DEVELOPMENT OF AN AUTOMATED DEVICE FOR SORTING SEEDS
Application on Sunflower Seeds

Vincent Muracciole, Patrick Plainchault
INRA;ESEO;LISA;GEVES-SNES, Université d'Angers, Angers, France
CER ESEO, Angers, France
vincent.muracciole@eseo.fr, patrick.plainchault@eseo.fr

Dominique Bertrand, Maria Rosaria Mannino
ENITIAA-INRA, Sensométrie - Chimiométrie, Nantes, France
GEVES-SNES, Beaucouzé, France
bertrand@enitiaa-nantes.fr, maria-rosaria.mannino@geves.fr

Keywords: Machine vision, computer vision, seed identification, grain quality, seed quality, embedded systems.

Abstract: Purity analysis and determination of other seeds by number are still made manually. It is a repetitive task based upon visual analysis. Our work objective is to create and use a simple and quick automated system to do this task. A first step of this machine has been reached by validating the image acquisition and feeding process. The principle of this machine is based on a seeds fall with stroboscopic effect image acquisition. This article presents the first step of creating a dedicated and autonomous machine which combines embedded constraints and real time processes.

1 INTRODUCTION

In most countries, seeds cannot be commercialized without an assessment of their quality. Quality control tests generally includes the evaluation of unwanted materials in the batches. This control is currently performed manually by operators able to separate the pure seeds from impurities.

In France, the official assessment of seeds is performed by GEVES (Groupe d'Etude et de contrle des Varits; Variety and seed study and control group). The standardized methods of the ISTA (International Seed Testing Association) and of the AOSA (Association of Official Seed Analysts) are applicable to two kinds of analyses of the seeds: purity analysis and determination of other seeds by number. In the analytical analysis of purity, all the elements in a sample (pure seeds (PS), inert matter (IM) and seeds of other plants (SOP)) are identified, separated and weighed. The result is then given as percentage of each part's weight.

In the determination of other seeds by number analysis, only the impurities are separated, identified and counted. The standardized methods of purity analysis are very requiring since a one hundred percent correct identification is expected. The visual inspection of seeds is time-consuming and needs operators to be trained. It is thus very desirable to develop an automaton making it possible to do such controls.

Some mechanical devices are commonly used to clean commercial seeds and separate them according to their qualitative nature. It seems possible to add an artificial vision system for completing the mechanical step. The artificial vision domain follows closely the quick growth of computer power. This evolution allows creating new applications which was not possible until today. (Jayas D.S and Bulley, 1999) (Moltó Enrique, 1998) (Egelberg P., 1994) (Pearson, 1994) (Wan, 2002) (Bennett, 2005)

Image analysis is not currently restrained to simple geometrical shapes and can be applied to complex objects such as biological materials.

All the work dealing with the identification of seeds proceed in two steps: extraction of features from a digital images of seeds, processing the features for building up a discriminant model. In most of the published studies, the proportions of correctly classified seeds reported by the authors hardly exceed 95 percent. Such figures are encouraging but probably not sufficient to allow the replacement of human operators by machines.

Most of the recent studies are based on the use of very complex discriminant methods involving neural or Bayesian approaches (Chtioui, 1997) (Granitto P.M. and Ceccatto, 2003) (Granitto Pablo M and Ceccatto, 2003) (Majumdar S., 2000b). These methods are often efficient but involve a very time-consuming training phase. The parameters of these predictive models (Kernels, neural weights) require to be stored in computer memories.

In a previous work, a vision system for the automatic acquisition of images of individual seeds has been designed and developed. Unlike other systems, our device is based on dedicated hardware. This dedicated hardware has some limitations. These limitations are directed by the embedded constraints: a limited code size implementation, a low memory size available, a limited binary mathematical operation, a low frequency compared to a PC processor, limited temperature dissipation and a limited integrated circuit size. Design and development of the vision system include these different constraints (Plainchault P., 2003). This is why it is not possible to implement very complex algorithms for discrimination into an embedded board and some choices have been made in the design and the development of system.

A linear discriminant analysis (LDA) has been chosen as a decision algorithm. LDA is applied on the features extracted from the individual image of seeds.

In the present work, the features include morphological, colour and texture characteristics. According to the application, the discriminating power of variables may change. Another way to reduce the algorithms complexity is to select a few predictive variables adapted to a given application. Moreover in practical applications, a large part of the sample can be accurately identified and it may remain only a small proportion in which the discrimination is more difficult. A relevant approach consists in using an automated system for a first screening and let the more difficult part of the sample to be manually analysed.

Instead of computers, we propose to use an FPGA (Field Programmable Gate Array). These hardware components allow the implementation of a dedicated architecture for vision chain. FPGA gives the possibility to apply massively parallel processes. We present below the chain of vision and the mechanism of image acquisition already in place. We finish by the presentation of the method and the results.

2 ARTIFICIAL VISION

The chain of artificial vision (Figure 1) includes several phases: acquisition, image processing, feature extraction and decision. Each element of this chain must be optimised in order to reduce computing time and increase efficiency.

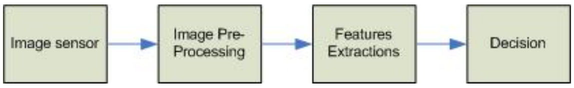


Figure 1: Vision chain.

2.1 Image Acquisition

Acquisition is a very important part of the performance system. It determines the quality of the image and the image processing time. A good image quality is without noise and without defect of illumination. These conditions are determinant for the image processing speed and the level of the results.

The quality depends on the choice of three principal types of sensor, on the choice of the lighting system and the selected background.

2.1.1 Sensor

The general principle of the sensors lies in the projection of a 3D scene on a 2D matrix where each cell carries out the summation of the photons reflected by the object into the scene.

- Mono CCD:
  A mono-CCD sensor is generally organized according to the Bayer mosaic. This matrix represents an alternation of separating cells of red, green and blue colours organized as shown in Figure 2. This matrix indeed leaves zones where the two others colours are not represented. In order to fill this absence, some methods of interpolation have been presented in this article (Gunturk Bahadir K., 2005) . Some methods can privilege certain features like anti-aliasing, shape enhancement or noise reducing.

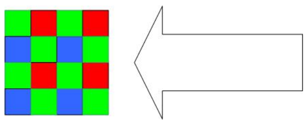


Figure 2: Mono CCD Sensor : Bayer Mosaic.

- Three CCD:
  Three ccd sensors do not need an interpolation phases. Colour separation is made by a prism which projects it onto three two-dimensional matrices, one red, one green, one blue as presented in Figure 3.
- Mono-chip Multichannel CCD:
  This sensor has been developed by the Foveon

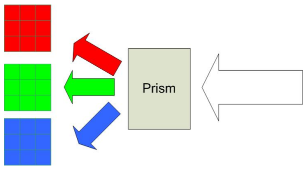


Figure 3: Three CCD Sensor.

company. It performs colour separation on the silicon. This sensor comes back to the photosensitive film principle as represented in Figure 4.

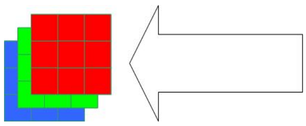


Figure 4: X3 Sensor.

2.1.2 Illumination Scene

Illumination is an important part to reduce image processing time. The more homogeneous illumination, we have. The quicker image processing, we have. In our case, a good separation between background and object is needed to maintain the real-time process. The illumination system can be composed of incandescent light, discharge lamps or light-emitting diodes. The choice of these systems is determined by the application and their characteristics.

2.1.3 Interpolation

The interpolation is a demosaicing phases for mono-CCD sensors. There are two methods to obtain images in three channels. The first method, is to reduce image size by calculating the average of green channel (1) and leave the red and blue channel as represented in the figure 5. The second is to use interpolation methods that can maintain sensor size.

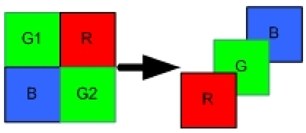


Figure 5: Representation of the image into the three channel.

$$\left\{ \; G = \tfrac{G1+G2}{2} \; \right\} \tag{1}$$

The requirements of embedded systems are elementary binary mathematical operators. For that reason, only bilinear and associated interpolation methods have been retained. These methods are totally independent and deterministic which respect the real time and parallel processing system objectives. These interpolations are reversible, avoid memory overloading and reduce image transfer time.

There are two possible methods : the bilinear interpolation represented in figure 6 and equation 2 and 5 (equation for red and blue channels are the same) and the constant-difference-based interpolation represented in Figure 7.

$$G22 = \frac{G12+G32+G21+G23}{4} \tag{2}$$

$$R22 = \frac{R11+R13+R31+R33}{4} \tag{3}$$

$$R12 = \frac{R11+R13}{2} \tag{4}$$

$$R21 = \frac{R11+R23}{2} \tag{5}$$

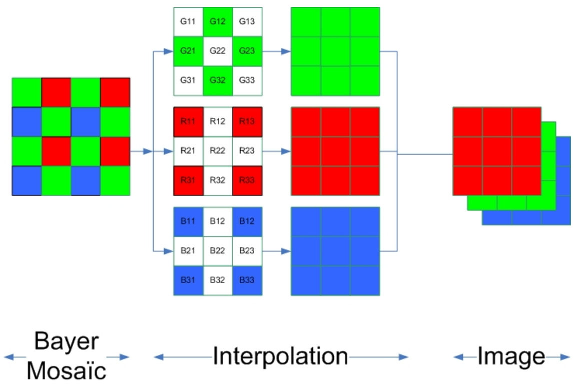


Figure 6: Bilinear Interpolation

.

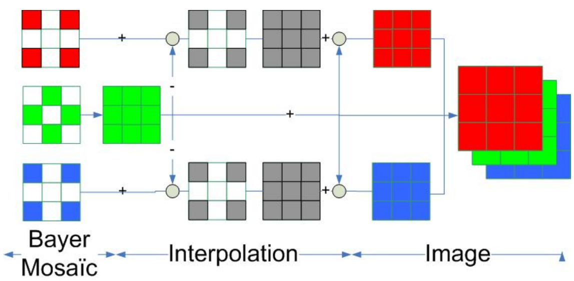


Figure 7: Constant-difference-base Interpolation.

2.1.4 Color Space

In general, the embedded systems operate using the sensor's RGB color space. Each channel in this space are strongly correlated. The choice of the colour space is dependant on the colour background. Generally a blue background is used for seed inspection. This background needs a new color space to perform seed extraction. The RGB space does not perform very well this process, so the conversion in the YCbCr (equation 6) space can be possible and is the best to perform this process. The choice of this space has been made for the following reasons:

1. It is a matrix based conversion
2. It is a reversible transformation (linear transformation)
3. The distance between seed and background is maximized

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.5 \\ 0.5 & -0.419 & -0.081 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (6)$$

2.2 Feature Extraction

The extraction aims at defining object based on its morphology, its colour and its texture. For this extraction we need to extract the object and define his position, we use for that a segmentation task and a labelling task.

The segmentation algorithms used to perform the seed positionning task is a variance-based algorithm (Otsu, 1979).

The extracted features defining seeds are detailled by many publications. These features are summarized in (Majumdar S., 2000a) (Majumdar S., 2000b) (Majumdar S., 2000c) and in the thesis of Younes Chtioui (Chtioui, 1997). These features are around 110. There are three groups: morphological, colour and textural. In the morphological group, there is dimension, invariant moment and fast fourrier transform type. In the colour group, there is standard deviation, mean, variance. In the textural group, there is descriptor of texture, co-occurence, entropy, kurtosis etc...

2.3 Pattern Recognition: Linear Discriminant Analysis

Many identification methods (Chtioui, 1997) (Visen N.S., 2002) (Majumdar S., 2000d) in the domain of seeds are presented. Some of them cannot be retained because they do not answer the embedded constraints. This is why we privilege the simple methods based on the distances between groups and the methods of the regression type.

Some previous work show the feasibility of pattern recognition in FPGA based systems (Miteran J., 2005). An LDA method has been chosen. This method is based on the principle of the minimal distance to a group of a given seed as shown in figure 8 and the equation. It is the type of decision algorithm easily parallelisable, where each class are independent of its neighbours. These methods and algorithm has been associated to the Mahalanobis distance (equation 7) in order to increase the distance between classes. The x vector is the unknown seed to be classified. The μ vector is the mean value of a labelled group features. The Σ value is the covariance matrix for multivariate vector.

$$d^2(x) = (x-\mu)^T \Sigma^{-1} (x-\mu) \quad (7)$$

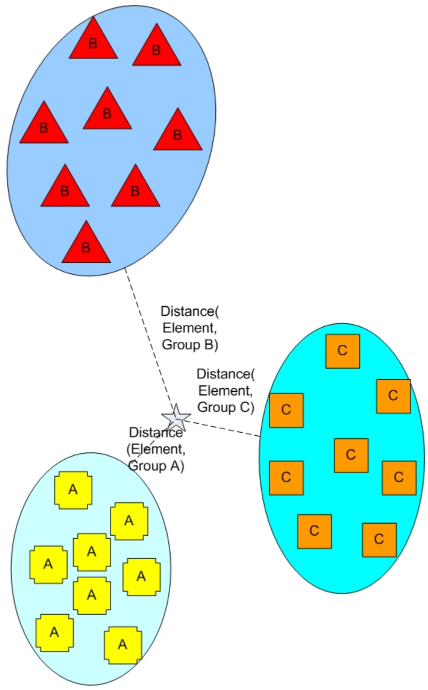


Figure 8: Principle of Linear Discriminant Analysis.

2.4 Hardware

The architectures presented in this section allow the creation of dedicated systems based on work from the computer world. They combine good computing efficiency with low power consumption, low memory and device monitoring constraints. These various points make them the ideal components for the embedded world.

2.4.1 MCU

Microcontrollers are circuits largely used for the management of the automated control processes. Although not intended for mathematical computation, the new evolutions of these components integrate more and more mathematical calculating units (example EP9315A from Cirrus Logic with mathematical coprocessor), and even cores DSP (example architecture OMAP Texas Instrument(TI)). However their performance is still limited for signal processing.

2.4.2 DSP

DSP are processors dedicated to signal processing. They combine the flexibility of a programmable processor with performance in real time signal processing. The new evolutions of these architectures saw the capacities of calculation increased to the detriment of consumption. However now they represent now a true general purpose processor for the embedded domain.

2.4.3 FPGA

FPGAs see their uses and their applicability grow. They combine the performances of a dedicated architecture and the flexibility of programming. FPGAs are composed of independent logical cells which have to be activated to create dedicated processes. Nowadays some signal and image processing are implemented into these components and represent new application for high efficiency and low cost systems.

3 MATERIALS AND METHOD

The automaton is based on a principle of free fall seed (Plainchault P., 2003). It acquires an image of each seed. It adds to it a stroboscopic effect breaking up the movement of fall. In the long term, the automaton must reach a sorting at the rate of one seed per second. The sorting in the broad sense covers acquisition, the image processing, the extraction of the parameters, the identification and, finally, the sorting at the same time.

3.1 Device

The automaton is composed of a sensor, a flash, a detection barrier, a background, an electronic system. Figure 9 shows this association.

3.1.1 Seed Sorting System

- Mechanical feeding:
  The mechanical feeding is based on an electromagnet which attracts a support heightened by blades of metal.
- Sensor :
  The sensor provides the frame grabbing function. It is a Kodak KAC-1310 sensor with a resolution of 1280*1024 and a clock frequency of 10MHz. It has the advantage of having a window activation authorizing the activation of only a part of the sensor. This function can increase speed of acquisition because this speed is proportional to the activated dimension.
- Flash Light :
  The flash is ensured by 8 diodes luxeon lumileds with a unit power of one Watt. This choice has been done for the response time and robustness of this component for the rate of one seed per second.
- Barrier :
  The detection barrier is based on infra red cells functioning in saturated mode. It determines the cut of the beam. For the moment, this barrier is able to treat only sunflower seeds, sunflower kernels, or sclerotia.
- Background :
  A blue background was selected because it allows a better segmentation of the object by using the YCbCr colour space and extent the seeds varieties
- Embedded Card :
  The electronic system lies on an embedded card. It is a sundance SMT 355 card composed of a TI DSP TMS320C32 and an ALTERA FPGA Flex 10K130E and 512 Kbytes of RAM. Its task is to control the peripherals and perform simple processing. Due to the architecture of the card, it is not simple to implement image processing into the FPGA. The FPGA does not have direct access to memory and the memory space is too low to implement image processing. But this card has tested and validates the feasibility of acquisition in free fall of object into embedded card.

3.1.2 Computer Validation of Embedded Algorithm

- Computer
  The electronic system also includes a computer. It is a DELL computer with an Intel Pentium IV 3GHz with HyperThreading processor, 1GB of DDR-SDRAM, a graphics board AMD/ATI X300 with dedicated memory and a 160GB hard disk. The computer provides the Human-Machine interface to control the automaton. It saves the image acquired in Bitmap format on the hard disk drive. The computer also plays a part in the algorithm validation. The identification algorithm

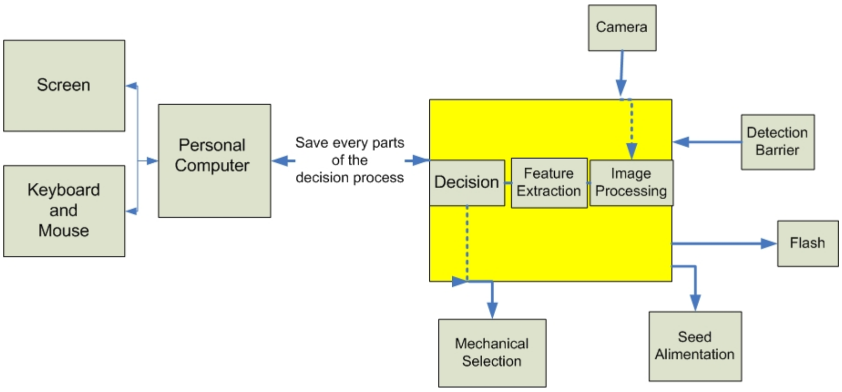


Figure 9: Presentation of the wanted system.

has been tested off line in order to determine its efficiency for the sorting.

- Software

  The automaton is controlled by a program implemented in the computer using Visual C++. It provides a window interface to control the acquisition process. The algorithm validation is carried out off line with the image acquired by the automaton under the environment of matrix algebra MATLAB 7. Firstly, we calculate the parameters for all the datasets. In order to have all the features algorithms we just use the bwmorph function from the matlab image processing toolbox. Secondly, we train and test the identification algorithm on the dataset to select the best algorithm.

3.2 Methods

Seeds and impurities observed by the vision system were extracted from 23 samples of sunflower seeds representing commercial seed lots of various origins and 18 varieties, differing by seed colour and size. The dataset for impurities with a low frequency in the samples was enlarged using the reference collection of the GEVES-SNES. A dataset including 1051 images has been created. This dataset includes 6 classes: sunflowers kernels, broken seeds (fragments of seeds with a size higher than 50 percent of the size of the seed), mutilated seeds (fragments of seeds with a size lower or equal to 50 percent), sunflower achenes (intact seeds), sclerotia and soil.

At this stage of the study, our dataset includes all the categories (seeds and impurities) that have to be identified in official analysis of commercial seed lots (table 1) with the exception of seeds belonging to other species than the sunflower. Figure 10 shows colour, texture and morphology differences among the different classes.

This dataset has been labelled by an expert of the

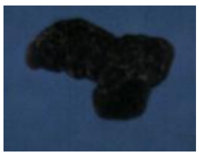
Table 1: Principle of the analysis in the framework of purity and counting seed.

Object	Purity	counting
sunflower kernels	PS	
sunflower achenes	PS	
broken seeds	PS	
mutilated seeds	IM	
soil	IM	
sclerotia	IM	counting
Seeds of other plants	SOP	counting


(a) Sunflower achenes


(b) Sunflower Kernels


(c) Soil


(d) Sclerotia


(e) Broken seeds

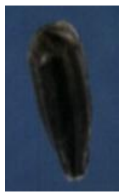

(f) Mutilated seed

Figure 10: Image of pure seeds (a,b,e) and impurities (c,d,f).

GEVES-SNES. With this dataset we have conducted a linear discriminant analysis using cross-validation. We have randomly selected 348 images for the training sets and 703 for the testing sets in accordance with table 2. We obtain a reduction of the feature space with a stepwise canonical discriminant analysis according to the article (Bertrand D., 1990). The selection of parameters is made incrementally by testing the increase of identification percentage.

Table 2: Dataset representation of each groups.

Object	training set	testing set	total
sunflower kernels	56	113	169
broken seeds	72	145	217
mutilated seeds	36	72	108
sclerotia	61	123	184
sunflower achenes	66	134	200
soil	57	116	173
total	348	703	1051

4 RESULTS

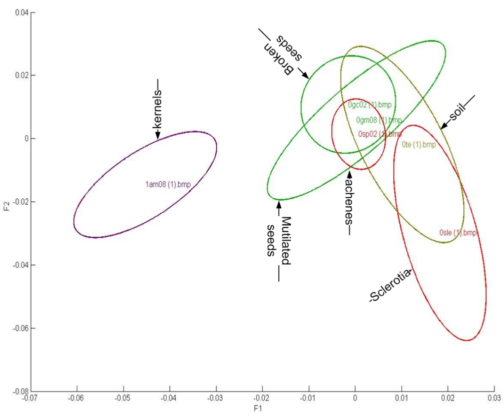


Figure 11: representation map of the sample.

The parameters selected by the LDA method show a great extent of texture (energy, entropy, occurrence), colour (mean, variance) and a few morphological (perimeter, area, moments, Fast Fourier Transform) parameters. Figure 11 shows the representation map of the sample and table 12 indicates the identification results for all the classes. A high percentage of each class has been correctly identified: from 75% of broken seeds to 97% of sclerotia. Some misclassification occurred, with the extent and importance that vary in relation to the classes. If we seek to identify independently each group, the selected features change but follow what we have said before in the part 3.2.1. The

	kernels	broken seeds	mutilated seeds	sclerotia	achenes	soil
kernels	108	5	0	0	0	0
broken seeds	0	111	19	0	15	3
mutilated seeds	0	8	56	0	1	4
sclerotia	0	0	4	119	0	0
achenes	0	6	8	0	116	3
soil	0	0	1	3	5	108

Figure 12: Differentiation sclerotia, achenes, broken seeds, mutilated seeds, soil.

results expressed in table 12 is in accord with the map of labelled groups (figure 11). The table 12 shows all of the misclassifications identified on the map. Many of the errors of classification are due to a lack of information in the image. The free fall principle with only one camera has the same problem as on a rolling carpet (Egelberg P., 1994): it is possible that the camera does not take the image of the face bringing information, which is the case after analysis of the badly identified seeds. As we can note in table 12, there is a strong confusion between broken seeds and mutilated seeds. The algorithm can not perform a virtual representation of the associated pure seeds as the human expert can. From a practical point of view, we could accept a misclassification when it affects pure seeds (kernels, achenes and broken seeds) identified as impurities (e.g. sclerotia classified as soil). In this case, to achieve the analysis, the analyst will observe only a fraction of the sample classified as impurities by the system. Following this principle, the system classification was less satisfactory for mutilated seeds and soil: 13% of mutilated seeds were classified as pure seeds (broken seeds or achene), 4% of soil as achene. This first result shows the interest of the system when a determination of other seeds is carried out in sunflower: the algorithm makes possible 100% of correct separation of pure seed (achenes or kernels of sunflower) from sclerotia that are searched for in this kind of analysis.

5 CONCLUSION

The widening of the dataset of image seeds to new species is under acquisition.

The remaining work will consist of performing acquisition, features extractions and decision in a complete automaton. Optimisation and testing on other varieties of the decision algorithm have to be conducted.

Moreover the creation of a multi-camera system of acquisition according to the same principle would allow on certain ambiguities at the time of the phase of decision. It is also possible to make this approach more robust by creating a unknown class for seeds at equal distance from the various groups.

The testing on the identification algorithm on seeds of other plants has to be made. System enhancements have to be made like changing the infrared barrier in order to extend seed varieties acquisition. The design and the development of a new hardware system have to be made in order to implement a three camera systems.

ACKNOWLEDGEMENTS

This work was supported by the National Institut of Agronomical Research (INRA), the variety and seed study and control group (GEVES), the ESEO graduate school of engineering in electronic, and the LISA laboratory of the University of Angers.

For their financial support the region Pays de Loire and the CER Vegetal (French contract).

REFERENCES

Batlle J., M. J. (2002). A new fpga/dsp-based parallel. *Real time imaging*, 8:345–356.

Bennett, T. D. K. F. M. M. M. (2005). A computer-based system for seed identification. *Seed Technology*, 27:190–202.

Bertrand D., Coucroux P. Autran J.-C., M. R. R. P. (1990). Stepwise canonical discriminant analysis of continuous digitalized signals : applicatoin to chromatograms of wheat proteins. *Journal of chemometrics*, 4:413–427.

Chtioui, Y. (1997). *Reconnaissance automatique des semences par vision artificielle basée sur des approches statistiques et connexionnistes*. PhD thesis, IRESTE.

Davies, E. (2005). *Machine Vision - Theory, Algorithms, Practicalities*.

Egelberg P., Mansson O., P. C. (1994). Assessing cereal grain quality with a fully automated instrument using artificial network processing of digitized color video images. In *Forestry and Biological Processing*, volume 2345, pages 146–156. SPIE congress.

Granitto Pablo M, P. F. V. and Ceccatto, H. A. (2003). Automatic identification of weed seeds. *ASAI*.

Granitto P.M., P.A. Garralda, P. V. and Ceccatto, H. (2003). Boosting classifiers for weed seeds identification. *JCS&T*.

Gunturk Bahadir K., e. a. (2005). Demosaicking : Color filter array interpolation. *IEEE Signal Processing Magazine*, pages 44–54.

Jayas D.S, M. C. and Bulley, N. (1999). An automated seed presentation device for use in machine vision identification of grain. *Canadian agricultural engineering*, 41(2):113–118.

Majumdar S., J. D. S. (2000a). Classification of cereal grains using machine vision i. morphological models. *American Society of Agricultural Engineers*, 43:1669–1675.

Majumdar S., J. D. S. (2000b). Classification of cereal grains using machine vision ii. color models. *American Society of Agricultural Engineers*, 43:1677–1680.

Majumdar S., J. D. S. (2000c). Classification of cereal grains using machine vision iii. texture models. *American Society of Agricultural Engineers*, 43:1681–1687.

Majumdar S., J. D. S. (2000d). Classification of cereal grains using machine vision iv.combined morphology, color, and texture models. *American Society of Agricultural Engineers*, 43:1689–1694.

Miteran J., M. J. (2005). Automatic hardware implementation tool for a discrete adaboost-based decision algorithm : Prototyping for machine perception on a chip. *EURASIP*, 2005:1035–1046.

Moltó Enrique, Blasco José, B. J. V. (1998). Computer vision for automatic inspection of agricultural produces. In SPIE, editor, *Precision agriculture and biological quality*, volume 3543, pages 91–100.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9:62–66.

Pearson, T. (1994). Machine vision system for automated detection of stained pistachio nuts. In SPIE, editor, *Optics in Agriculture, Forestry, and Biological Processing*, volume 2345.

Plainchault P., Demilly D., F. A. T. S. B. D. (2003). Development of an imaging system for the high speed identification of seeds. *IEEE Sensor*.

Visen N.S., J. D. (2002). Specialist neural network for cereal grain classification. *Biosystems Engineering*, 82:151–159.

Wan, Y.-N. (2002). Automatic grain quality inspection with learning mechanism. *Information technology in Agriculture*.