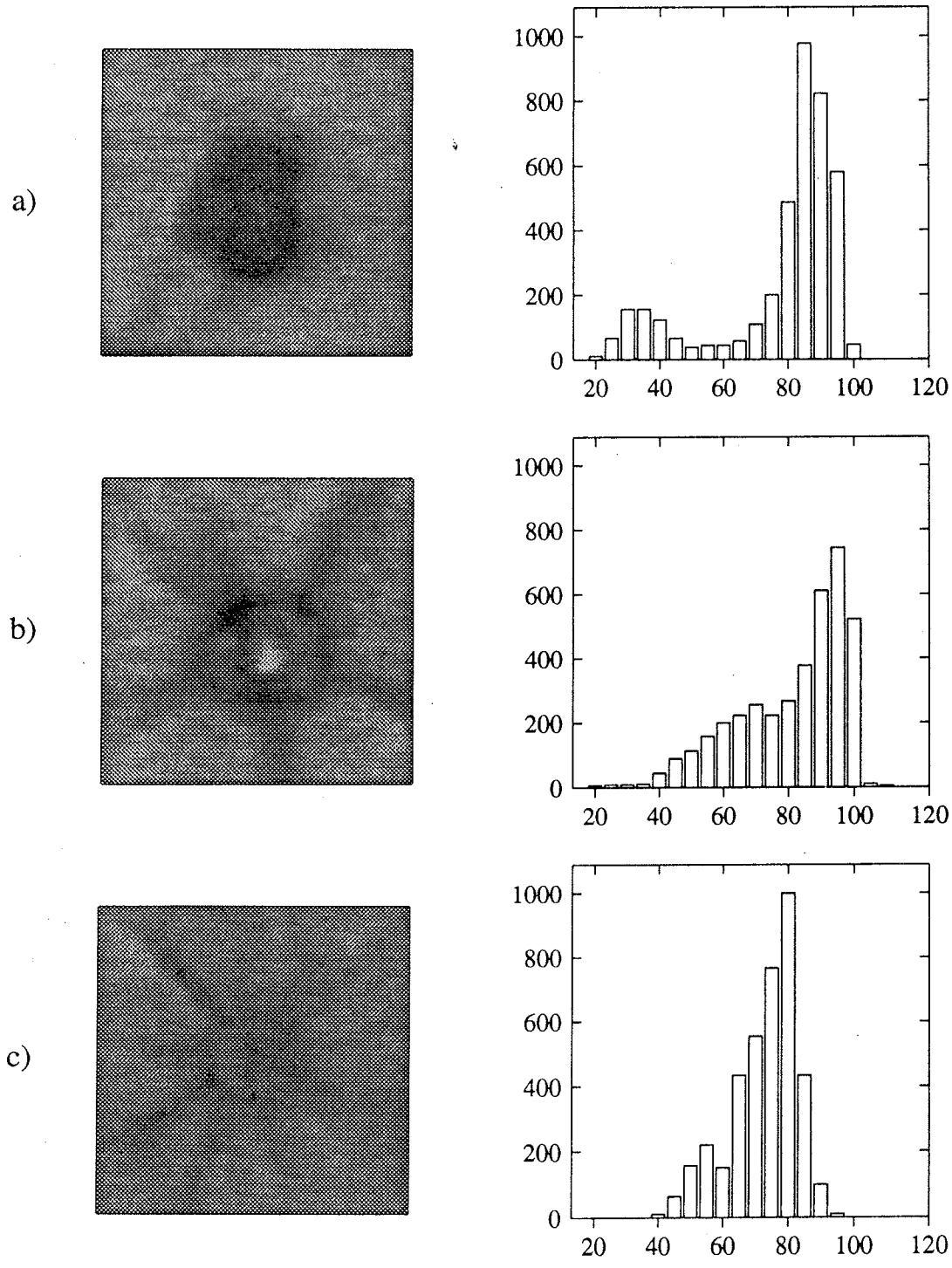


Figure 1. Examples of images and corresponding grey-levels histograms: a) defect; b) calix with stem; c) calix without stem.

the population $\omega^{(i)}$ and $L(\omega_{hk} \mid \omega^{(i)})$ be the likelihood under $\omega^{(i)}$. The Bayes test for minimum error assigns an observed grey level ω_{hk} to that population $\omega^{(i)}$ which maximizes the likelihood (see e.g. Mardia and Hainsworth (1988), Fukunaga (1990)); then $\omega_{hk} \in \omega^{(0)}$ if and only if

$$\frac{p_0 L(\omega_{hk} \mid \omega^{(0)})}{p_1 L(\omega_{hk}) \mid \omega^{(1)})} > 1 . \tag{1}$$

The hypotheses stated above imply that:

$$L(\omega_{hk} \mid \omega^{(i)}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{ - \frac{(\omega_{hk} - \mu_i)^2}{2\sigma_i^2} \right\} . \tag{2}$$

Suppose $\sigma_0 = \sigma_1 = \sigma$. In this case (1) holds if and only if

$$\omega_{hk} \geq t \tag{3}$$

where

$$t = \frac{1}{2}(\mu_0 + \mu_1) + \frac{\sigma^2}{\mu_1 - \mu_0} \log (p_0/p_1) . \tag{4}$$

The general case $\sigma_0 \neq \sigma_1$ is approached in Mardia and Hainsworth (1988).

Many methods have been proposed in order to find a threshold t for segmentation (see e.g. Weszka et al. (1974), Ridler and Calvard (1978), Otsu (1979), Mardia and Hainsworth (1988)). We used an adaptation of the iterative method of Ridler and Calvard (1978) proposed in Mardia and Hainsworth (1988) which makes use of spatial threshold values. Denote with $\Phi_2^0(\omega_{hk})$ the 3×3 neighbourhood of ω_{hk}, that is the set $\{\phi_0(\omega_{hk}), \phi_1(\omega_{hk}), \phi_2(\omega_{hk})\}$, where the sets $\phi_p(\omega_{hk})$, with $p = 0, 1, 2$ are given in Figure 3. The main steps of this algorithm are:

1. segment the image into two regions, $\omega^{(0)}$ and $\omega^{(1)}$, using the mean value;
2. calculate the respective mean grey level $\overline{\omega^{(0)}}$ and $\overline{\omega^{(1)}}$ and the associated number of pixels n_0 and n_1;
3. calculate the spatial threshold value

$$t = \frac{1}{2}(\overline{\omega^{(0)}} + \overline{\omega^{(1)}}) + \frac{\sigma^2}{\overline{\omega^{(1)}} - \overline{\omega^{(0)}}} \log(n_0/n_1) ; \tag{5}$$

4. for each $\omega_{hk} \in \omega$: if the mean value of $\Phi_2^0(\omega_{hk})$ is less than the threshold value t in (5), then assign ω_{hk} to $\omega^{(1)}$;

5. repeat steps 2-4 until the solution is stable.

Two steps of median filtering (see e.g. Mardia and Hainsworth (1988), Rosenfeld and Kak (1982)) complete the image segmentation: each $b_{hk} \in b(\omega)$ the median filter is classified as the median of the set $\Phi_2^0(b_{hk})$. In our case with the notation introduced above we have:

$$b_{hk} = \begin{cases} 1 & \text{if } \sum_{b_{ij} \in \Phi_2^0(b_{hk})} b_{ij} \geq 5 \\ 0 & \text{otherwise} \end{cases}$$

3. Neural networks for data classification.

Neural-network models are algorithms for cognitive tasks, such as learning and optimization, which are in a loose sense based on concepts derived from research into the nature of the brain. One of the most important applications of neural networks concerns classification of elements of a statistical population into classes; in other words, given a numerical pattern $x \in \mathcal{X}$ the network should indicate to which of say l classes x belongs.

The literature about neural networks is now quite considerable. Here we sketch main ideas, the interested reader, for example, is referred to the paper of Ripley (1992), which highlights statistical aspects of neural networks, or to the recent monograph of Pao (1989) and bibliography cited therein.

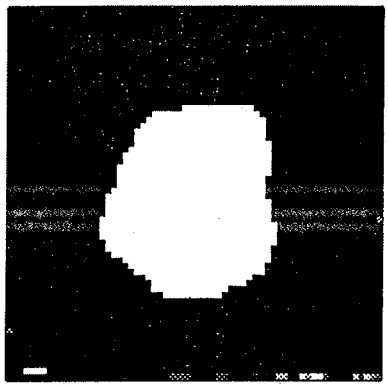
In mathematical terms a *neural network model* is defined as a directed graph with the following properties:

1. a state variable n_i is associated with each node i;
2. a real-valued *weight* w_{ij} is associated with each edge $\{i, j\}$ between the nodes i and j;
3. a real-valued *bias* θ_i is associated with each node n_i;
4. a transfer function $f_i[n_p, w_{ij}, \theta_i, (j \neq i)]$ is defined, for each node i, which determines the state of the node as a function of its bias, of the weight of its incoming edges, and of the states of the nodes connected to it by these edges.

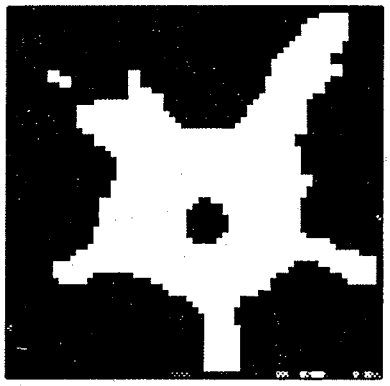
In the standard terminology, the nodes are called *neurons*, the edges are called *synapses*, and the bias is known as the *activation threshold*. The transfer function usually is taken to have the form $f(\sum_p w_{ij} n_p - \theta_i)$ where $f(x)$ is either a discontinuous step function or its smoothly increasing generalization known as a logistic function:

$$f(x) = \frac{1}{1 + e^{-(x-\theta)}} .$$

a)



b)



c)

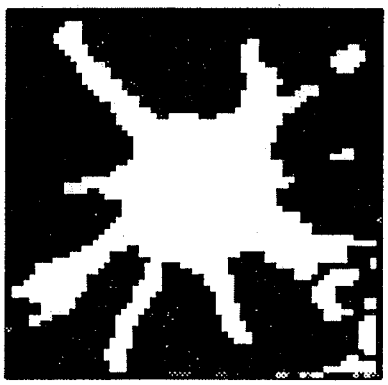


Figure 2. Binary versions of the images given in Figure 1.

Nodes without edges towards them are called *input* neurons; *output* neurons are those with no edges leading away from them. A *feed-forward* network is one whose topology admits no closed paths. It has been proved to be an effective system for learning discriminants for patterns from a set of examples. Such nets consist of sets of nodes arranged in layers: input nodes, output nodes and nodes in hidden layers. Connection of nodes are only allowed between different layers

but not inside a layer.

The learning procedure has to select both all the weights $\{w_{ij}\}$ and the biases $\{\theta_i\}$. This is done by presenting the training examples in turn several times, whilst aiming to minimize a certain cost function, for example the total squared error

$$E = \frac{1}{2} \sum_q \| y^q - \xi^q \|^2 \tag{6}$$

where $y^q \in \mathscr{Y}$ is the output for input x^q, ξ^q is the target output and q is the index of the training set. In our case $\mathscr{Y} = [0, 1]$ and $\xi^q \in \{0, 1\}$ for each pattern q.

4. Input patterns and results.

In this section we discuss the choice of the vector representation of the images $\omega \in \Omega$ and give the results of the discrimination via the neural network.

The statistical description of the data given in Section 2, suggests the idea of a first discrimination between defects and calyces based on grey levels. The input patterns are based on the frequencies of 15 classes. First we considered the 15-component vector giving the frequencies of these classes. Afterwards, according to the data analysis of Section 2, we noticed that the information is mainly contained not in the values of the frequencies but in their variations. The final input patterns have been binary ones with 1 if $f_{i+1} > f_i$ and 0 otherwise, where f_i $(i = 1, \ldots, 14)$ are the frequencies of the classes introduced above.

The discrimination function ψ between Ω_{md} and $\Omega_{\text{cws}} \cup \Omega_{\text{cns}}$ has been by considering a feed-forward neural network having the 15 neurons in the input layer, 5 neurons in the hidden layer and only one neuron in the output layer.

Furthermore better results have been obtained by adding a new field in the vector x introduced above, which takes into account the shape of the object represented in the picture. In fact, as showed in Section 2, the difference of shape between thresholded images of mechanical defects and calyces is quite clear: pictures concernig calyces are, more or less, star-shaped. This feature helps in the discrimination (obviously defects with strange shape can be encountered); the problem is to perform a numerical field able to point out this feature. Our idea is the following. Given the thresholded image of any $\omega \in \Omega$, let us introduce the matrix $N = (N_p(b_{hk}))_{hk}$ defined as follows:

$$N_p(b_{hk}) = \begin{cases} 0 & \text{if } b_{ij} = 0 \\ \displaystyle\sum_{b_{ij} \in \Phi_p^0(b_{hk})} b_{ij} & \text{if } b_{ij} = 1 \end{cases}$$

where $\Phi_p^0(b_{ij}) = \bigcup_{l=0}^{p} \phi_l^0(b_{ij})$, with $\phi_0^0(b_{ij}) = b_{ij}$ is the neighbourhood of order l of the pixel ω_{ij}. Thus any element of the matrix N gives the number of pixels put on 1 in the neighbourhood of b_{hk} if $b_{hk} = 1$ and 0 otherwise.

Afterwards, let us consider the part of the image in which only the "rays" appear: in this case, if $N_p(b_{hk})$ has a very small value then $b_{ij} = 1$ for a very few $b_{ij} \in \Phi_p^0(b_{hk})$ and ω_{hk} could belong to either a spot or the center of a calyx; otherwise, if $N_p(b_{hk})$ has a suitably large value, then $b_{ij} = 1$ for many $b_{ij} \in \Phi_k^0(b_{hk})$ and ω_{hk} could belong to a ray of a calyx. Hence the average number $\overline{N_p}$ of the set of the pixels equal to 1 in the neighbourhoods of order p, for a suitable p, gives a shape index.

			17	16	15	16	17			
	14	13	12	11	10	11	12	13	14	
	13	9	8	7	6	7	8	9	13	
17	12	8	5	4	3	4	5	8	12	17
16	11	7	4	2	1	2	4	7	11	16
15	10	6	3	1	•	1	3	6	10	15
16	11	7	4	2	1	2	4	7	11	16
17	12	8	5	4	3	4	5	8	12	17
	13	9	8	7	6	7	8	9	13	
	14	13	12	11	10	11	12	13	14	
			17	16	15	16	17			

Figure 3. Sets $\phi_p(\bullet)$ for $p = 1, \ldots, 17$: the numbers indicate the order of the model of neighbourhood of $\bullet$.

Now let us give a formal description of the method. First we must highlight the rays and the lengthened shape from a thresholded image. Choose a threshold c: if $N_p(b_{hk}) > c$ then set to zero all the pixels belonging to $N_p(b_{hk})$ (we have chosen neighbourhoods up to order 17, see Figure 4 for details). Afterwards, for the image thus obtained, let us compute the new matrix N and then the

average number $\overline{N_p}$ of $\{N_p(b_{hk})\}_{h,k=1,\ldots,n}$. If $\overline{N_p}$ is quite large, then the picture should represent a calyx, otherwise it should represent a mechanical defect. The discriminant function ψ has been designed by means of the back-propagation neural networks described above. In order to have a total squared error (6) less than 0.01, about 300 iterations were needed in both networks. The results have been summarized in Figure 4 which gives the values of the discriminant functions. We see that for the threshold values $c = 0.4$ there are five misclassifications: two mechanical defects three calyces (two with stem and one without stem) are misclassified. As we stated in the introduction, the two kinds of errors have a different cost: misclassification of a defect is considered a worse error than misclassification of a calyx.

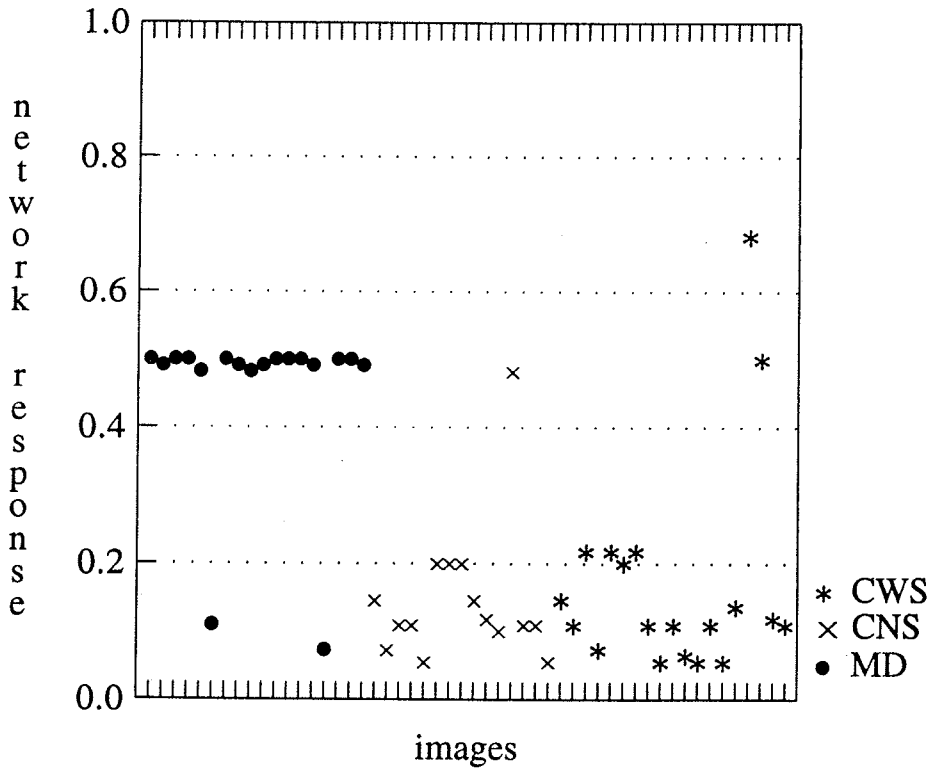


Figure 4. Values of the discriminant function. Where CWS: calyx with stem; CNS: calyx without stem; MD: mechanical defect.

We also performed some experiments with input patterns based directly on frequencies: more or less the results were the same, but in this case about 12,000 iterations were needed for the training phase.

Other numerical experiments have been made for the discrimination between Ω_{md} and $\Omega_{cws} \cup \Omega_{cns}$ considering only the grey-levels and the index $\overline{N_p}$. In the first case the simple grey-levels analysis gave a misclassification rate of about 15 %, while the simple shape analysis gave a misclassification rate of about 30 %.

5. Further remarks and conclusions.

Finally we give remarks and comments both on the procedures here proposed and on the results obtained. First of all, as we said in the introduction, we remember that we need only one bit of information; hence the problem is to have sufficient information to be able to discriminate between defects and calyces, with an acceptable error rate, in a short time (compared with human work). In the method here proposed, which performs a good compromise between the two goals, the central role is played by the choice of the vector representation of any image.

Undoubtely other choices of vector representations of an image are possible, for example via image segmentation based on random Markov fields models (see e.g. Geman (1991)), but we didn't follow this approach for two main reasons. First we saw that the analysis of grey levels give good results in discrimination and it does not require much computational effort. Moreover we remember that we are interested in only one bit of information (defect, no-defect) and, in order to reach this target, precise texture analysis seems too expensive. In fact, algorithms approaching such problems (e.g. Gibbs sampler, ICM, etc.) require a lot of computational resources and hence the time required for obtaining the input pattern for the network could be not competitive with human work.

Moreover the approach based on frequency distribution of grey levels is easily implemented and highly parallelizable: the image can be partitioned into some parts, each of them is analyzed by a different processor and then the results are collected.

The shape of the object reproduced in the thresholded image is also quite informative: it has been taken into account following the same criteria given above. Many other approaches are possible by the data analysis of Section 2. For example the grey-levels histograms of Figure 1-4 suggest the description of the data as a two-component normal mixture distribution and an automatic classification by means of the estimated parameter of the mixture. This is an outline for further work.

Image analysis by methods of moments constitutes another approach (see e.g. Teh and Chin (1988)). The same problem has been recently approached

(Capizzi, 1992) by methods based on circle polynomials of Zernike which have certain invariance properties (e.g. see Section 9.2 in Born and Wolf (1975)). More exactly, the vector representation of any image is constituted by some Zernike moments (the discriminant function is again designed by a neural network). This method gave excellent results in discrimination between Ω_{md} and Ω_{cns}, but it was not tested with the other classes (which give more problems). Zernike approach is not new in image analysis (see e.g. Teh and Chin (1988), Khotanzad and Hong (1990)), even if it must be much further explored.

Acknowledgements. The authors thanks Proff. Angelo M.Anile and Giuseppe Lunetta for their suggestions and encouragements. Thank also are due to Marcello Iacono Manno and Salvatore Capizzi for helpful discussions.

REFERENCES

[1] M. Born - E. Wolf, *Principles of optics,* Pergamon Press, New York, 1975.

[2] S. Capizzi, *Technical Report,* Consorzio per la Ricerca in Agricoltura nel Mezzogiorno, Catania, 1992.

[3] G.R. Cross - A.K. Jain, *Markov random field texture models,* IEEE Transaction on Pattern Analysis and Machine Intelligence, n. 1, January 1983, pp. 25–39.

[4] K. Fukunaga, *Introduction to Statistical Pattern Recognition,* Academic Press, 1990.

[5] D. Geman, *Random fields and inverse problems in imaging,* "École d'Été de Probabilités de Saint-Flour XVIII", Lecture Notes in Mathematics n. 1427, Springer-Verlag, Berlin, 1991, pp. 115–193.

[6] U. Grenander - R. Manbeck, *A stochastic shape and color model for defect detection in potatoes,* Rep. Patt. Th. 156, Division of Applied Mathematics Brown University, Providence R.I., 1992.

[7] D.J. Hand, *Discrimination and classification,* John Wiley & Sons, Chichester, 1981.

[8] A. Khotanzad - Y.H. Hong, *Invariant Image Recognition by Zernike Moments,* IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. PAMI-12, n. 5, May 1990, pp. 489–497.

[9] K.V. Mardia - T.Y. Hainsworth, *A spatial thresholding method for image segmentation,* IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. PAMI-10, n. 6, November 1988, pp. 919–927.

[10] N. Otsu, *A threshold selection method from grey-level histograms,* IEEE Transaction on Systems, Man and Cybernetics, vol. SMC-9, n. 1, January 1979, pp. 62–66.

[11] Y.H. Pao, *Adaptative pattern recognition and neural networks,* Addison-Wesley, 1989.

[12] T.W. Ridler - S. Calvard, *Picture thresholding using an iterative selection method,* IEEE Transaction on Systems, Man and Cybernetics, vol. SMC-8, n. 8, August 1978, pp. 630–632.

[13] B.D. Ripley, *Statistical aspects of neural networks. Technical report,* Department of Statistics, University of Strathclyde, Glasgow, 1992.

[14] A. Rosenfeld - A.C. Kak, *Digital Picture Processing* vol. 2, Academic Press, Orlando, 1982.

[15] C.H. Teh - R.T. Chin, *On image analysis by the method of moments,* IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. PAMI-10, n. 4, July 1988, pp. 496–513.

[16] J.S. Weszka - R.N. Nagel - A. Rosenfeld, *A threshold selection technique,* IEEE Transaction on Computers, December 1974, pp. 1322–1326.

Salvatore Ingrassia,
Istituto di Statistica,
Facoltà di Economia e Commercio,
Corso Italia 55,
95129 Catania (Italy)
and
Enrico Commis,
Dipartimento di Matematica,
Viale A. Doria 6,
95125 Catania (Italy)