Weed seeds identification by machine vision

Pablo M. Granitto, Hugo D. Navone, Pablo F. Verdes, H.A. Ceccatto *

Abstract

The implementation of new methods for reliable and fast identification and classification of seeds is of major technical and economical importance in the agricultural industry. As in ocular inspection, the automatic classification of seeds should be based on knowledge of seed size, shape, color and texture. In this work, we assess the discriminating power of these characteristics for the unique identification of seeds of 57 weed species. Using the performance of a naïve Bayes classifier as selection criterion, we identified a nearly optimal set of 12 (six morphological + four color + two textural) seed characteristics to be used as classification parameters. We found that, as expected, size and shape characteristics have larger discriminating power than color and textural ones. However, all these features are required to reach an identification performance acceptable for practical applications. In spite of its simplicity, the naïve Bayes classifier reveals itself surprisingly good for the identification of seed species. This might be due to the careful selection of the feature set, leading to nearly independent parameters as assumed by this method. We also found that, using the same feature set, a more sophisticated classifier based on an artificial neural network committee performs only slightly better than this simple Bayesian approach. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Machine vision; Seed identification; Classification; Neural networks

* Corresponding author. Tel.: + 54-341-4853222; fax: + 54-341-4821772. E-mail address: ceccatto@ifir.edu.ar (H.A. Ceccatto).
1. Introduction

The analysis and classification of seeds are essential activities contributing to the final added value in crop production. These activities are performed at different stages of the global process, including seed production, cereal grading for industrialization or commercialization purposes, during scientific research for improvement of species, etc. For all these purposes, specialized technicians are employed. In most cases, these methods are slow, have low reproducibility, and possess a degree of subjectivity hard to quantify, both in their commercial as well as in their technological implications. It is then of major technical and economical importance to implement new methods for reliable and fast identification and classification of seeds. Like ocular identification work, automatic classification should be based on knowledge of seed size, shape, color and texture (i.e. greytone variations on the surface, see Haralick et al., 1973; Galloway, 1975). Numerous image-processing algorithms are available for extracting these features from seed images, which make machine vision suitable for such a task.

Most previous attempts to identify seeds by machine vision have concentrated on cultivated varieties. Initially it was assumed that varietal differences could be extracted from the structure of the kernel, so different geometrical measurements were used to describe a seed variety (Draper and Travis, 1984; Keefe and Draper, 1986). Other investigations have been conducted to separate different species of cereal grains (Sapirstein et al., 1987; Chen et al., 1989), wheat from non-wheat components (weed seeds and stones) (Zayas et al., 1989), different types of wheat (Symons and Fulcher, 1988; Zayas et al., 1986; Neuman et al., 1989a,b) and special grading classes (Neuman et al., 1987, 1989b), etc. In these studies image analysis was essentially restricted to basic geometrical measurements to obtain different parameters (shape factor, aspect ratio, length/area, etc.). In addition, color was successfully used to separate red-, amber- and white-colored wheat, but could not separate them into grading classes. More recent studies have used color images to establish seed quality and hardseededness of some annual pasture legumes (Jansen, 1995), to characterize fungal damage, viral diseases and immature soybean seeds (Ahmad et al., 1999), etc.

Besides varietal identification and cereal grain grading, early identification of weeds from the analysis of strange seeds is also of major interest in the agricultural industry. This can be done for the purpose of chemically controlling weed growth or, as occurs in many countries, it can be routinely performed as part of official requirements before a seed batch can be made commercially available (purity analysis). Weed seeds are also identified by seed testing stations and seed corporations to measure the purity of the harvest, and by research stations to detect changes in seed banks in the soil. Automatic identification of seeds of wild species is different from the identification of seeds of varieties of a single species. To be approved as a variety, the cultivated plants have to be homogeneous with respect to certain plant characters. Wild species, on the contrary, tend to have larger intra-species variations. Moreover, the variation between weed species will be in general larger, but seeds of some closely related species can be very similar. From
the color point of view, most weed seeds are light to dark brownish or black. All these characteristics make the automatic identification of weed seeds a priori a difficult classification problem. Consequently, a successful approach should include parameters associated with all the relevant characteristics of size, shape, color and texture mentioned above.

An early attempt to identify weed seeds (Petersen and Krutz, 1992) showed the importance of using color instead of black and white images to improve classification accuracy. More recently, Chtioui et al. (1996) compared the capabilities of linear discriminant analysis and artificial neural networks (ANNs) to identify weed seeds from morphological and textural parameters. However, these investigations considered only four different species, which does not provide a good characterization of inter-species seed variations.

In this work, we assess the discriminating power of different seed characteristics for the unique identification of seeds of weed species. We use a simple Bayesian approach (naïve Bayes classifier) to evaluate morphological, color and textural characteristics measured from video images, establishing their importance as classification features for weed seeds identification. The final classifier based on the optimal set of features shows a remarkable good performance. In addition, we present classification results obtained using the same feature set as input of a committee of ANNs. Although the classification parameters selected using the naïve Bayes classifier are probably not the best suited for the ANN approach, these results provide a preliminary comparison of both methodologies. These studies were conducted on a much larger basis than previous ones (Petersen and Krutz, 1992; Chtioui et al., 1996), including seed images of frequent weeds found in Argentina’s commercial seed production industry. To avoid introducing a bias in the selection of species to be considered, we restricted ourselves to the 58 species listed by the Secretary of Agriculture as prohibited and primary- and secondary-tolerated weeds. From this list we finally considered 57 species for which a good number (~40) of young exemplars were available in the seed bank of the Seed Analysis Laboratory at the Oliveros Experimental Station of the National Institute for Agricultural Technology (INTA).

Argentina’s law regulations require the analysis by registered laboratories of a small batch sample before a seed batch can be made commercially available. In these analyses, commercial and strange seeds present are separated, and the latter ones identified one by one. The studies in the present work are part of a development to avoid the continuous training of new technicians to perform this task, providing an automatic classifier that can be used by less skilled operators. Another possible application is the identification of very strange seeds in botanical gardens, although this would require a very large database.

This work is organized as follows. First, we describe the hardware and the experimental conditions used to capture seed images. Then, we define the morphological, color and textural parameters measured from these images, and discuss the selection of the most relevant ones for identification purposes. We present next the results obtained with the naïve Bayes classifier, and compare its performance with that of the ANN committee. Finally, in the last section we draw some conclusions.
2. Materials and methods

2.1. Image acquisition

We have built a database containing 3163 images of the 57 species considered (a list of these species is available on request). To acquire the images we used a 2/3 in. CCD video camera (XC-711P, Sony Corp, Japan) connected to a color frame grabber (IC-PCI, Imaging Technology Inc, USA) with 8-bit look-up tables per color channel. Illumination was provided by a 150 W light source (Fostec Inc, USA) through a quadruple fiber optic bundle of 12.7 mm diameter, with the four guides in a symmetric arrangement to produce an even illumination with good texture enhancement.

Images were taken with a 768 × 512 pixel resolution on a blue background, which can be easily subtracted by standard segmentation routines because of the color difference versus the seeds (Petersen and Krutz, 1992). The segmented images consist of arrays whose entries are 24-bit records, corresponding to the 256 pixel intensity levels (8 bits) for each of the red (R), green (G) and blue (B) channels. In order to reduce effects associated with illumination changes, we also considered the normalized red \( r = R/I \) and green \( g = G/I \) pixel values, where \( I = (R + G + B)/3 \) is the average intensity.

Regardless of seed size, all images were taken to approximately fill the camera field of view by adjusting a 6.5 × parfocal zoom (6000 System, Navitar Inc, USA) with 0.5 and 2 × lens attachments. This is necessary given the large differences in seed sizes considered—from 0.2 to 15 mm, approximately—since, otherwise, the images of the smallest seeds would have shown very little texture details. The calibration of the entire zoom range was performed at the beginning of the database acquisition, and the computer automatically sets the correct length scale according to the zoom level (we used ten steps). Light intensity was regulated with an iris diaphragm in order to adjust the illumination to the changing field of view while keeping a constant color temperature (corresponding to a standard 20 V–150 W halogen projector lamp (Ushio Inc, Japan)).

The adjustment of the light intensity to have a good image definition was performed manually; a better control of illumination conditions (for instance, keeping the background color constant by an electronic control of the light source) would have enhanced the classification capabilities of color and texture parameters. This could be required for a commercial system, but the experimental procedure just described was considered sufficient for the purposes of the present work. Notice also that, for the kind of application we have in mind (purity analysis of seed batches), the re-adjustments of zoom and light-intensity levels above described is not a serious inconvenience.

For the sake of illustration, in Fig. 1 we show images of eight different species selected to exemplify how similar/dissimilar shapes and textures can be. These images have been taken with different magnifications (as described in the figure caption) to have good detail of textures. Color is not particularly important for classification of these seeds since most of them are medium to dark brownish.
2.2. Feature extraction

We measured a number of features from the raw seed images to be later used for classification purposes. As stated above, these features correspond to morphological, color and textural characteristics of the seeds. We briefly describe below the different parameters considered.

2.2.1. Morphology

Size and shape characteristics of seeds can be easily obtained from binarized images. In particular, we measured the lengths of the principal axes and several moments of the planar mass distribution with respect to those axes, the size of the
minimal rectangular box containing the seed and the ratio of its area to the seed
area (compactness), etc. All these quantities were made dimensionless by conve-
niently normalizing them by the required powers of the square root of the seed area
(which was taken as the only dimensional quantity). Furthermore, since we used the
principal axes as a reference frame for all measurements, the resulting values are
independent of image orientation. In total, we measured 21 morphological features.

2.2.2. Color
We calculated gray level histograms in the \( I, r, g \) channels. From these histograms
we measured standard features such as average, variance and skewness. In addition,
we considered three ratios of average histogram values in the RGB channels:
\( E[R]/E[G] \), \( E[R]/E[I] \) and \( E[G]/E[I] \) (where \( E[.] \) means the average pixel value in
the corresponding channel). We measured in total 12 different color characteristics.

2.2.3. Texture
Following Petersen and Krutz (1992), two different matrices were used to
describe seed surface texture.
1. Gray level co-occurrence matrix: a two-dimensional matrix with entries \( A_{ij} \),
where \( i, j \) are gray levels and the entry value gives the number of nearest-neighbor
pixels in the image having these gray levels along a given direction (we used
alternatively both principal axes directions to have rotational invariant textural
features). In practice, we considered a coarse-grained version of this two-dimen-
sional histogram. First, we performed a dynamic equalization of the gray level
histogram on each color channel using 16 bins in order to eliminate illumination
intensity variations (Haralick et al., 1973). Then, the indices \( i, j \) were made to
correspond to these bin levels. From the resulting \( 16 \times 16 \) matrix 17 textural
features were obtained. The precise definition of these parameters and the
interpretation of their discriminating properties can be found in Haralick et al.
2. Gray level run length matrix: the two dimensions in this matrix are the gray
level and the so-called run length, i.e. the base 2 logarithm of the number of
adjacent pixels in a given direction with the same gray level. We have considered
both principal axes directions to compute the run lengths. In this case the matrix
dimension was reduced by taking the same 16 gray level intervals used before.
The resulting matrix measures four new textural features. The precise definition
of these parameters and the interpretation of their discriminating properties can
be found in Galloway (1975).
In total we considered 42 textural characteristics, corresponding to the 17
features measured from the gray level co-occurrence matrices plus the four features
obtained from the gray level run length matrices in both principal axes directions.

2.2.4. Feature selection
Thus, from each color image we measured 75 features to be used for classifica-
tion. By simple statistical analysis (comparison of intra-species fluctuations with
inter-species changes) we determined that several of them could be discarded as
classification parameters. Doing so, we retained 15 morphological, eight color and 17 textural properties. Of course, this large set of features still contained redundant, noisy or even irrelevant information for classification purposes. To choose the best features in each group (those with the largest discriminating power), we implemented standard sequential forward and backward selection algorithms (Jain and Zongker, 1997), using the performance of a naïve Bayes classifier as the selection criterion. The naïve Bayes classifier fits the class conditional probabilities with a product of normal distributions of the individual features and, in spite of its simplicity, it performs well on this problem (see next section). The selection algorithms reduced the parameters to nearly optimal sets of ten morphological, seven color and seven textural features. The same procedure applied to these 24 remaining parameters selected 12 (six morphological, four color and two textural) features, which were finally used to build the classifiers. A list of these final parameters is given below.

### 2.2.5. Classification parameters

The final 12 parameters selected for classification are the following:

- **Morphology and size** (see Fig. 2).
  - Square root of seed area \([\text{SQRT}(A)]\).
  - Ratio of semi-axis lengths of the main principal axis \([h_1/h_2]\).
  - Ratio of seed and enclosing box areas \([A/(h_1 + h_2) \times (v_1 + v_2)]\).
  - Moments of the planar mass distribution with respect to the principal axes \([M_{20} , M_{21} , M_{22}]\).

- **Color**.
  - Variance of the intensity histogram \([M_2(I)]\).
  - Skewness of the intensity histogram \([M_3(I)/M_2(I)^{3/2}]\).
  - Ratios of average pixel values in RGB channels \([E(R)/E(I), E(G)/E(I)]\).

- **Texture**.
  - Contrast (Haralick et al., 1973) along the main principal axis direction.
  - Cluster Prominence (Conners et al., 1984) along the secondary principal axis direction.
3. Results and discussion

3.1. Discriminating power

To compare the discriminating power of the different sets of features, in Table 1 we present the generalization capabilities of naïve Bayes classifiers built solely in terms of either the ten morphological, seven color or seven textural features. Here, and in all the subsequent tables, the results correspond to an average over 30 independent experiments as described next. In each experiment, we split the 3163 images of the 57 species considered in training and test sets, randomly choosing, for each species, 80% of the images to build the classifier and including the remaining 20% in the test set. This leaves 2530 images for training and 633 images for testing the system. Table 1 gives the average performances and standard deviations (S.D.) over the 30 experiments for both the training and test sets. It also shows how performance increases when the system assigns a test image to any of the \( n \) most probable classes. The different cases considered are indicated in the table as first option \( (n = 1, \) standard classification), First two options \( (n = 2) \) and first three options \( (n = 3) \). Notice that for \( n > 1 \) the classification is considered as correct if the test image corresponds to any of the \( n \) classes with the largest probabilities. This possibility is very useful in practice, since untrained operators can easily select the correct option by simple visual comparison with stored representative seed images of the \( n \) classes suggested by the classifier.

A quick look at Table 1 shows, as expected, the large discriminating power of morphological features. As anticipated, color is not particularly good because many species are light to dark brownish or black, and texture characteristics are even less reliable as classification parameters. Furthermore, if we combine any two sets of features (see Table 2), morphology plus color features have an edge over the combined use of morphology and texture characteristics. Notice, however, that in this last case it would be enough to consider black and white images, which constitutes an important simplification and a reduction in hardware cost. Finally, the performance of the naïve Bayes classifier built in terms of the optimal set of 12 features listed above are given in Table 3.

Table 1
Naïve Bayes classifier performances as percentage of correct seed identifications using only one particular set of features at a time

<table>
<thead>
<tr>
<th>Features</th>
<th>First option</th>
<th>First two options</th>
<th>First three options</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Test</td>
<td>Training</td>
</tr>
<tr>
<td>Morphology</td>
<td>86.7 ± 0.3</td>
<td>95.8 ± 0.5</td>
<td>98.2 ± 0.1</td>
</tr>
<tr>
<td>Color</td>
<td>60.2 ± 0.7</td>
<td>73.6 ± 1.5</td>
<td>81.6 ± 0.4</td>
</tr>
<tr>
<td>Texture</td>
<td>55.3 ± 0.5</td>
<td>69.3 ± 1.3</td>
<td>76.9 ± 0.5</td>
</tr>
<tr>
<td></td>
<td>84.5 ± 1.1</td>
<td>94.4 ± 0.9</td>
<td>66.7 ± 2.2</td>
</tr>
</tbody>
</table>

Mean values and standard deviations are estimated from 30 independent experiments, as described in the main text.
Table 2
Naïve Bayes classifier performances as percentage of correct seed identifications using different combination of two sets of features

<table>
<thead>
<tr>
<th>Features</th>
<th>First option</th>
<th>First two options</th>
<th>First three options</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Test</td>
<td>Training</td>
</tr>
<tr>
<td>Morphology + Color</td>
<td>96.6 ± 0.2</td>
<td>94.8 ± 0.8</td>
<td>99.1 ± 0.1</td>
</tr>
<tr>
<td>Morphology + Texture</td>
<td>91.7 ± 0.3</td>
<td>89.1 ± 0.9</td>
<td>97.6 ± 0.3</td>
</tr>
<tr>
<td>Color + Texture</td>
<td>83.2 ± 0.3</td>
<td>80.4 ± 1.3</td>
<td>91.0 ± 0.7</td>
</tr>
</tbody>
</table>

Mean values and standard deviations are estimated from 30 independent experiments, as described in the main text.

It is important to stress that differences in performance scores in Tables 1 and 2 are significant in all cases, with at least 29 out of the 30 experiments performed behaving consistently with these differences. For instance, color features have larger discriminating power than textural ones in 29 out of 30 experiments [remember that ideally 22/30 would correspond to a 99% confidence level, although there are risks in applying the paired-difference t-test to different random train-test splits of the data (Dietterich, 1996)].

3.2. Neural network classifier

For comparison, we also developed a classifier based on ANNs (Bishop, 1995) using the same feature set selected in the Bayesian approach. To this end, we trained feedforward networks with 12 input, \( h \) hidden, and 57 output units. The numbers of input and output units correspond to the number of parameters used and seed species to be identified, respectively. The number of hidden units was varied from \( h = 20 \) to 80, monitoring the performance on cross-validation samples set aside from the training data; the results presented below correspond to \( h = 40 \)

Table 3
Performances of different classifiers as percentage of correct seed identifications using the optimal set of 12 features

<table>
<thead>
<tr>
<th>Classifier</th>
<th>First option</th>
<th>First two options</th>
<th>First three options</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Test</td>
<td>Training</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>97.5 ± 0.2</td>
<td>95.8 ± 0.9</td>
<td>99.4 ± 0.1</td>
</tr>
<tr>
<td>Single ANN</td>
<td>100</td>
<td>95.6 ± 0.8</td>
<td>100</td>
</tr>
<tr>
<td>Committee</td>
<td>MR</td>
<td>96.6 ± 0.7</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>AP</td>
<td>96.7 ± 0.7</td>
<td>100</td>
</tr>
</tbody>
</table>

Mean values and standard deviations are estimated from 30 independent experiments, as described in the main text.
units, which lead to the smallest classification error on these samples. We employed output units with softmax (normalized exponential) activation functions to allow the interpretation of outputs as class probabilities. Furthermore, a cross-entropy error measure was used, which is the standard choice for classification problems. We trained the ANNs with the usual backpropagation rule until convergence, since only negligible overfitting problems were observed. This avoided the use of part of the training set for validation purposes (except for the initial selection of the optimal number of hidden units).

The performance of a single (generic) ANN and the results obtained by structuring ten networks in a committee are shown in Table 3. In the case of the ANN committee we considered two options, (i) each network votes for the class with the largest probability according to its own outputs, and the image is finally assigned to the class with the majority of votes (Majority Rule, MR); and (ii) the class probabilities output by the ten networks are added and the image is assigned to the class with the largest sum value (Added Probabilities, AP). Again in this case, all the results quoted correspond to an average over 30 independent realizations of the whole procedure.

A complete comparative description of the different methods’ performances is given in Table 4, where the entry \((i,j)\) gives the number of experiments in which the method in row \(i\) produced better results (at first option) than the method in column \(j\). Using standard paired \(t\)-tests with the proviso mentioned above concerning random splits of the data, these figures show that the two ANN committee implementations are better than the naïve Bayes and single ANN classifier with more than a 99% confidence level. Moreover, the strategy of adding probabilities in the committee is better than the majority-rule vote with a confidence level also above 99%. Notice that, for simplicity, feature selection was performed using the naïve Bayes classifier, which may not necessarily produce an optimal set for the ANN approach.

Several comments are in order at this point. First, we stress the excellent performance of the naïve Bayes classifier, which might be related to an effective near independence of the selected classification parameters. Second, since in the ANN approach the performance of a single network is already very good, there is little room left for improvement by adding several predictors in a committee.
Notice that when the system is allowed to suggest three options for class membership, from the 633 images in the test set only five images are misclassified by the naïve Bayes classifier and four images by the single ANN (for both methods the performance reaches 100% with five options). Of course, for a much larger number of species the classification problem would be more demanding and the ANN committee might have a substantial advantage over other simpler methods. Furthermore, feature selection should be performed using this classifier as selection criterion to obtain an optimal set for the ANN approach. In passing we mention that there are more sophisticated feature selection methods that the one implemented in this work (Jain and Zongker, 1997; Chtioui et al., 1998). Finally, we stress the fact that different realizations of training and test sets do not substantially change performance (Tables 1–3), as indicated by the low S.D.s observed in the 30 independent runs.

4. Conclusions

We performed an analysis of the discriminating power of different characteristics of weed seeds measured from color images, using the performance of a naïve Bayes classifier as evaluation criterion. First, a careful selection of the features best suited for classification purposes reduced these characteristics to a nearly optimal set of 12 parameters—six morphological, four color and two textural properties, listed in the Section 2. Then, we established the relative importance of different kind of features in the identification of weed seeds. As expected, size and shape are the principal characteristics for seed identification (see Table 1), although color and texture must be also considered to obtain a good performance. These last properties have approximately the same discriminating power when considered independently of morphology.

The implementation of a naïve Bayes classifier on the basis of the selected parameters produced excellent results, as shown in Table 3. This is in part due to the careful parameter selection, which lead to a set that approximately fulfills the independence of classification features assumed by the naïve Bayes approach. Table 3 also gives the performances of classifiers based on ANNs for comparison. In particular, the best ANN committee correctly identified 99.5% of the test images within the three most probable classes for each case. Furthermore, we determined with a 99% confidence level that this ANN committee performs better than the simple Bayesian approach.

For the number of species considered in this study, the preprocessing of images and the careful selection of measured features reduced considerably the complexity of the classification problem. However, one might expect that this problem will become more demanding for databases containing several hundreds of species, as required for a commercial system. In such a case, several important improvements on classifier development can be implemented. As already mentioned, feature selection for the ANN classifier should be performed using the ANN as the selection criterion, which would lead to the set best suited for this approach. In
addition, implementation of optimal ensemble techniques instead of a simple committee could take advantage of the ANN variance (nonidentifiability of the model) on complex problems (Navone et al., 2000). Work in this direction requires the lengthy acquisition of an extended database, which is currently in progress.

**Acknowledgements**

We acknowledge the constant assistance of Eng. Roque Craviotto and technicians of the Seed Analysis Laboratory at EEA Oliveros of INTA. This project was partially financed through grant PICT 11-03834 from ANPCyT.

**References**


