

Feature Selection by a Genetic Algorithm. Application to Seed Discrimination by Artificial Vision

Younes Chtioui,^{1*} Dominique Bertrand¹ and Dominique Barba²

¹ Institut National de la Recherche Agronomique, Laboratoire de Technologie Appliquée à la Nutrition, BP 1627, Rue de la Géraudière, F-44316 Nantes Cedex 03, France

² Institut de Recherche et d'Enseignement Supérieur aux Techniques de l'Électronique, Laboratoire des Systèmes Electroniques et Informatiques, La Chantrerie, CP 3003, 44087 Nantes Cedex 03, France

(Received 2 October 1996; accepted 24 June 1997)

Abstract: Genetic algorithms (GAs) are efficient search methods based on the paradigm of natural selection and population genetics. A simple GA was applied for selecting the optimal feature subset among an initial feature set of larger size. The performances were tested on a practical pattern recognition problem, which consisted on the discrimination between four seed species (two cultivated and two adventitious seed species) by artificial vision. A set of 73 features, describing size, shape and texture, were extracted from colour images in order to characterise each seed. The goal of the GA was to select the best subset of features which gave the highest classification rates when using the nearest neighbour as a classification method. The selected features were represented by binary chromosomes which had 73 elements. The number of selected features was directly related to the probability of initialisation of the population at the first generation of the GA. When this probability was fixed to 0.1, the GA selected about five features. The classification performances increased with the number of generations. For example, 6.25% of the seeds were misclassified by using five features at generation 140, whereas another subset of the same size led to 3% misclassification at generation 400. The present work shows the great potential of GAs for feature selection (dimensionality reduction) problems. © 1998 SCI.

J Sci Food Agric **76**, 77–86 (1998)

Key words: feature selection; genetic algorithm; seed; colour image analysis; classification; discrimination

INTRODUCTION

In many artificial vision studies, morphological features are extracted from the images. These features are then used as variables for the discrimination of the nature of the objects. The performances of the pattern recognition system are directly related to the relevance of the measured features, and to the method of classification. However, measured features usually present mutual interactions (correlations). The selection of the 'optimal'

subset of features among a set of available ones is a necessary and important step in pattern recognition, because it allows the reduction of redundancy. Since the introduction of a new feature in the predictive model may increase the computation time, it is generally strongly needed to implement systems with few but relevant features.

Feature selection is a difficult task in pattern recognition, because it requires searching through a space which may be high-dimensional. An exhaustive search is computationally prohibitive especially when there are a large number of features. This has led to the development of a wide range of feature selection methods (dimensionality reduction). The most widely applied

* To whom correspondence should be addressed.
Contract/grant sponsor: National Seed Testing Station of France.

methods for feature selection introduce (or remove) variables in a stepwise manner (Siedlecki and Sklansky 1988). For instance, *sequential direct selection* starts by selecting the best feature among the whole set of available ones. At each step, it keeps the latest selected subset of features and adds to it the best feature among the remaining features. This method stops when the introduction of a new variable does not significantly improve the performances. In contrast to sequential direct selection, *sequential backward selection* starts by selecting the complete set of features and rejects one feature at a time. At a given step, the feature which causes the smallest decrease of the performances is permanently removed. The process stops when the removal of each of the remaining features causes a high decrease of performances. The main weakness of these sequential methods is that they may reject a variable at an early step of the search process which is then unavailable for a later possible improvement in performance. This means that this kind of feature selection methods may lock at a local minimum during the search process. An alternative is the use of *genetic algorithms* (GAs) which are able to explore, in an efficient way, high-dimensional spaces for function optimisation (Holland 1975). They derive their name from the fact that they try to mimic the natural evolution of living organisms. GAs are based on the observation that the evolution of natural species is very efficient at adapting to changing environments. By simulating the evolution process, GAs may provide a good way for optimising artificial systems.

GA is a population-based algorithm. Each individual of this population represents a candidate solution to the given problem. Each individual is assigned a fitness value on the basis of a fitness function. In pattern recognition, a common fitness function is the percentage of correctly classified learning patterns. GA operates in three stages: *selection*, *crossover* and *mutation*. In the selection step, the most adapted individuals (which have a high fitness value) are given more chance to be selected for reproduction. During the crossover, portions of the parent solutions are exchanged in the hope of generating more adapted solutions. Mutation operates by randomly changing one or more components of a selected individual. It acts in order to keep some diversity in the currently tested solutions, and prevent premature convergence of the algorithm. GAs have been successfully applied to various scientific and engineering problems. For example, they have been used to model various aspects of the natural immune system (Bersini and Varela 1991), to design and control robots (Davidor 1991), and to predict protein structure (Schulze-Kremer 1992).

The purpose of the present work was to provide better understanding of the limitations and the strengths of the GAs for feature selection in artificial vision. The practical application consisted of the dis-

crimination of seeds according to their species. In seed stations, seed batches cannot be commercialised without passing a quality control test, which consists of the detection of adventitious seeds. Seed analysis is currently a manual task in most seed laboratories, which is time consuming, and is increasingly difficult because of the growing number of registered seed species. For this reason, other techniques have been attempted for automatic seed identification, such as artificial vision (Sapirstein *et al* 1987; Myers and Edsall 1989). The discrimination rates are closely dependant on the seed species. In most of the published studies, stepwise discriminant analysis was applied for feature selection. However, this method makes the assumption that classes are normally distributed in the feature space. This assumption is not usually met in the case of seed discrimination, because seeds belonging to the same species often present wide variations in their morphological characteristics.

In the present work, GA was applied in order to explore the strengths and the limitations of this kind of search technique for feature selection. The performances were tested on a practical problem, which consisted on the discrimination between four seed species.

THEORY AND OPTIMISATION PROCEDURES

The number of features which can be safely introduced in a predictive model is dependant on the number of learning patterns. Only a few features are necessary to correctly describe a data set including a small number of learning patterns. With a limited number of learning patterns, adding new features to a pattern recognition system may degrade rather than increase the performances. Hence, feature selection is necessary to separate adequately qualitative classes with a limited number of learning patterns.

Genetic algorithms

GAs are computational models of evolution. They work on the basis of a set of candidate solutions. Each candidate solution is called a 'chromosome', and the whole set of solutions is called a 'population'. The algorithm allows movement from one population of chromosomes to a new population in an iterative fashion. Each iteration is called a 'generation'. There are various forms of GAs, a simple version, which is called *static population model* (Whitley 1989), was used in all the experiments. In the static population model, the population is ranked according to the fitness value of each chromosome. At each generation, two (and only two) chromosomes are selected as parents for reproduction. The algorithm works as follows.

#1

- Define a representation of the chromosomes for the given problem. Each chromosome is a string of alphabets representing a candidate solution to the problem. Binary chromosomes are commonly used. Some GAs use chromosomes formed from real numbers.
- Choose an objective function to optimise (called the *fitness* function). This function is directly related to the given problem. In pattern recognition, the usual criterion to evaluate the performances is the percentage of correct classification of the learning data.
- Define the parameter values of the GA, such as the size of the population, 'initialisation probability', 'crossover probability' and 'mutation probability'.

#2

Generate an initial population of chromosomes. This initialisation is often achieved at random, but the population may be initialised by chromosomes which are already known to perform well. When random initialisation of binary chromosomes is used, each bit of the chromosomes is randomly set to 0 or 1 according to a probability which is called the *initialisation probability*.

#3

Evaluate the fitness value of each chromosome. The more adapted chromosomes will receive higher fitness values. When the GA is used for function minimisation, a transformation is necessary to derive a function maximisation problem.

#4

Select 2 individuals from the whole population of individuals. The selection is dependant on the value of the fitness function of each individual. The well-adapted individuals have a greater chance at being selected. The two selected chromosomes are called 'parents'.

#5

Eventually crossover is applied on the two selected individuals according to a probability which is called the *crossover probability*. The crossover operation consists of randomly selecting a position in the two selected chromosomes. Each of the two parents are then divided into two segments. The segments of the two selected chromosomes are exchanged in order to create two new chromosomes (two *children*). The second segment of parent 1 becomes the second segment of child 2, and *vice versa*. As an example of crossover, consider the two binary parents:

Parent 1: 1010010111

Parent 2: 1000010010

Suppose the random crossover occurs after the sixth bit. Then each new child receives one segment of bits originated from each parent:

Child 1: 1010010010

Child 2: 1000010111

The crossover is performed in the hope that the combination of two well-adapted chromosomes may give two new better adapted ones. Crossover is the main operation which allows the exploration of the feature space in order to find a near to optimal solution. When all the individuals of the population are identical, the application of the crossover operator on two chromosomes will generate the same chromosomes. This means that the crossover operator is not able to generate diversity within a population. The diversity can be maintained by the mutation operator.

#6

Mutate the two new chromosomes. The operation of mutation consists in randomly altering the value of each element of the chromosome according to a probability which is called the *mutation probability*. This probability usually has a very low value. When binary chromosomes are used, mutation consists in inverting the value of a component (from 0 to 1, or the inverse). Mutation acts as a population perturbation operator and is a means of avoiding premature convergence of the algorithm. The random perturbation of a chromosome may be of great importance when there is a lack of diversity in the population.

#7

Evaluate the fitness of the two created chromosomes, and replace the two less adapted chromosomes, which have the two lowest fitness values in the population, by these two new chromosomes. Go back to #4. The algorithm stops when the maximum number of generations is reached.

Discrimination by the nearest neighbour method

The nearest neighbour method was applied for the discrimination of seeds from their features. This method affects an unknown pattern to the class of its nearest neighbour in the learning set. The main advantage of nearest neighbour classification is that results are directly related to the learning data, and no training time is required. In the present case, the distance between two patterns is assessed by using only the selected features. GA was applied with binary chromosomes. The size of each chromosome was equal to the total number of available features. The value 1 of an element means that the corresponding feature is selected. In order to evaluate the performance of each

chromosome, it was necessary to assess a fitness value. In the present case, the fitness value of each chromosome was defined as the percentage of correct classification of the learning set when using the corresponding selected features. When a learning pattern is to be classified, it is temporarily removed from the learning set. This method is called *leave-one-out nearest neighbour*.

MATERIAL AND METHODS

Sample collection

Samples of four seed species were supplied by the national seed testing station of France (Station Nationale d'Essais de Semences, Beaucauzé, France). Seeds of red fescue (*Festuca rubra* L), perennial rye grass (*Lolium perenne* L), rumex (mixture of three rumex varieties: *Rumex crispus* L, *Rumex longifolius* L, *Rumex obtusifolius* L) and wild oat (mixture of three wild oat varieties: *Avena fatua* L, *Avena pubescens* L and *Avena sterilis* L) were randomly picked from batches. Red fescue and perennial rye grass are cultivated seed species, whilst rumex and wild oat are adventitious seed species. Rumex and wild oat seeds are very devastating for the crop, and current regulations require that they should be rigorously detected and removed from seed batches. Each seed species was represented by a set of 200 seeds.

Image acquisition

Images were acquired via a high-resolution colour camera (KY-F55B, JVC Corp, Japan). A lens was fitted on the camera in order to adjust the magnification (Nikkor AF, Nikon, Japan). The lighting was assured by two 18 W neon lamps placed at each side of the camera. All the images were digitised by a frame grabber

(VP1300-768-E-AT, Imaging Technology Inc, Bedford, USA) to the size of 512 rows and 768 columns. Seeds were placed in random positions and orientations in front of the camera. The distance between the camera and the working surface was 45 cm. The size of the vision field of the camera was 6 cm × 9 cm, and hence the spatial resolution of the images was 0.11 mm.

In this study, a colour image referred to three separate grey scale images each corresponding to a *primary* channel (red, green and blue channels). The number of seeds in a single image depends on the seed species. A single image of red fescue, or perennial rye grass, or rumex contained up to 100 seeds, whilst an image of wild oat seeds included about 20 seeds. The numerical images were acquired and stored for further processing and analysis.

Feature measurement

The images were first binarised by an adaptive technique (Mardia and Hainsworth 1988). The main goal of the binarisation was to separate the objects (seeds) from the background. A wide range of features were then extracted in order to characterise each seed (Tables 1 and 2). Twenty five size and shape features were extracted from the binary image. The textural features were measured independently from each colour channel. Three texture analysis methods were applied: *grey level distribution* method (Ballard and Brown 1982; Press *et al* 1992), *co-occurrence matrices* method (Haralick *et al* 1973), *run length matrices* method (Galloway 1975). The grey level distribution method allows the extraction of simple statistical descriptors from the histogram of the pixel luminance within the region of each seed. In this way, features such as mean grey level, variance and energy were extracted. Co-occurrences and run lengths methods made it possible to extract features to characterise the two-dimensional distribution of the grey level.

TABLE 1

Size and shape features used for seed characterisation. All these features were measured from binary images

Feature	Interpretation
Area	Sum of all pixels in the region of a selected seed
Perimeter	Sum of all pixels on the boundary of a selected seed
Length	Maximum length of the seed through the centroid
Width	
Thinness ratio (or circularity)	
Elongation	Ratio of the width to the length of a seed
First ten magnitude Fourier descriptors	
Seven invariant moments	
Eccentricity	
Spread	

TABLE 2
Textural features used for seed characterisation. Each of these features is a three-dimensional vector

	<i>Feature</i>	<i>Interpretation</i>
Local histogram features	Mean grey level	
	Variance	Gives the variance of the histogram. It is minimised when histogram elements are as equal as possible.
	Energy	
	Entropy	Measures the homogeneity of the histogram.
	Kurtosis	Characterises the degree of asymmetry of the histogram around its mean.
Grey level co-occurrence matrix features	Skewness	Measures the relative peakedness or flatness of the histogram.
	Energy (or Angular Second Moment)	Gives the variance of the matrix. It is minimised when matrix elements are as equal as possible.
	Contrast	Gives the inertia of the matrix according to its main diagonal.
	Correlation	Measures the resemblance between lines (respectively columns).
	Entropy	Gives a measure of the homogeneity of the matrix elements.
Grey level run length matrix features	Inverse different moment	Give high values if elements are concentrated around the main diagonal of the matrix.
	Short-run emphasis	Emphasises the short run existing in the image.
	Long-run emphasis	Emphasises the long run lengths of an image.
	Grey level distribution	Gives high values if frequencies of occurrence of run length are distributed over very few grey levels.
	Run-length distribution	Gives high values if frequencies of occurrence of run length are distributed over very few run lengths.
	Run percentages	Low for homogeneous texture.

Sixteen texture features were extracted from each colour channel, leading to a total of 48 (16×3) texture features. Including the size and shape features, the total number of features was 73. All the features were normalised in order to have an invariance pertaining to the rotation and the translation of the seeds. All the data were gathered into a matrix of 800 rows (learning patterns) and 73 columns.

Genetic algorithm

GA was performed by means of programs written in C++ language. GA was run with 500 generations, the crossover probability was 0.6, the mutation probability was 0.001, and the size of the population was 50. The values of these parameters were fixed after a scanning through the literature. The chosen empirical values of these parameters had proven their efficiency in previous optimisation problems, and are recommended by several authors (Davis 1989; Goldberg 1989). There is no method to automatically optimise the values of the

GA parameters for a given problem, but many authors suggest the use of a high value of crossover probability, and a low value of the mutation probability. The initialisation probability determines the proportion of selected variables of the initial population. As the number of selected variables is an important element of robustness of a given predictive model, the initialisation probability deserved to be studied. For this goal, independent tests were achieved by varying the value of the initialisation probability.

The natural representation of each chromosome was a binary string of length 73 representing the absence or presence of each of the 73 available features. The main advantage of this representation was that it made it possible to use the standard GA operators (binary crossover and binary mutation). Hence, no definition of new GA operators was required. Great care was taken in order to better understand the strengths and the limitations of GAs in feature selection. The convergence properties of this algorithm were studied. Moreover, the number and the kind of selected features were analysed as a function of the number of generations.

RESULTS AND DISCUSSION

Seed appearance

Seed discrimination by artificial vision is based on the physical appearance of the seed species to be discriminated. In the present case, seeds of red fescue and perennial rye grass, which are cultivated seed species, have the same colour features. They both present a yellowish-green colour. Seeds of red fescue are slightly smaller than those of perennial rye grass (Fig 1). These two species are usually mixed in commercial seed lots for lawn cultivation. Visual discrimination between these two species is uneasy, and it might be supposed that their automatic identification would also be difficult. Seeds of rumex had a small size and a dark brown colour and were homogeneous in size and colour, whilst wild oat seeds were heterogeneous in their morphological appearance. Indeed, there were small, brown wild oat seeds as well as large, yellow ones. Some wild oat seeds presented a stalk, although most of these stalks had usually been broken off during transportation.

The histogram of area and width are shown on Figs 2 and 3, respectively. The seeds of wild oat had the largest and the most variable area. Seeds of perennial rye grass had an area which was slightly larger than that of seeds of rumex and red fescue. Seeds of rumex and red fescue had approximately the same area. Areas therefore did not allow the discrimination between rumex and red fescue. Moreover, the width, which represented the length of the minor principal axis of each seed, did not make it possible to discriminate between rumex and perennial rye grass (Fig 3). Seeds of red fescue had the lowest width. The width of wild oat seeds varied from 17 pixels to 45 pixels. In spite of the heterogeneity of width, this feature allowed the discrimination of wild oat seeds from the three other seed species.

Feature selection by genetic algorithm

All the parameters of the GA were fixed to empirical values except the initialisation probability parameter. This parameter operates during the initialisation stage of the population. In the first experiment, the initialisation probability was set to 0.1. This meant that 10% of the elements of each initial chromosome were randomly set to 1. Figure 4 shows the evolution of the proportion of misclassification of the learning set by the leave-one-out nearest neighbour method pertaining to the number of generations (iterations) of the GA. At each generation, the average performance of the corresponding population and those of the best and the worst chromosomes are given. The worst chromosome is the chromosome which gave the highest proportion of misclassification in the population. It could be observed that the proportion of misclassification of the

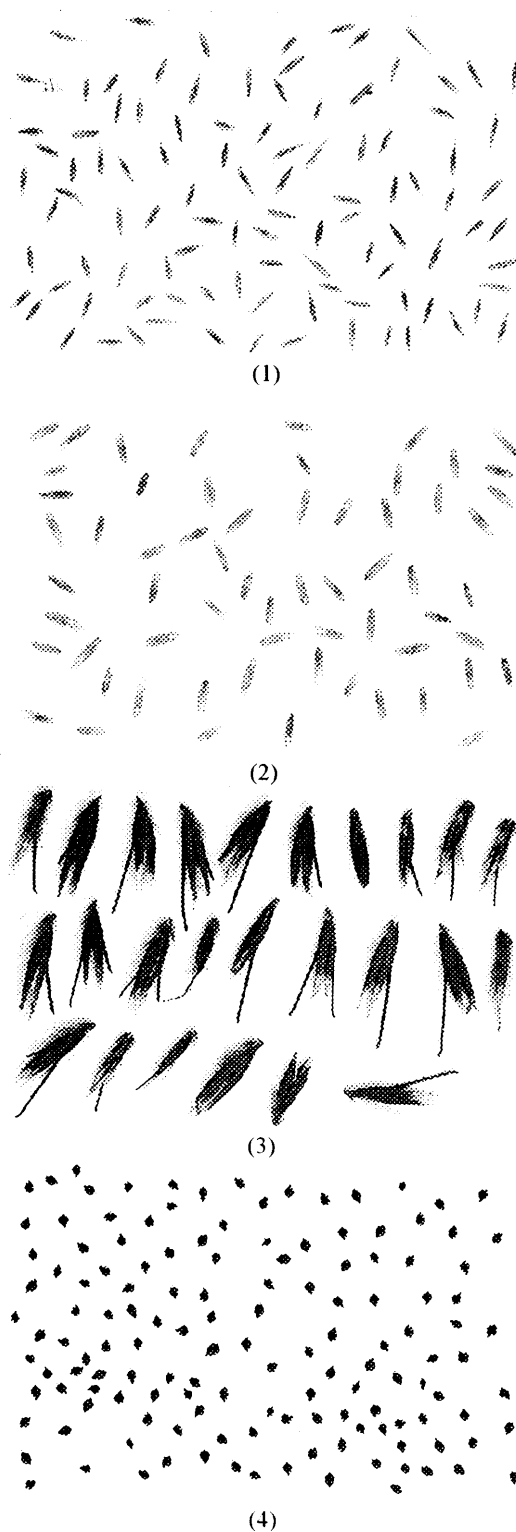


Fig 1. Example of typical images of seeds. (1) Green channel image of red fescue seeds, (2) green channel image of perennial rye grass seeds, (3) green channel image of wild oat seeds and (4) green channel image of rumex seeds.

best chromosome decreased as the number of generations increased. At generation 20, 4.75% of the seeds were misclassified by the best chromosome, whereas at generation 400, the percentage of misclassification was

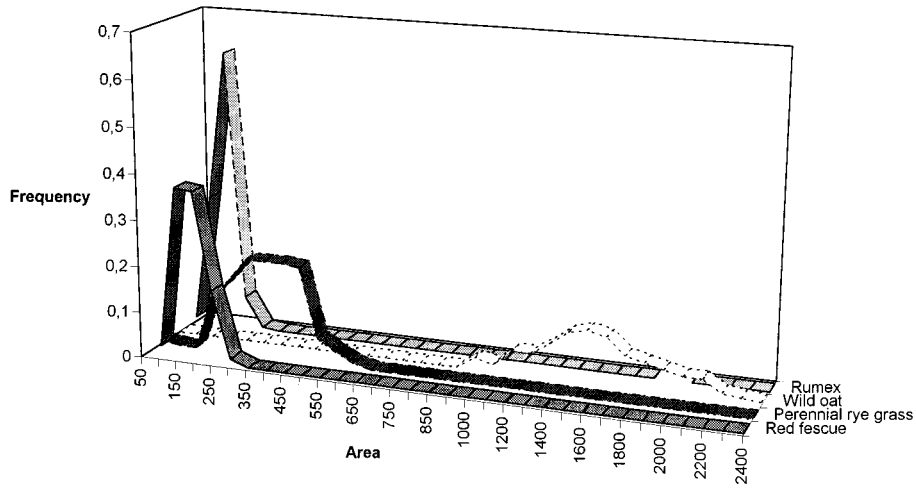


Fig 2. Histogram of the apparent area of seeds.

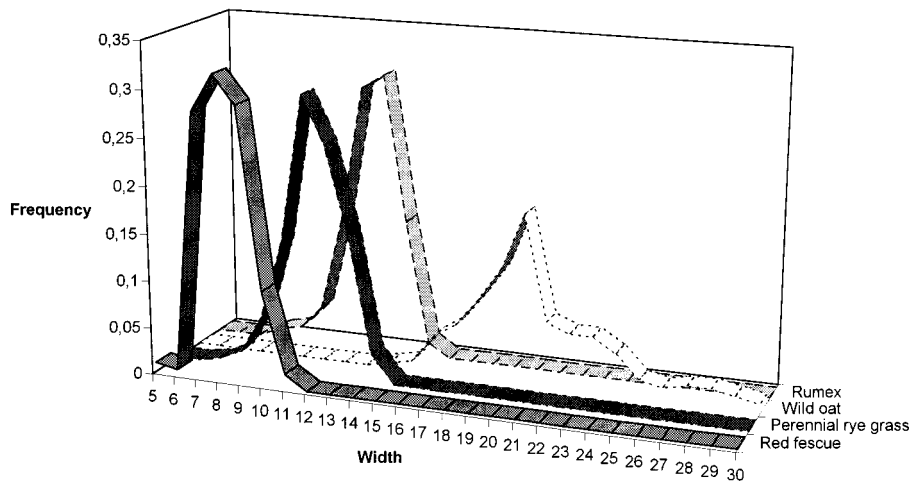


Fig 3. Histogram of the width of all the seeds.

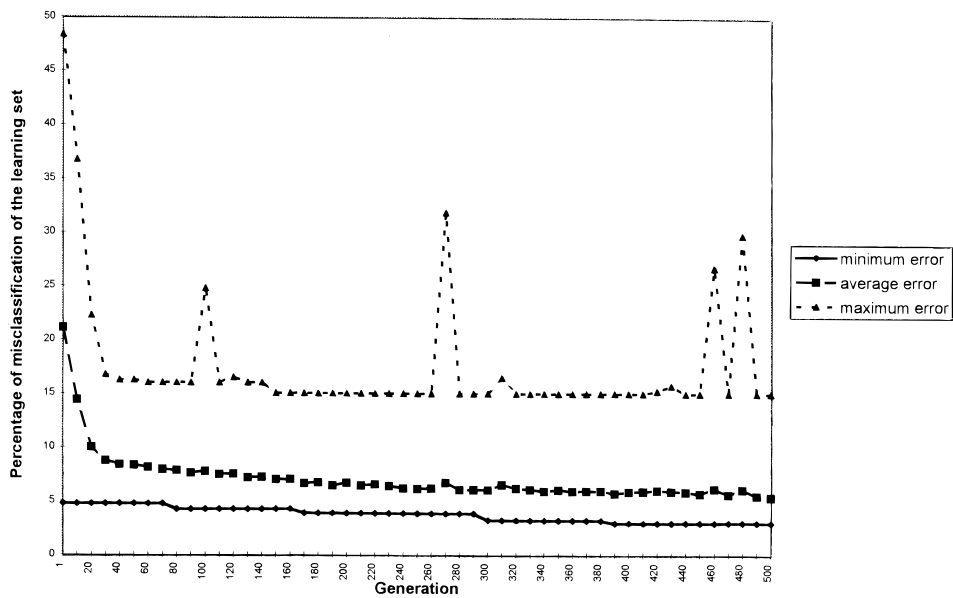


Fig 4. Proportion of misclassification as a function of the number of generations (initialisation probability = 0.1).

equal to 3.00. The average proportion of misclassification of each population slowly decreased as a function of the number of generations. The worst chromosome gave very instable results. For example, it gave a percentage of misclassification of 68.22 for the generation 270 and 15.00 at generation 400. The number of features selected by the best chromosome is represented as a function of the number of generations (Fig 5). The number of selected features by the best chromosome was almost invariant to the number of generations, and had values around 5 (almost 10% of 73). The features which were most often selected were: the width, the entropy in the blue channel of the histogram of each seed, the magnitude of the second and the sixth Fourier

descriptors, and the grey level distribution in the green channel of the run length matrix of each seed. It was observed that GA did not tend to drastically modify the number of selected features, initially determined by the initialisation probability.

The same results were obtained when the initialisation probability was set to 0.5 (Figs 6 and 7). The number of selected features was again almost independent on the number of generations, and took values around 35 (about 50% of 73). Initially, 5.38% of the seeds were misclassified with 34 features selected by the best chromosome. At generation 360, the percentage of misclassification decreased to 5.00 with 31 selected features. There was no obvious relationship between the

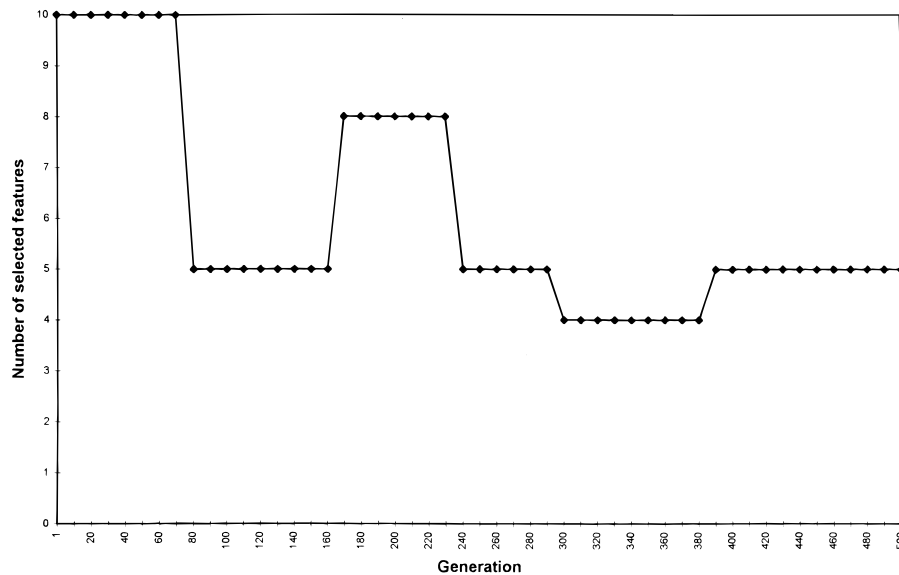


Fig 5. Number of features introduced by the genetic algorithm as a function of the number of generations (initialisation probability = 0.1).

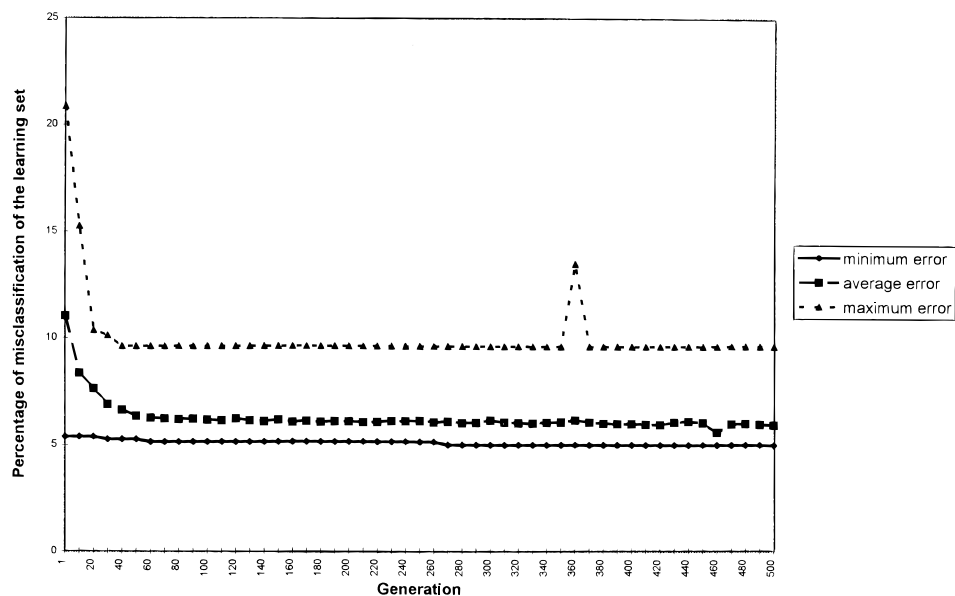


Fig 6. Proportion of misclassification as a function of the number of generations (initialisation probability = 0.5).

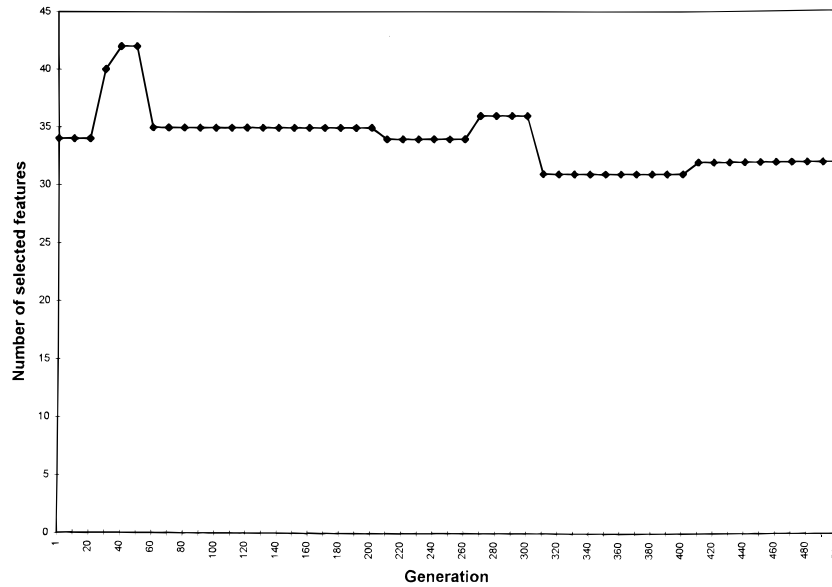


Fig 7. Number of features introduced by the genetic algorithm as a function of the number of generations (initialisation probability = 0.5).

proportion of misclassifications and the number of introduced variables. With only five selected features (obtained from the study with an initialisation probability of 0.1 and generation 400) the results were slightly better with only 3.00% misclassifications. It therefore seems that it would be necessary to process a larger number of generations when the value of the initialisation probability is higher.

GA is based on the selection of the best individuals for reproduction. The chromosomes of the best individ-

uals were supposed to include the most relevant features. In order to study the feature selection during the search process, the percentage of selection of some representative variables in the whole population was analysed. At each generation, the percentage of selection of a variable refers to the ratio of the number of times the corresponding variable was selected in the whole population to the total size of the population (here, 50 chromosomes). At a given iteration, a variable which was selected by a large number of chromosomes has a large percentage of selection. Some variables, such as

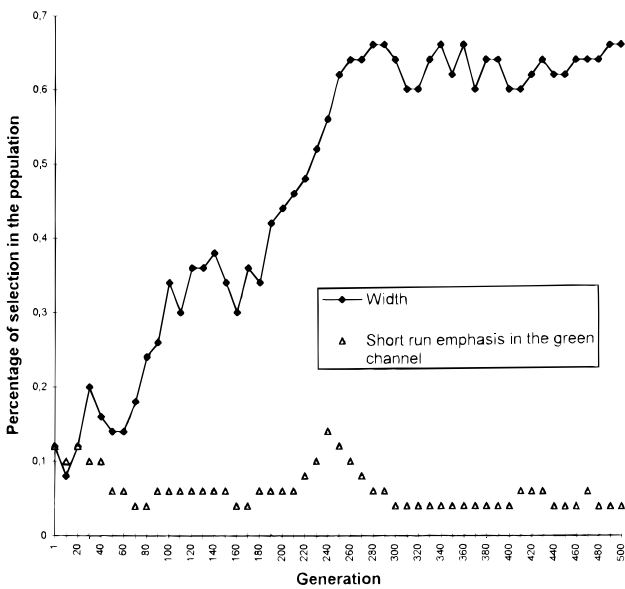


Fig 8. Evolution of the percentage of selection of variables 'width' and 'short run emphasis in the green channel' as a function of the number of generations (initialisation probability = 0.1).

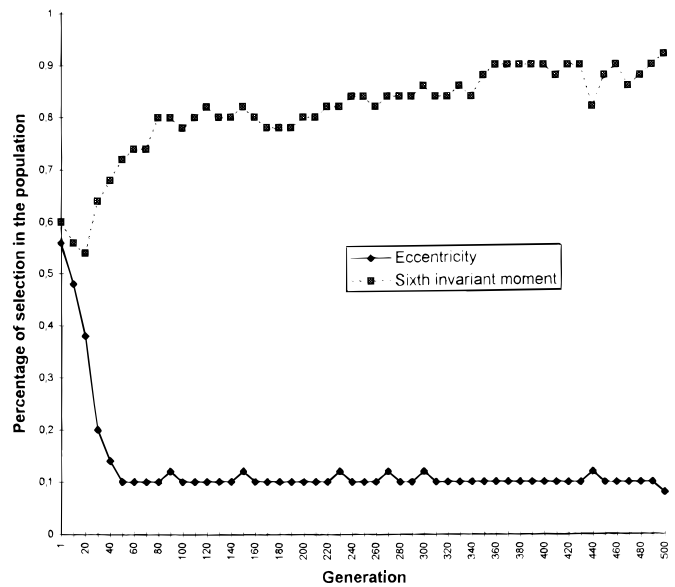


Fig 9. Evolution of the percentage of selection of variables 'eccentricity' and 'sixth invariant moment' as a function of the number of generations (initialisation probability = 0.5).

the width and the magnitude of the sixth Fourier descriptor, which were initially seldom selected (low value of the percentage of selection), became more and more important as the algorithm progressed (Figs 8 and 9). The variable 'width' was initially present in only 6 chromosomes (Fig 8). At generation 500, it was selected by 33 chromosomes. On the contrary, other variables were rejected as the number of generations increased. This was the case, for example, for eccentricity (Fig 9). This means that GA was able to efficiently select or reject features according to their relevance.

CONCLUSIONS

In the present experiment, GA was combined with the nearest neighbour method for the discrimination of seeds by artificial vision. The value to be maximised was the percentage of correct classification by the leave-one-out nearest neighbour method. GA made it possible to select the more relevant features among an initial set of large size. The best results were obtained with an initialisation probability of 0.1, at generation 400, leading to 3.00% of misclassifications between the four seed species. The number of selected features was quite independent on the number of generations of the GA.

Surprisingly, GA did not tend to significantly vary the number of features as a function of the number of generations. It would be interesting to implement a GA which minimises simultaneously two values: the proportion of misclassification and the number of introduced features. However, the standard version of GA cannot be used for the optimisation of more than one attribute. Multiobjective problems can be transformed into single objective problems. This can be achieved by defining a single fitness function which takes into account all the objectives. For example, a linear combination (weighted sum) of the multiple attributes may be used, but the choice of the weights is not trivial. Studies directly dealing with multiobjective optimisation by GAs have been recently published. These works have introduced new concepts in the definition of the basic operators of GA (Mahfoud 1992; Horn and Nafpliotis 1994). In further work we intend to investigate this alternative.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Y Dattée, D Demilly and M Mannino for their help. This research was supported by the National Seed Testing Station of France.

REFERENCES

- Ballard D H, Brown C M 1982 *Computer Vision*. Prentice-Hall Inc, NJ, USA.
- Bersini H, Varela F J 1991 The immune recruitment mechanism: a selective evolutionary strategy. In: *Proc of the Fourth Inter Conf on Genetic Algorithms*, eds Belew R K & Booker L B. Morgan Kaufmann, San Mateo, CA, USA, pp 520–526.
- Davidor Y 1991 *Genetic Algorithms and Robotics*. Robotics and Automated Systems, World Scientific, Singapore.
- Davis L 1989 Adapting operator probabilities in genetic algorithms. In: *Proc of the Third Inter Conf on Genetic Algorithms*. Morgan Kaufmann, La Jolla, CA, USA, pp 60–69.
- Galloway M M 1975 Texture analysis using grey level run lengths. *Comput Graph Image Proc* 4 (2) 172–179.
- Goldberg D E 1989 Sizing populations for serial and parallel genetic algorithms. In: *Proc of the Third Inter Conf on Genetic Algorithms*. Morgan Kaufman Publishing, San Mateo, CA, USA, pp 70–79.
- Haralick R M, Shanmugan K, Dinstein I 1973 Texture features for image classification. *IEEE Trans Syst Man Cyb* 3 (6) 610–621.
- Holland J H 1975 *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, USA.
- Horn J, Nafpliotis N 1994 Multiobjective optimisation using the niched pareto genetic algorithm. *Proc of the First IEEE Conf on Evolutionary Computation, IEEE World Congress on Computational Intelligence* (Vol 1). IEEE, Piscataway, NJ, USA.
- Mahfoud S M 1992 *Crowding and Preselection Revisited*. Parallel Problem Solving from Nature, 2, North Holland.
- Mardia K V, Hainsworth T J 1988 A spatial thresholding method for image segmentation. *IEEE Trans Pattern Machine Intell* 10 (6) 919–927.
- Myers D G, Edsall K J 1989 The application of image processing techniques to the identification of Australian wheat varieties. *Plant Varieties Seeds* 2 109–116.
- Press W H, Teukolsky S A, Vetterling W T, Flannery B P 1992 *Numerical Recipes in C. The Art of Computing* (2nd edn). Cambridge University Press, Cambridge, UK.
- Sapirstein H D, Newman M R, Wright E H, Shwedyk E, Bushuk W 1987 An instrumental system for cereal grain classification using digital image analysis. *J Cereal Sci* 6 3–14.
- Schulze-Kremer S 1992 Genetic algorithms for protein tertiary structure prediction. In: *Parallel Problem Solving from Nature*, eds Männer R & Manderick B. Springer-Verlag, Amsterdam, The Netherlands, pp 391–400.
- Siedlecki R, Sklansky J 1988 On automatic feature selection. *Int J Pattern Recog Artificial Intell* 2 197–220.
- Whitley D 1989 The GENITOR algorithm and selection pressure: why rank-based allocation of reproductive trials is best. *Proc Third Inter Conf on Genetic Algorithms*. Washington, DC, USA.