

Cereal Grain Classification by Optimal Features and Intelligent Classifiers

A. Douik, M. Abdellaoui

Ali Douik, Mehrez Abdellaoui

Ecole Nationale d'Ingénieurs de Monastir (ENIM)

Département de Génie Electrique

Laboratoire ATSI

Rue Ibn El Jazzar, 5019 Monastir Tunisie

E-mail: {Ali.douik,mehrez.abdellaoui}@enim.rnu.tn

Abstract: The present paper focused on the classification of cereal grains using different classifiers combined to morphological, colour and wavelet features. The grain types used in this study were Hard Wheat, Tender Wheat and Barley. Different types of features (morphological, colour and wavelet) were extracted from colour images using different approaches. They were applied to different classification methods.

Keywords: morphological, colour, wavelet transform, neural networks, statistical classifier, fuzzy logic.

1 Introduction

The past few years was marked by the development of researches that contribute to reach an automatic classification of cereal grains which is perceived as a possible solution to prevent human errors in the quality evaluation process. Computer vision system which is a promising technology in the quality control can replace the human operator. After hours of working the operator may loose concentration which in turn will affect the evaluation process. So a computer vision system proved to be more efficient at the level of precision and rapidity. But, the natural diversity in appearance of various cereal grains varieties makes classification by computer vision a complex work to achieve. Many researches were carried out to classify cereal grains. Characterization models were based on morphological features ([1–9]), colour features ([10–13]) or textural features ([14]). Other researchers ([15–18]) have tried to combine these features for the sake of improving the efficiency of classification. Recently, wavelet technique was integrated in cereal grains characterization ([19,20]). This technique, developed by Mallat [21], is used in textural image analysis to make object classification more precise. The present paper is divided into four main parts. The first one will deal with the cereal image acquisition system, the second part will be devoted to present the classification features with its morphological, colour and wavelet components, the third section will focus on the different methods used in the classification process and the last one will compare the different methods accompanied with their performance evaluation.

2 Cereal image acquisition system

2.1 Image acquisition device

A high resolution colour camera (VIVITAR) with a USB 2.0 cable was used to acquire grain images. The acquired images were of 3.1 mega pixel resolution. Light sources were placed symmetrically over and under a glass plate over which the grains are spread out. All the samples were taken at constant camera settings, i.e., exposure time, saturation and gamma. The images obtained were pre-processed to eliminate background pixels using image subtraction. Indeed, the active image containing grain sample is compared to image containing background. The image we got contains the grains and a uniform background (black). This step of pre-processing makes the gains segmentation easier and more efficient.

Table 1: Freeman code features and their abbreviations

Region	Direct1	Direct2	Direct3	Direct4	Direct5	Direct6	Direct7	Direct8
Region1	Vz_{11}	Vz_{12}	Vz_{13}	Vz_{14}	Vz_{15}	Vz_{16}	Vz_{17}	Vz_{18}
Region2	Vz_{21}	Vz_{22}	Vz_{23}	Vz_{24}	Vz_{25}	Vz_{26}	Vz_{27}	Vz_{28}
Region3	Vz_{31}	Vz_{32}	Vz_{33}	Vz_{34}	Vz_{35}	Vz_{36}	Vz_{37}	Vz_{38}
Region4	Vz_{41}	Vz_{42}	Vz_{43}	Vz_{44}	Vz_{45}	Vz_{46}	Vz_{47}	Vz_{48}

2.2 Image database samples

A database of images was created from various samples of several cereal varieties obtained from different sources and for different crop years from laboratories of the Tunisian Cereal Office. Tunisian Hard Wheat (HW), Tunisian Tender Wheat (TW) and Tunisian Barley (B) are the main classes of the samples considered.

3 Classification features

For each grain type, 152 parameters are extracted from the colour images of the database (122 morphological, 18 colour, and 12 wavelet features).

3.1 Morphological features

After isolating the grain, the region of interest was selected around the boundary of the edge. The morphological features were obtained from the binary images containing only pixels of the grain edge. We can classify these features as follows:

- **Grain size measurements:** Length (L), width (l), width by length ratio (R_1), area (S), perimeter (P), area by perimeter ratio (R_2), angles (GrA, PtA) and radius of curvature (Rr, Rl) of the two extremities, likelihood between the grain and the nearest ellipsis for the grain (E), mean (Sx, Sy) and standard deviation (σ_x, σ_y) of horizontal and vertical symmetry.
- **Freeman code features:** After dividing the grain image in four regions as shown in the figure (1.a). We perform for every region the freeman code ([22]); it's the oldest contour descriptor and the most used today; it's mainly based on the position of the pixels set that are the nearest neighbours (NN-set) of the actual pixel. In fact, every region is coded starting from a given origin and according to the directions of the nearest neighbour that are represented in 8-connectivity (coded on 3 bits) as demonstrated in figure (1.b). The features extracted from the Freeman code are 32; eight for every region. These features are summed up in table 1.

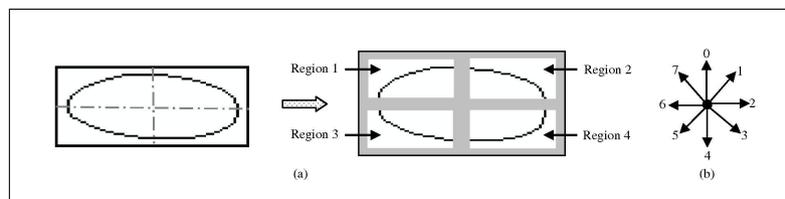


Figure 1: Freeman code extraction, (a) Dividing image in four regions to compute the Freeman code, (b) Direction codes

- **Fourier transform features:** The Fourier Transform is an important image processing tool which is used to decompose an image into its sine and cosine components ([23]). The application of

Table 2: List of colour parameters and their abbreviations

Colour Components	Mean value	Mean square value	Variance	Standard deviation	Kurtosis	Skewness
Red	RM_1	RM_2	RV	RSD	RM_3	RM_4
Green	GM_1	GM_2	GV	GSD	GM_3	GM_4
Blue	BM_1	BM_2	BV	BSD	BM_3	BM_4

this transform on the contour pixels creates a set of complex coefficients that represents the shape of the contour. From these coefficients we extract the morphological descriptors using different signatures (1).

$$a(u) = \frac{1}{N} \sum_{k=0}^{N-1} s(k) \exp \left[\frac{-j2puk}{N} \right] \quad (1)$$

Where:

$u \in [0, N-1]$ (N : number of points in contour)

$s(k)$: the signature chosen.

$a(u)$: harmonic descriptors.

The signatures used are complex, radial distance and polar. From each signature, we selected the first 25 harmonic coefficients that can be added to the set of the morphological features.

The three signatures used are invariant by translation and consequently their Fourier descriptors (FD), but it was proved that they are sensitive to rotation. Invariance by rotation is then realized by ignoring the FD phase and by considering only modules of these Fourier descriptors.

For the complex signature all descriptors except the first (DC component) are needed to index the form. The DC component describes only the contour position, and it is useless with the form description. The descriptors standardization consists in dividing their modules by the one of second descriptor. The vector which indexes the form is given by the (2).

$$F = \left[\frac{|FD_2|}{|FD_1|}, \frac{|FD_3|}{|FD_1|}, \dots, \frac{|FD_{N-1}|}{|FD_1|} \right] \quad (2)$$

The radial distance function and the polar coordinates are real. They have $N/2$ different frequencies for that half of the FD is necessary to index the form. The invariant vector (3) is obtained by dividing the module of the $N/2$ first descriptors by the module of the first descriptor.

$$F = \left[\frac{|FD_1|}{|FD_0|}, \frac{|FD_2|}{|FD_0|}, \dots, \frac{|FD_{N/2}|}{|FD_0|} \right] \quad (3)$$

3.2 Colour features

For each colour image that contains an isolated grain, we perform statistical parameters on values of pixels belonging to the grain. Color parameters included: Mean value, Mean square value, Variance, Standard Deviation, Kurtosis and Skewness of the Red, Green and Blue primaries. In table 2 we present these parameters and their abbreviations.

Table 3: List of wavelet parameters and their abbreviations

Matrix type	Average value	Variance	Standard deviation
Matrix of approximation image	<i>MVAP</i>	<i>VAP</i>	<i>SDAP</i>
Matrix of horizontal details	<i>MVHD</i>	<i>VHD</i>	<i>SDHD</i>
Matrix of vertical details	<i>MVVD</i>	<i>VVD</i>	<i>SDVD</i>
Matrix of diagonal details	<i>MVDD</i>	<i>VDD</i>	<i>SDDD</i>

3.3 Wavelet features

The wavelet analysis of an image is a multi resolution analysis which is defined by linear operators allowing analyzing a signal on various frequencies. Indeed, the signal is projected on a scale function that gives a representation of the original signal at higher scale. This projection causes a back zoom of the original signal, where the approximation is performed. In order to rebuild the signal, starting from approximation coefficients, we must also project the original signal on a wavelet to recover information lost during the first projection. The second projection contains the details of the original signal.

The details of wavelet features have been reported earlier in [20]. Table 3 resumes the chosen features. They were statistically tested to extract the best parameters leading to an optimal classification.

The tests done on the 12 parameters proved that only two parameters are judged like not-significant (ADH and ADD). Thus; the number of parameters which is going to be retained for the characterization phase is 10: SDAP, AAP, VAP, SDDH, VDH, SDDV, ADV, VDV, SDDD and VDD.

4 Classification methods

Starting from the classification features extracted, we developed many methods using different approaches. The first approach is a statistical classification method that uses only morphological and colour features. The second approach is a classification using a fuzzy logic based method. The third is a combination between the first and the second. The last approach is an artificial neural network classification method that exploits all features leading to the best classification result.

In what follows, we present these different approaches and their contribution to the classification of cereal grains.

4.1 Statistical classification method

From the set of samples, we achieved statistics related to morphological and color features extracted from color images of grains. From these statistics we obtained a distribution curve of every feature. This method operates directly on the distributions intervals of the morphological and color parameters. The classification is made by successive tests on parameters according to their ranks. Conceived algorithm has been tested on images containing a mixture of grains collected from treated samples.

To classify the grain types using a statistical method we considered the morphological and color features in this approach. Classification results for the grain types using this method are illustrated in figure 2.

We notice that the recognition rate for TW is weak while working with morphological or colour features (morph. 56%; color 51%). For HW, colour features gave an optimal recognition rate exceeding 99,4%, but does not exceed 67% when working with morphological features. For Barley grains, due to their form that is different from other types of grains, the morphological features gave us a good classification result reaching 98,7%. The global recognition rate for the statistical classification method

Table 4: Test of best parameters

Parameters	Barley	Hard wheat	Tender wheat	Total
Lsb	82,43%	77,02%	90,28%	80,94%
E	77,82%	20,75%	90,91%	46,28%
GrA	72,80%	70,57%	67,13%	68,62%
RM2	64,02%	50,52%	39,18%	50,25%

is limited to 76%. This is explained by the overlapping that exists between the distribution curves of grain classes when working with all the morphological and colour features.

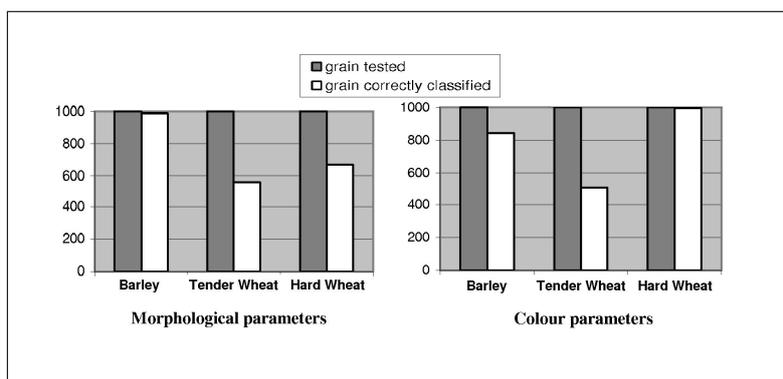


Figure 2: Results of the statistical classification method when applied to morphological and colour features

4.2 Fuzzy logic based classification method

Due to the overlaps of distribution curves of grains types we implement a classification method based on fuzzy logic techniques to improve the recognition rate issued from the statistical classification method. Classification using the fuzzy logic is made according to the following steps:

- Classes's definition.
- Generation of the membership functions for every parameter.
- Development of inference rules.
- Decision making.

It results three classes corresponding to the different grain types considered. Membership functions are deduced from the distribution curves of the different parameters of every grain type. The membership functions were conceived by normalization of the curves and then by a Gaussian approach for every curve. The number of rules depends on the number of parameters considered. The chosen norm is the max-prod. Then, the rules form is: "IF (condition1) AND (condition2) THEN (decision)".

The choice of entries is based on a test of identification parameters. Table 4 illustrates the test of the four best parameters for the classification from the set of morphological and colour features associated to the fuzzy logic method.

From the possible combinations of the four parameters we select the best ones according to its recognition rates. The combinations selected are illustrated (Lsb and GrA : 83,42%; Lsb, GrA and RM2 : 72,71%; Lsb, GrA, E and RM2 : 68,23%). From this test we chose the parameters Lsb and GrA since

when combining them it gives the best recognition rate. The result of this method using the combination Lsb and GrA is shown in figure 3.



Figure 3: Results for Fuzzy Logic based classification method

Concerning the hard wheat and tender wheat grains, this method gives us a best recognition rate than the statistical one. On the other hand for the barley grains, the first method is more reliable. So that, we opt to use a method which combines the two previous methods and gives the best recognition rate.

4.3 Statistical and Fuzzy Logic combined classification method

It consists in making a decision about the grain type by the fuzzy logic method in the cases where the statistical method cannot make a decision. The fuzzy logic is used in the combined method in the cases of overlaps of all morphological and colour parameters. The improvement concerns the hard wheat and tender wheat grains only; the barley grains possess an optimal recognition rate. The results of this method are illustrated in figure 4.

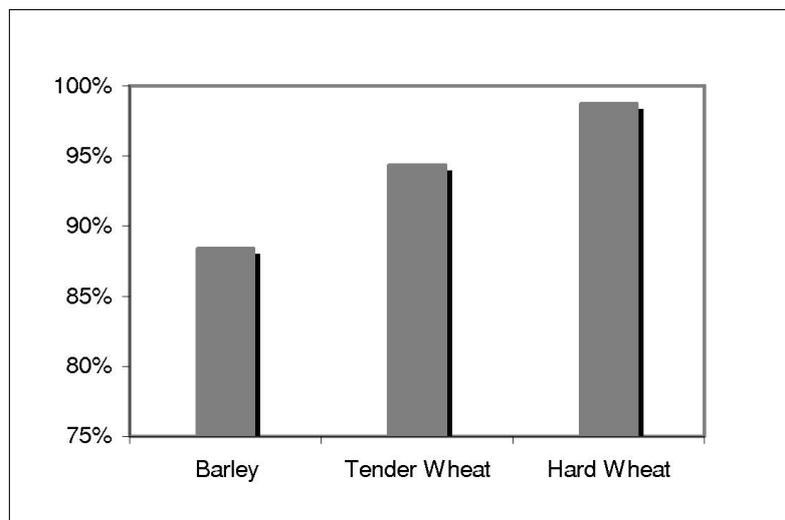


Figure 4: Results for Statistical and Fuzzy Logic combined classification method

Table 5: Number of neurons in the hidden layer

	Morphological features				Colour features	Wavelet features
	SM	FC	FT	AMF		
Number of features	15	32	25	122	18	10
Number of neurons	3	7	5	5	10	3

Table 6: Classification results for ANN classifier

Grain type	Rates (%)								
	Morphological features			Colour features			Wavelet features		
	CR	RJR	RCR	CR	RJR	RCR	CR	RJR	RCR
B	1,1	0	98,9	2,8	0	97,2	2,5	0,7	96,8
HW	1,6	0,5	97,9	7,6	0,7	91,7	1,7	1	97,3
TW	7,7	2,8	89,5	3,5	1,3	95,2	0	0	100
Mean rates	3,5	1,1	95,4	4,6	0,7	94,7	1,4	0,6	98

4.4 Artificial Neural network classification method (ANN)

Training phase and network architecture

The network architecture is a multi-layer neural network MLP. The training is done using the function "TRAINLM" from the Matlab neural network toolbox. Activation functions are hyperbolic tangent and the linear Matlab functions "tansig" and "purelin". During the training phase, we varied the neurons number in the hidden layer and we determine the training error. We chose 40000 as training iterations number since this value leads to a minimum training error. The number of neurons in the hidden layer depends of the type of features considered as entries of the network in the table 5 we illustrate the variation of the number of neurons in the hidden layer when using different types of features (for the morphological features SM means Size Measurements, FC : Freeman Code, FT : Fourier Transform and AMF : All Morphological Features) .

Classification results

For this test we used 3000 grains (1000 grains of each class), 600 grains for characterization and 400 grains for validation. The training of each class is done using 1800 grains, 600 grains will be used to learn the true membership and the others 1200 will be used to learn the system the false membership to the class. This technique seems to be very original and will make it possible to enlarge the classification space, to refine space collates and to reduce the conflict rate between various classes.

Thus this test will determine the conflict rate (CR), the rejection rate (RJR) and recognition rate (RCR). Table 6 represents the results obtained during the first test.

5 Evaluation and discussion

Figure 5 shows the classification recognition rates of the four developed methods. The ANN classifier lead to the best recognition rates for Barley (98,9%) using morphological features and Tender wheat (100%) using wavelet features whereas the statistical and fuzzy logic combined classifier was the best for Hard wheat classification (98,7%). These two methods gave better results than the first and second one.

The tables 7, 8, 9 and 10 present the confusion matrixes of the four developed classifiers. When we

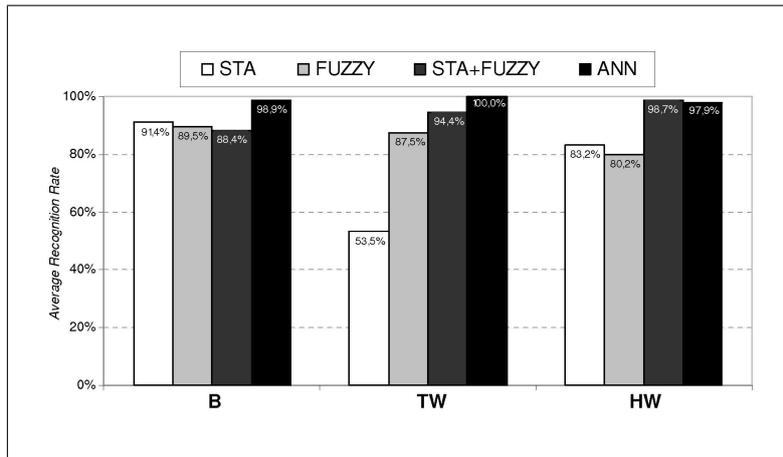


Figure 5: Comparison of the classification methods based on average recognition rates

Table 7: Confusion Matrix (%) for the statistical method

	B	TW	HW
B	91,4	1,5	7,1
TW	5,2	53,5	41,3
HW	2,9	13,9	83,2

observe these matrixes, we note that the major confusions are between Tender Wheat and Hard Wheat in the Statistical classification method (41,3% for HW and 13,9 TW), Fuzzy Logic classification based method (12% for HW and 15,7% for TW) and Statistical and Fuzzy Logic Combined classification method (5% for HW and 0,8% for TW) this is due to the similarities that exists in the morphology and the texture of these two cereal grain classes. This problem is resolved using the ANN classification method (0% for HW and 1,6% for TW).

Barley grains are more confused with Hard Wheat (STA: 7,1% ; FUZZY: 9,6% ; STA+FUZZY: 10,5% and ANN : 0,6%) than with Tender Wheat (STA: 1,5% ; FUZZY: 0,9% ; STA+FUZZY: 1,1% and ANN : 0,5%) this is due to the size that is larger than Tender Wheat and colour features.

To evaluate the time performance for each classification method we count the time in seconds that takes every algorithm to classify grains in a sample of 300 grains containing 100 grain of each type. The algorithms are developed on a Toshiba Satellite (Intel Core 2 / 1,6 Ghz) Laptop under Windows Vista environment. Table 11 presents the cost time for each method.

Table 8: Confusion Matrix (%) for the fuzzy logic based method

	B	TW	HW
B	89,5	0,9	9,6
TW	0,5	87,5	12,0
HW	4,1	15,7	80,2

Table 9: Confusion Matrix (%) for the statistical and fuzzy logic cobined method

	B	TW	HW
B	88,4	1,1	10,5
TW	0,6	94,4	5,0
HW	0,5	0,8	98,7

Table 10: Confusion Matrix (%) for the ANN method

	B	TW	HW
B	98,9	0,5	0,6
TW	0	100	0
HW	0,5	1,6	97,9

Table 11: Time performance of the different methods

Method	Time(s)
STA	72
FUZZY	43
STA+FUZZY	84
ANN	177

We have noticed that the Fuzzy Logic based classification method appears to run about 60% faster than the second fastest (Statistical classification method). The method leading to best recognition results is 4 times slower than the fastest methods. The Statistical and Fuzzy Logic combined classification method can be considered as the most performing method as it have a good recognition rate (94%) and take 50% less time than the method leading to the optimal recognition rate. While the reported execution times depend on the implementation language, we note that we have used Matlab 2007.

6 Conclusion

As dealt above, the classification of different grain types was successfully achieved using different parameters based on different types of features (morphological, colour and wavelet). These parameters were tested on different classification methods; the statistical classification method gave an average recognition rate of 76%. The second method based on fuzzy logic techniques gave an average recognition rate of 85,73%. The hybrid method, which is a combination of the two fore mentioned methods gave an average recognition rate of 93,83%. Finally, the ANN classification method was tested on all features and gave the best recognition rate reaching 98%.

Bibliography

- [1] M. Abdellaoui, A. Douik, M. Annabi, Détermination des critères de forme et de couleur pour la classification des grains de céréales, *Proc. Nouvelles Tendances Technologiques en Génie Electrique et Informatique, GEI'2006*, Hammamet, Tunisia, 2006, pp. 393-402.
- [2] D. A. Barker, T. A. Vouri, M. R. Hegedus, D. G. Myers, The use of ray parameters for the discrimination of Australian wheat varieties. *Plant Varieties and Seeds* 5(1) (1992) 35-45.

-
- [3] D. A. Barker, T. A. Vouri, M. R. Hegedus, D. G. Myers, The use of slice and aspect ratio parameters for the discrimination of Australian wheat varieties, *Plant Varieties and Seeds* 5(1) (1992) 47-52.
- [4] D. A. Barker, T. A. Vouri, M. R. Hegedus, D. G. Myers, The use of Fourier descriptors for the discrimination of Australian wheat varieties, *Plant Varieties and Seeds* 5(1) (1992) 93-102.
- [5] D. A. Barker, T. A. Vouri, M. R. Hegedus, D. G. Myers, The use of Chebychev coefficients for the discrimination of Australian wheat varieties, *Plant Varieties and Seeds* 5(1) (1992) 103-111.
- [6] P. D. Keefe. A dedicated wheat grain image analyzer, *Plant Varieties and Seeds* 5(1) (1992) 27-33.
- [7] H. D. Sapirstein, J. M. Kohler, Physical uniformity of graded railcar and vessel shipments of Canada Western Red Spring wheat determined by digital image analysis, *Canadian Journal of Plant Science* 75(2) (1995) 363-369.
- [8] J. Paliwal, N. S. Shashidhar, D. S. Jayas, Grain kernel identification using kernel signature, *Transactions of the ASAE* 42(6) (1999) 1921-1924.
- [9] S. Majumdar, D. S. Jayas, Classification of cereal grains using machine vision. I. Morphology models, *Transactions of the ASAE* 43(6) (2000) 1669-1675.
- [10] M. Neuman, H. D. Sapirstein, E. Shwedyk, W. Bushuk, Wheat grain colour analysis by digital image processing: I. Methodology, *Journal of Cereal Science* 10(3) (1989) 175-182.
- [11] M. Neuman, H. D. Sapirstein, E. Shwedyk, W. Bushuk, Wheat grain colour analysis by digital image processing: II. Wheat class determination, *Journal of Cereal Science* 10(3) (1989) 182-183.
- [12] X. Y. Luo, D. S. Jayas, S. J. Symons, Identification of damaged kernels in wheat using a colour machine vision system. *Journal of Cereal Science* 30(1) (1999) 49-59.
- [13] S. Majumdar, D. S. Jayas, Classification of cereal grains using machine vision. II. Color models, *Transactions of the ASAE* 43(6) (2000) 1677-1680.
- [14] S. Majumdar, D. S. Jayas, Classification of cereal grains using machine vision. III. Texture models, *Transactions of the ASAE* 43(6) (2000) 1681-1687.
- [15] S. Majumdar, D. S. Jayas, Classification of cereal grains using machine vision. IV. Combined morphology, color, and texture models, *Transactions of the ASAE* 43(6) (2000) 1689-1694.
- [16] J. Paliwal, N. S. Visen, D. S. Jayas, N. D. G. White, Comparison of a neural network and a non-parametric classifier for grain kernel identification, *Biosystems Engineering*, 85(4) (2003) 405-413.
- [17] N. S. Visen, D. S. Jayas, J. Paliwal, N. D. G. White, Comparison of two neural network architectures for classification of singulated cereal grains, *Canadian Biosystems Engineering* 46 (2004) 3.7-3.14.
- [18] M. Abdellaoui, A. Douik, M. Annabi, Hybrid method for cereal grain identification using morphological and color features, *Proc. 13th IEEE International Conference on Electronics, Circuits, and Systems*, (Nice, France, 2006), pp. 870-873.
- [19] R. Choudhary, J. Paliwal, D. S. Jayas, Classification of cereal grains using wavelet, morphological, colour, and textural features of non-touching kernel images, *Biosystems engineering* 99 (2008) 330 - 337.
- [20] A. Douik, M. Abdellaoui, Cereal varieties classification using wavelet techniques combined to multi-layer neural networks, *Proc. 16th Mediterranean Conference on Control and Automation*, (Ajaccio, France, 2008) pp1822-1827.

- [21] S. G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7) (1989) 674-693.
- [22] H. Freeman, On the encoding of arbitrary geometric configurations, *IEEE Trans on Electr. Comput.* 10 (1961) 260-268.
- [23] D. Zhang, G. Lu, A Comparative Study on Shape Retrieval Using Fourier Descriptors with Different Shape Signatures, *Proc. IEEE International Conference on Multimedia and Expo*, (2001), pp. 1139-1142.

Ali Douik was born in Tunis, Tunisia. He received the Master degree from the “Ecole Normale Supérieure de l’Enseignement Technique de Tunis”, in 1990 and the Ph.D. degree in Automatic from the “Ecole Supérieure des Sciences et Techniques de Tunis, Tunisia”, in 1996. In 2010, he received the ability degree from the “University of Monastir, Tunisia”. He is presently “Maitre assistant” in the “Ecole Nationale d’Ingénieurs de Monastir”. His research is related to Automatic Control and Image Processing.

Mehrez Abdellaoui was born in Tunis in 1979. He received his Electrical Engineering Diploma from Electrical Engineering Department in ENIM-Monastir in 2003 and the Master degree in Automatics from the ENIM-Monastir in 2005. He is currently a PhD student in the Electrical Engineering Department at the ENIM-Monastir. His research interests include Image Processing and Video Analysis.