Large-Scale Investigation of Weed Seed Identification by Machine Vision

Pablo M. Granitto, Pablo F. Verdes, H. Alejandro Ceccatto *

Instituto de Física Rosario, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) and Universidad Nacional de Rosario (UNR), Boulevard 27 de Febrero 210 Bis, 2000 Rosario, Argentina.

Abstract

We explore the feasibility of implementing fast and reliable computer-based systems for the automatic identification of weed seeds from color and black and white images. Seeds size, shape, color and texture characteristics are obtained by standard image-processing techniques, and their discriminating power as classification features is assessed. These investigations are performed on a database much larger than those used in previous studies, containing 10310 images of 236 different weed species. We consider the implementation of a simple Bayesian approach (naïve Bayes classifier) and (single and bagged) artificial neural network systems for seed identification. Our results indicate that the naïve Bayes classifier based on an adequately selected set of classification features has an excellent performance, competitive with that of the comparatively more sophisticated neural network approach. In addition, we discuss the possibility of using only morphological and textural characteristics as classification features, which would reduce the operational complexity and hardware cost of a commercial system since they can be obtained from black and white images. We find that, under particular operational conditions, this would result in a relatively small loss in performance when compared to the implementation based on color images.

Key words: machine vision, seed identification, classification, artificial neural networks

* Corresponding author. Tel: +54-341-4853222; Fax: +54-341-4821772.
Email address: ceccatto@ifir.edu.ar (H. Alejandro Ceccatto).

1 Introduction

The process of manual identification of seeds by specialized technicians is slow and possess a degree of subjectivity hard to quantify, both in its commercial as well as in its technological implications. It is then of major technical and economical importance to implement computer-based methods for reliable and fast identification and classification of seeds. Automatic systems can be based on seed images, from which classification features associated to seed size, shape, color and texture (i.e., greytone variations on the surface) are readily obtained. Thus, the field of machine vision, *i.e.*, image-processing algorithms complemented with classification methods, seems a suitable framework for automatic seed identification.

Most previous attempts to identify seeds by machine vision have concentrated on cultivated varieties (Draper and Travis, 1984; Keefe and Draper, 1986) (Zayas et al., 1986; Sapirstein et al., 1987; Chen et al., 1989; Symons and Fulcher, 1988; Zayas et al., 1989; Neuman et al., 1987, 1989a,b), with image analysis essentially restricted to basic geometrical measurements (shape factor, aspect ratio, length/area, etc.). More recently, color images were also successfully used to establish seed quality and characterize damages and diseases (Jansen, 1995; Ahmad et al., 1999). 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). 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.

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 a previous work (Granitto et al., 2002) we have assessed the discriminating power of different seed characteristics for the unique identification of seeds of several weed species. We used a simple Bayesian approach to evaluate morphological, color and textural characteristics measured from video images, establishing their importance as classification features for weed seeds identification. In addition, we presented classification results obtained using the same feature set as input of a committee of ANNs. These studies were conducted on a much larger basis than previous ones, (Petersen and Krutz, 1992; Chtioui et al., 1996) including seed images of 57 frequent weeds found in Argentina's commercial seed production industry. These species are those listed by Argentina's Secretary of Agriculture as prohibited and primary- and secondary-tolerated weeds.

Regulations currently enforced in Argentina require the analysis of a small sample out of a seed batch before it can be made commercially available. In these analysis, strange seeds are separated from commercial ones and identified one by one. The studies in Granitto et al. (2002) were part of a development to avoid the continuous training of new technicians to perform this task, providing an automatic classifier that could be used by less skilled operators. In spite of the good performances of the classifiers developed in that work, the number of species considered was still too small to draw definite conclusions on the viability of our approach. Here we present a more extensive investigation of this problem by considering a much larger database, consisting of 10310 seed images of 236 common weed species. This is not yet enough for a commercial system, which would require a recognition capability of at least twice this number of species. However, in comparison with our previous investigation (Granitto et al., 2002) considering 57 different species, the current database is already large enough to reassess the feasibility and limitations of the proposed development.

This work is organized as follows. First, we give, for completeness, a brief description of the hardware used to capture the seed images and list the more relevant morphological, color and textural parameters for identification purposes. More details on image acquisition and a discussion on how these parameters were selected are given in Granitto et al. (2002). We present next the results obtained with naïve Bayes (Mitchell, 1997) and ANN classifiers and compare their performances. In addition, we investigate their capabilities for seed identification without using color features, a possibility of interest since black and white images are easier to process and the required hardware is much cheaper. Finally, in the last section we summarize our work and draw some conclusions.

2 Image Acquisition and Classification Features

2.1 Hardware

The database contains 10310 images of 236 different species (a list of these species is available on request). Images with 768×512 pixel resolution were obtained using a 2/3" 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 for red (R), green (G) and blue (B) channels. Illumination was provided by a 150W light source (Fostec Inc, USA) with a standard 20V-150W halogen projector lamp (Ushio Inc, Japan)), through a quadruple fiber optic bundle of 12.7mm diameter. All images were taken to approximately fill the camera field of view by adjusting a 6.5X parfocal zoom (6000 System, Navitar Inc, USA) with 0.5X and 2X lens attachments. This is necessary given the large differences in seed sizes considered –from 0.2mm to 15mm, approximately– since, otherwise, the images of the smallest seeds would have shown very little texture details.

2.2 Features

We initially measured 75 features from the raw seed images to be later used for classification purposes. They can be grouped in morphological, textural and color characteristics, and correspond to common measures used in previous studies in the literature (Granitto et al., 2002). From these studies we expect *a priori* a large redundancy among features in each group; consequently, to choose the features with the largest discriminating power we implemented standard sequential forward and backward selection algorithms (Jain and Zongker, 1997). As selection criterion we considered the performance of a naïve Bayes classifier (Mitchell, 1997), using normal distributions to fit the class-conditional probabilities. This selection reduced the parameters within each group to nearly optimal sets of 10 morphological, 7 color and 7 textural features. The same procedure applied to these 24 remaining parameters selected 12 (6 morphological, 4 color and 2 textural) features, which were finally used to build the classifiers. A list of these final parameters is given below. Notice that we did not attempt a single-step parameter selection from the 75 initial measures to the final 12 features since it would have been too costly computationally.

- Morphology and size (see Fig. 1)
  - · Square root of seed area ($\sqrt{A}$)
  - · 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 $\eta_{20}, \eta_{21}, \eta_{22}$ (Chtioui et al., 1996) of the planar mass distribution with respect to the principal axes
- Color
  · Variance of the intensity histogram $[\eta_2(I)]$
  · Skewness of the intensity histogram $[\eta_3(I)/\eta_2^{3/2}(I)]$
  · 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

All morphological quantities were made dimensionless by conveniently 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. For color features, we considered the intensity $I = (R + G + B)/3$ and $E[\cdot]$ in the above expressions means mean pixel value. Texture parameters were obtained from gray level co-occurrence and gray-level run-length matrices, and their precise definitions can be found in Haralick et al. (1973) and Conners et al. (1984).

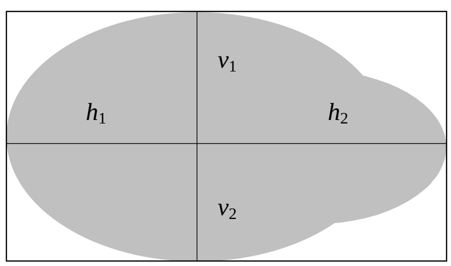

Fig. 1. Definition of quantities related to seed shape used to compute morphological features.

3 Classification Results

3.1 Color images

Following the previous experience in Granitto et al. (2002), we implemented a simple Bayesian approach to the classification problem (naïve Bayes classifier) and compared its performance with that of the more sophisticated ANN technique (see Mitchell, 1997, for an introduction to both methodologies). The naïve Bayes classifier fits the class conditional probabilities with a product of normal distributions of the individual features, considered as independent classification parameters. For the ANN approach we trained feedforward networks with 12 input, h hidden, and 236 output units. The number of input and output units correspond, respectively, to the number of parameters used and seed species to be identified. The number of hidden units h was selected by a "trial and error" approach. For this, h was varied from 20 to 100, monitoring the performance on cross-validation samples set aside from the training data. The results presented below correspond to $h = 80$ 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 (Bishop, 1995).

Both for the Bayesian and ANN approaches we split the 10310 images of the 236 species considered in training and test sets. For this, we randomly chose, for each species, 80% of the images to build the classifier and the remaining 20% for testing purposes. This left a large database with 8250 images for training and also a fairly large test set with 2060 images. The ANNs were trained 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). In Granitto et al. (2002) we found that there is little gain over the performance of a single ANN by aggregating several of them in a simple committee. Here we have explored the slightly more sophisticated "bagging" approach (Breiman, 1996) to build a composite ANN classifier. According to this aggregation method, several ANN are trained on bootstrap re-samples of the training data and their individual predictions are then averaged to produce the final classification (see Breiman, 1996, for details). All the results quoted below correspond to an average over 30 independent experiments.

Table 1 gives average performances and standard deviations for the test sets. It also shows how performance increases when the system assigns a test image to any of the n most probable classes, starting from $n = 1$ (standard classification) to $n = 5$. That is, for $n > 1$ the classification is considered as correct if the test image corresponds to any of the n classes with the largest probabilities output by the classifier. 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. We give the average performances of single classifiers ("Single") and those obtained by ensembling 100 classifiers according to the bagging technique ("Bagging"). In this last case, classifiers were trained on bootstrap re-samples of the training set, which should give them some diversity. Then, the class probabilities output by the 100 ensemble members were added and the image was assigned to the class with the largest sum value.

Table 1
Performances of Naïve Bayes (NB) and ANN classifiers as percentage of correct seed identifications using the optimal set of 12 features. "Single" refers to the average performance of a single classifier and "Bagging" to the performance of the ensemble obtained by aggregating 100 classifiers according to the bagging technique. Mean values and standard deviations are estimated from 30 independent experiments, as described in the main text.

Classifier	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
NB Single	92.4 ± 0.4	97.0 ± 0.3	98.4 ± 0.2	98.9 ± 0.1	99.1 ± 0.1
NB Bagging	92.6 ± 0.4	97.3 ± 0.2	98.6 ± 0.1	99.0 ± 0.1	99.2 ± 0.1
ANN Single	92.5 ± 0.4	96.9 ± 0.3	98.2 ± 0.2	98.7 ± 0.2	99.1 ± 0.2
ANN Bagging	93.0 ± 0.4	97.4 ± 0.3	98.5 ± 0.1	99.0 ± 0.1	99.3 ± 0.1

ANNs are known to be a "unstable" learning algorithm since it produces fairly different predictors for small changes in the learning data (or even in two realizations of the learning process without any changes at all in the learning data). On the contrary, the nave Bayes classifier is "stable", since it is based on parameterizing the class conditional probability for each (assumed) independent feature by a normal distribution. Fitting the mean and variance of a normal distribution is only marginally sensitive to small changes on the data. The dispersion of the results presented in Table 1 for the nave Bayes classifier are then due only to the bootstrap re-sampling of the learning data and not to intrinsic instabilities of this algorithm. These dispersions are almost the same as those of the ANN classifiers, despite the fact that this last method is potentially much more unstable, an indication that the fitting process of the ANN's multiple parameters is robust and not prone to overfitting problems. We stress the excellent performance of the nave Bayes classifier, in both single and bagged versions, which might be related to an effective selection of nearly-independent classification parameters. This point had already been remarked in Granitto et al. (2002). However, there we (wrongly) speculated that for a much larger number of species the classification problem would be more demanding and ANN ensembles might have an advantage over simpler methods. We also see that the performances of single classifiers leave not much room for further improvement by bagging them (differences between single and bagged algorithms are, however, statistically significant).

From Table 1 we see that bagged classifiers reach a performance of 99% for the 236 different weed species considered when they are allowed to suggest four options for class membership. For comparison, in our previous work (Granitto et al., 2002) this performance was obtained with $n = 3$ but for only 57 different species in the database. Notice also that different realizations of training and test sets do not substantially change performances, as indicated by the low standard deviations observed in the 30 independent runs.

3.2 Black and white images

In general the largest discriminating power corresponds to morphological features. This was established in Granitto et al. (2002), where it was also shown that, as expected, color features are not particularly good classification parameters since many species are light to dark brownish or black. On the other hand, the combined use of morphology and texture characteristics alone (without color features) requires only considering black and white images, which constitutes an important simplification in system's operation and leads to a reduction in cost. In fact, color images require a much better control of illumination conditions than black and white ones, and the required acquisition hardware (RGB camera, frame grabber, etc.) is substantially more expensive.

To explore in more detail the possibility of working with black and white images to identify weed seeds, one can consider using only the 8 morphological and textural features listed in the previous section. However, since two of the color features (variance and skewness of the intensity histogram) are related only to the intensity channel, they can also be obtained from black and white images. Consequently, we have explored the classification capabilities of a system built in terms of the resulting set of 10 features, *i.e.*, excluding the ratios $E[R]/E[I], E[G]/E[I]$ of average pixel values in RGB channels from the 12 parameters listed in the previous section. The corresponding results are given in Table 2.

Table 2
Performances of Naïve Bayes (NB) and ANN classifiers as percentage of correct seed identifications using the 10 features measured from black and white images. Mean values and standard deviations are estimated from 30 independent experiments, as described in the main text.

B&W (10 f.)	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
NB Single	85.2 ± 0.7	93.4 ± 0.6	96.0 ± 0.5	97.4 ± 0.4	98.1 ± 0.3
NB Bagging	85.5 ± 0.6	93.8 ± 0.7	96.5 ± 0.4	97.8 ± 0.3	98.4 ± 0.3
ANN Single	85.9 ± 0.8	93.3 ± 0.6	95.8 ± 0.5	97.1 ± 0.5	97.9 ± 0.4
ANN Bagg.	87.3 ± 0.6	94.1 ± 0.5	96.5 ± 0.4	97.6 ± 0.4	98.2 ± 0.3

Unfortunately, we see that for $n = 1$ (standard classification) there is an important loss in classification accuracy from 93% to $\sim 87\%$ due to the elimination of the two color features. This lack of discriminating information could be perhaps compensated by adding some of the features eliminated in the previous selection process, since they might become relevant in the absence of color information. In order to check this possibility and improve the results in Table 2, we have carried out a new selection process for all the morphological and textural features plus the variance and skewness of the intensity histogram. The final quasi-optimal set obtained contains again 12 features; they correspond to replacing η_{20}, Contrast, $E[R]/E[I]$ and $E[G]/E[I]$ from the features listed in the previous section by the following four parameters:

- Moments η_{02} and η_{12} of the planar mass distribution with respect to the principal axes (Chtioui et al., 1996).
- Gray Level Distribution (Galloway, 1975) along the main principal axis direction.
- Run Percentage (Galloway, 1975) along the main principal axis direction.

In Table 3 we present the results obtained using this new set of 12 features measured from black and white images. By comparison with Table 2, we see that there are no important changes in performances by the backtracking strategy of adding some of the discarded features and performing a new selection. In any case, for $n = 4$ now both bagged classifiers have nearly 98% of accuracy, only $\sim 1\%$ less accurate than the corresponding one including color features. These performances are still acceptable for the concrete application we are considering, since presenting four options to the operator is not unreasonable. Notice, however, that for $n = 1$ (standard classification) there is still a sensible loss from 93% to 88%, which would require implementing more efficient classification methods to overcome it.

Table 3
Same as Table 2 using the new set of 12 features measured from black and white images.

B&W (12 f.)	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
NB Single	86.3 ± 0.4	93.5 ± 0.7	96.2 ± 0.5	97.4 ± 0.3	98.1 ± 0.2
NB Bagging	86.5 ± 0.4	94.1 ± 0.7	96.7 ± 0.5	97.8 ± 0.4	98.4 ± 0.3
ANN Single	86.5 ± 0.8	93.7 ± 0.5	96.1 ± 0.4	97.4 ± 0.4	98.1 ± 0.4
ANN Bagging	88.1 ± 0.6	94.6 ± 0.6	96.8 ± 0.4	97.8 ± 0.4	98.4 ± 0.2

4 Conclusions

We extended our previous work (Granitto et al., 2002) on the feasibility of using machine vision algorithms for the identification of weed seeds. We considered a much larger database than those used in previous studies, comprising

10130 images of 236 species, and discussed the discriminating power of weed seed characteristics measured from color images. We considered the same features as in Granitto et al. (2002), which had been carefully selected to maximize the classification performance of a naïve Bayes classifier. This led to the nearly optimal set of 12 characteristics -6 morphological, 4 color and 2 textural properties- listed in Section 2.

For the large database considered in this work, the naïve Bayes classifier produced again excellent results, as shown in Table 1. This seems to confirm the conclusion advanced in Granitto et al. (2002), namely, that due to the careful parameter selection the retained set approximately fulfills the independence assumed by this simple Bayesian approach. Table 1 also shows that the performances of this method and of ANN-based classifiers are essentially the same (except perhaps for $n = 1$, where the latter are slightly better). In particular, for the bagged versions these performances reach 99% for $n = 4$ classification options.

We have also investigated the possibility of using only black and white images of weed seeds. This alternative is appealing since in this case illumination conditions are less critical and the required hardware much cheaper, which are important advantages for a potential commercial system. Our results in Table 3 show that, by using a set of 12 features obtainable from black and white images, the bagged classifiers are able to reach approximately 98% of classification accuracy for $n = 4$. This is only a 1% loss with respect to the use of color features, and is still an acceptable performance for the application under consideration. We stress that even on this more difficult problem the ANN approach does not outperform the simpler Bayesian approach.

The number of species considered in this study is large enough to draw some definite conclusions. Firstly, for applications involving the identification of a few hundred species the present approach to weed seed identification is effective since the results do not deteriorate sensibly by scaling the database from 57 to 236 species. Secondly, for not too critical applications where the system is required only to suggest a few classification options, this approach can even be implemented using black and white seed images. In particular, this is the case for the intended application in commercial seed purity analysis. Finally, for more critical problems where presenting several options to an operator is not feasible and/or the number of species involved is extremely large –like, for instance, for botanical gardens use– the implemented classifiers have to be improved. There are several ways of doing it: As already mentioned, feature selection for the ANN classifiers should be performed using an ANN as the selection criterion, although this would probably not make them much more efficient; more sophisticated feature selection methods (Jain and Zongker, 1997; Chtioui et al., 1998) than the one used in Granitto et al. (2002) could be implemented; a better control of illumination conditions (for instance, keeping the background color constant by an electronic control of the light source) would enhance the discriminating power of color and texture parameters; etc. In addition, "boosting" techniques (Freund and Schapire, 1997) can be applied to improve the capabilities of the basic learners studied here and other classification methods may be considered (boosted decision trees, for instance). Finally, the whole classification strategy might be changed to a method better suited for large multiclass problems like this application. Some of these possibilities are currently under consideration.

Acknowledgments

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

I.S. Ahmad, J.F. Reid, M.R. Paulsen, J.B. Sinclair, "Color classifier for symptomatic soybean seeds using image processing," *Plant Disease* **83** (1999) 320-327.

C.M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford (1995).

L. Breiman, "Bagging predictors," *Machine Learning* **24** (1996) 123-140.

C. Chen, Y.P. Chiang, Y. Pomeranz, "Image analysis and characterization of cereal grains with a laser range finder and camera contour extractor," *Cereal Chem.* **66**(6) (1989) 466-470.

Y. Chtioui, D. Bertrand, Y. Dattée, M.F. Devaux, "Identification of seeds by color imaging: Comparison of discriminant analysis and artificial neural networks," *J. Sci. Food Agric.* **71** (1996) 433-441.

Y. Chtioui, D. Bertrand, D. Barba, "Feature selection by a genetic algorithm. Application to seed discrimination by artificial vision," *J. Sci. Food Agric.* **76** (1998) 77-86.

R.W. Conners, M.M. Trivedi, C.A. Harlow, "Segmentation of a high-resolution urban scene using texture operators," *Computer Vision, Graphics and Image Processing* **25** (1984) 273-310.

S.R. Draper, A.J. Travis, "Preliminary observations with a computer based system for analysis of the shape of seeds and vegetative structures," *Journal of National Institute of Agricultural Botany* **16** (1984) 387-395.

Y. Freund, R.E. Schapire, "A decision theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences* **55**(1) (1997) 119-139.

M.M. Galloway, "Textural analysis using gray level run length," *Computer Graphics Image Processing* **4** (1975) 172-179.

P.M. Granitto, H.D. Navone, P.F. Verdes, H.A. Ceccatto, "Weed Seeds Identification by Machine Vision," *Computers and Electronics in Agriculture* **33** (2002) 91-103.

R.M. Haralick, K. Shanmugam, I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics* **3**(6) (1973) 610-621.

A. Jain, D. Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(2) (1997) 153-158.

P.I. Jansen, "Seed production quality in Trifolium balansae and T. resupinatum: The role of seed color," *Seed Sci. Technol.* **23** (1995) 353-364.

P.D. Keefe, S.R. Draper, "The measurement of new characters for cultivar identification in wheat using machine vision," *Seed Sci. Technol.* **14** (1986) 715-724.

T.M. Mitchell, *Machine Learning*, WCB/McGraww-Hill, Boston (1997).

M.R. Neuman, H.D. Sapirstein, E. Shwedyk, W. Bushuk, "Discrimination of wheat class and variety by digital image analysis of whole grain samples," *J. Cereal Sci.* **6** (1987) 125-132.

M.R. Neuman, H.D. Sapirstein, E. Shwedyk, W. Bushuk, "Wheat grain color analysis by digital image processing I. Methodology," *J. Cereal Sci.* **10** (1989a) 175-182.

M.R. Neuman, H.D. Sapirstein, E. Shwedyk, W. Bushuk, "Wheat grain color analysis by digital image processing II. Wheat class discrimination," *J. Cereal Sci.* **10** (1989b) 183-188.

P.E.H. Petersen, G.W. Krutz, "Automatic identification of weed seeds by color machine vision," *Seed Sci. Technol.* **20** (1992), 193-208.

H.D. Sapirstein, M. Neuman, E.W. Wright, E. Shwedyk, W. Bushuk, "An instrumental system for cereal grain classification using digital image analysis," *J. Cereal Sci.* **6** (1987) 3-14.

S.J. Symons, R.G. Fulcher, "Determination of wheat kernel morphological variation by digital image analysis: I. Variation in Eastern Canadian Milling Quality Wheats," *J. Cereal Sci.* **8** (1988) 211-218.

I. Zayas, F.S. Lai, Y. Pomeranz, "Discrimination between wheat classes and varieties by image analysis," *Cereal Chem.* **63**(1) (1986) 52-56.

I. Zayas, Y. Pomeranz, F.S. Lai, "Discrimination of wheat and nonwheat components in grain samples by image analysis," *Cereal Chem.* **66**(6) (1989) 233-237.