Bulgarian Journal of Agricultural Science, 21 (No 1) 2015, 225-236
Agricultural Academy

GRAIN SAMPLE QUALITY ASSESSMENT FUSING THE RESULTS FROM COLOR IMAGE AND SPECTRA ANALYSES

M. I. MLADENOV[1], M. P. DEJANOV[2] and R. TSENKOVA[3]
[1] *University of Ruse, Department of Automatics and Mechatronics, BG – 7017 Ruse, Bulgaria*
[2] *University of Ruse, Department of Automatics and Mechatronics, BG - 7017 Ruse, Bulgaria*
[3]*Biomeasurement Technology Laboratory, Kobe University, Japan*

Abstract

MLADENOV, M. I., M. P. DEJANOV and R. TSENKOVA, 2015. Grain sample quality assessment fusing the results from color image and spectra analyses. *Bulg. J. Agric. Sci.*, 21: 225-236

The paper presents the approaches, methods and tools for assessment of main quality features of grain samples which are based on color image and spectra analyses. Visible features like grain color, shape, and dimensions are extracted from the object images. Information about object color and surface texture is obtained from the object spectral characteristics. The categorization of the grain sample elements in three quality groups is accomplished using two data fusion approaches. The first approach is based on the fusion of the results about object color and shape characteristics obtained using image analysis only. The second approach fuses the shape data obtained by image analysis and the color and surface texture data obtained by spectra analysis. The results obtained by the two data fusion approaches are compared.

Key words: grain quality assessment, color image analysis, spectra analysis, classification, data fusion

Abbreviations: ANN – Artificial Neural Networks; CA - Cluster Analysis; CDRBE - classifier with decomposing RBEs; CRBEP - classifier with RBEs which takes into account the class potentials; CSRBE - classifier with standard RBEs; CVS – Computer Vision System; HIS - Hyperspectral Imaging System7; INTECHN – Intelligent Technologies for Assessment of Quality and Safety of Food Agricultural Products; KNN - K-Nearest Neighbors; LDA – linear discriminant analysis; NIR spectra analyses – Near infrared spectra analyses; NIRS - A NIR spectroscopy; PCA - Principle Component Analysis; QDA - quadratic discriminant analysis; RBEs - Radial Basis Elements; SIMCA - Soft Independent Modeling of Class Analogy; SVM - Support Vector Machines; VIS spectra analyses – Visible spectra analyses; 1cc – first color class; 1csh – first shape class; 1cst – first class standard

Introduction

One of the main factors of human life quality is the food quality and safety. The food provides the energy, needed for the human body for movement, physical and intellectual activity. It is a source of proteins, fats, carbohydrates, vitamins and minerals, due to them the cells and tissues are renovated. As a result of the feeding the human organism produces hormones, enzymes and other regulators of the metabolic processes.

The assessment of food quality and safety is an important part of food production chain. The grain is a main part of the human and animal food. The higher food quality requirements demand development of new, objective, intelligent technologies, methods and tools for assessment of main food quality and safety features.

The grains and the cereals are an essential part of the human food. The cereals assure the half of the daily energy ration of the people in the developed countries and 80% in the developing countries. The grains of the wheat, maize, rice, barley, oats and millet contain about 60 - 80% carbohydrates, 8 - 15% proteins and 1.5 - 2% fats according to Emes et al. (2003).

The problem for rapid, objective, automated, express and nondestructive grain quality assessment is a complex and

Corresponding author: mladenov@uni-ruse.bg

multilevel task, related to the analysis of the appearance, the visible features as well as the grain contents, smell, flavor, moisture content, infections, non-grain impurities, etc. of the grain sample elements.

This investigation is focused on the assessment of main corn grain quality features. There are proposed and investigated methods and tools for feature extraction and data dimensionality reduction, analysis and identification of the grain sample elements. They are based on the analysis of color images and spectral characteristics of the investigated objects, as well as the fusion of the results of the two kinds of analyses. This approach is realized in the frame of INTECHN project DO 02-143/16.12.2008 (2008) "Development of Intelligent Technologies for Assessment of Quality and Safety of Food Agricultural Products" founded by Bulgarian National Science Fund.

According to the Bulgarian national standards the main quality features are as follow: the appearance, shape, color, smell, taste, moisture and impurities typical for the variety. Some of these features of a corn grain sample are presented in Table 1. The assessment of grain quality features presented in Table 1 is mainly related to the visible features of the grain sample elements and features related to the grain content, dry matter content, moisture content, starch, protein, glutenin, vitamins, toxins and mineral content according to Bulgarian Government Standard 607-73 (1973)**.**

It is obvious that all of the features mentioned above cannot be evaluated using the information extracted from one sensor source only. A huge part of these features (for example grain appearance, shape and color) are evaluated by an expert on the base of visual assessment only. That's why such features can be efficiently evaluated using a Computer Vision System (CVS). A review of the progress of computer vision in the agricultural and food industry is given by Brosnan and Sun (2003).

Some features like grain composition, content, infections etc. cannot be evaluated by means of CVS. Spectra analysis is mainly used for assessment of such features. Other features like moisture content, specific weight etc., are evaluated by other standard physicochemical methods. Acording to Mladenov (2011) obtaining a complex assessment of the grain quality using data about color, shape and dimensions of the grain sample elements is a complicated and multilevel task. This is because the color, the shape and the dimensions of the elements in a sample vary within a wide range.

There are many publications related to the assessment of some particular quality features using color image analysis. A digital image analysis algorithm based on the textural features is developed for classification of individual kernels of cereal grains (Mahesh et al., 2010). Color analyses are used to assess variety (Majumdar and Jayas, 2000a,b), infections (Ning et al., 1988; Mladenov et al., 2011b), germination (Mladenov and Dejanov, 2008a), weed identification (Aitkenhead et al., 2003), etc.

The grain variety is usually assessed by means of different morphological features related to the shape and geometrical parameters. A set of eight morphological features namely area, perimeter, length of major axis, length of minor axis, elongation, roundness, Feret diameter and compactness are used to recognize five different kinds of cereal grains (Paliwal et al., 2001). A broader investigation about classification of barley, Canada Western Amber Durum wheat, Canada Western Red Spring wheat, oats, and rye is presented in (Paliwal et al., 2003a). It is based on a total of 230 features (51 morphological, 123 color and 56 textural). A profile analysis through one-dimensional digital signals (Liao et al., 1993), by modeling the shape by means of a set of morphological features (Paliwal et al., 2003b) and by shape curvature analysis (Mladenov et al., 2011) is performed for assessment of grain purity. Computer vision methods are also used to determine kernel mechanical damage, mold damage (Ng et al., 1998), broken kernels in threshing process (Schneider et al., 1999), etc.

A preliminary investigation (Mladenov et al., 2011) shows that we can't get a precise assessment of some of the grain sample elements like smutty grains, infected grains and non-grain impurities, using an image analysis. It is difficult to detect the small changes of surface texture through a CVS. That's why we expect a more accurate assessment of such

Table 1
Corn grain quality groups

Grain quality groups	Grain quality features
First group - standard kernel	Whole grains and broken grains bigger than the half of the whole grain, with appearance, shape and color typical for the variety
Second group -grain impurities	Broken grains smaller than the half of the whole grain, heat-damaged grains, small grains, shriveled grains, green grains, sprouted grains, infected (with Fusarium) grains, smutty grains.
Third group – non-grain impurities	Corn-cob particles, leaf and stem fractions, pebbles, soil and sand, as well as harmful elements

features to be gotten by spectra analysis. Unfortunately information about shape and dimensions cannot be extracted from spectra.

Visible (VIS) and Near Infrared (NIR) spectra analyses are applied in the assessment different food products (Tsenkova et al., 2010; Nakakimura et al., 2012; Damyanov et al., 2006; Mladenov et al., 2011), as well as of grain quality features like grain composition, dry matter content, moisture content, starch, protein, glutenin, vitamins, toxins, mineral content, etc (Huang et al., 2008; Mladenov et al., 2011; Huang et al., 2008; Dowell et al., 2002; Girolamo et al., 2009). Different calibration models are developed for predicting grain composition and content (Peiris, 2001; Paulsen et al., 2003; Paulsen et al., 2004; Huang et al., 2008; Wesley et al., 2001; Miralbés et al., 2003).

Modified partial least squares models on NIR spectra (850 − 1048.2 nm) are developed to predict grain quality features (Paulsen et al., 2003). The best models are obtained for protein, moisture, wet gluten, and dry gluten with r^2 = 0.99, 0.99, 0.95, and 0.96, respectively.

The spectra analysis is also used for detection of different grain infections. Determination and prediction of the content of ergosterols and different kinds of mycotoxins like aflatoxin, fumonisin and others are very important tasks because mycotoxins are toxic for animals and humans. Reflectance and transmittance VIS and NIR spectroscopy are applied to detect fumonisin in single corn kernels infected with Fusarium verticillioides (Dowell et al., 2006). A method for determination of Fusarium graminearum infection is proposed in (Paliwal et al., 2003b). The classification accuracy reaches to 100% for individual samples. Transmittance spectra (500 to 950 nm) and reflectance spectra (550 to 1700 nm) are suggested as tools for aflatoxin determination in single whole corn kernels (Pearson et al., 2001). The authors use discriminant analysis and partial least squares regression for spectral data processing. The best results are obtained using two feature discriminant analyses of the transmittance data. A NIR spectroscopy (NIRS) method for estimation of sound kernels and Fusarium-damaged kernels proportions in grain and for estimation of deoxyinivalenol levels is proposed in (Peiris et al., 2010). The method classifies Fusarium damaged kernels with an accuracy of 99.9%. A neural network based method is developed for deoxynivalenol levels determination in barley using NIRS from 400 to 2400 nm (Ruan et al., 2002). Fourier transform NIRS is applied for rapid and non-invasive analysis of deoxynivalenol in durum and common wheat (Girolamo et al., 2009). A qualitative model for discrimination of blank and naturally contaminated wheat samples is developed. Classification accuracy of the model is 69% of the 65 validation samples.

A comparatively new approach for grain quality assessment is based on the Hyperspectral Imaging System (HIS). A HIS get data about object spectra at some regions (pixels) of the object area. Every pixel contains spectral reflection data for many narrow situated spectral bands usually in VIS and NIR spectrum. Spectral data is normally presented as a hyperspectral cube. HIS could be considered as a variant of a color image analysis, where the object image is divided into pixels, every pixel is analyzed using multiband spectral analysis instead of three band analysis (R, G, B).

The hyperspectral analysis is applied for assessment of different grain features. For example, Mahesh et al. (2010) use HIS for developing class models of different wheat varieties in Western Canada. The grain samples are scanned in NIR spectrum (960 – 1700 nm) at an interval of 10 nm. 75 different values of the intensity of the reflection are obtained from hyperspectral images and they are used for class model development. These models assure about 90% classification accuracy.

NIR spectroscopy is applied for assessment of grain moisture level too (Mahesh et al., 2008a,b; Mahesh et al., 2010). The authors present a new method using NIR hyperspectral imaging system (960 – 1700 nm) to identify five western Canadian wheat classes at different moisture levels. They are found that the linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) could classify moisture contents with classification accuracies of 89 – 91% and 91 – 99% respectively, independent of wheat classes. Once wheat classes are identified, classification accuracies of 90 – 100% and 72 – 99% are observed using LDA and QDA, respectively, when identifying specific moisture levels.

The HIS (350 – 2500 nm) is used for assessment of protein content in wheat grains (Wang, 2004) too.

Materials and Methods

Color image analysis. Grain groups and subgroups

Some features of grain sample elements, which are in principle evaluated by an expert on the basis of visual estimation, are assessed using CVS within the framework of this investigation. These features are related to the appearance, the color, the shape and the dimensions of the grain sample elements.

Groups (classes) and subgroups (subclasses) in which the corn grain sample elements are distributed are presented in Table 2. The tree normative quality classes are based on the corresponding color and shape subclasses presented in the same row of the Table 2.

Because the color and shape features are extracted and represented in a different manner, it is expedient their assessment to be made separately. After that the results from the two assessments have to be fused to obtain the final classification to one of the normative classes. Color and shape groups are divided in several subgroups in order to simplify the classification procedure. Color features are divided in 8 basic classes corresponding to the typical of different sample elements color zones and one additional class that corresponds to the non-grain impurities (it is impossibly to define a compact class for non-grain impurities). The sample elements are divided in 3 basic shape classes corresponding to the whole grains, broken grains bigger then the half of whole grain and broken grains smaller than the half of the whole grain, and one additional class that corresponds to the non-grain impurities. Each of the three basic shape classes is divided in 6 shape subclasses.

Features extraction from images

RGB, HSV, XYZ, NTSC and YCbCr color models are used for extracting the object area from background and for different color zones extraction in the frame of object area. These zones are typical of the standard grains, heat-damaged grains, green grains, smutty grains, infected (with Fuzarium) grains, bunt and non-grain impurities. Furthermore four color texture models (Mladenov, 2008b) are development for this purpose. It is expected they will better underline the difference between the color zones in the input RGB image.

Ten–dimensional descriptions are applied to represent the shape of the grain sample elements (Mladenov et al., 2011a,c). The following procedure is realized to obtain the shape description. First, the binary image of the object area is created. After that the object's peripheral contour is extracted and the bisection line of the object contour is found. An odd number of cross-sections perpendicular to the bisection line are built (Figure 1).

The relative length $h_i = s_i/D$ of the cross–sections, as well as the size and the sign of the difference between two neighbor cross–sections $\Delta_i = \Delta_1 - \Delta_2$ are calculated. Finally, the object shape description is presented in the following form:

$$X_{sh} = (h_1, h_2, \ldots, h_n, \Delta_1, \Delta_2, \ldots, \Delta_n) \qquad (1)$$

The contour line of the corn kernels has a huge asymmetry along to the bisection line. It is easy to locate the germ in the whole grain and to build contour descriptions and models with proper orientation. For broken grain, depending on what part of the whole grain is remained (with the germ or without

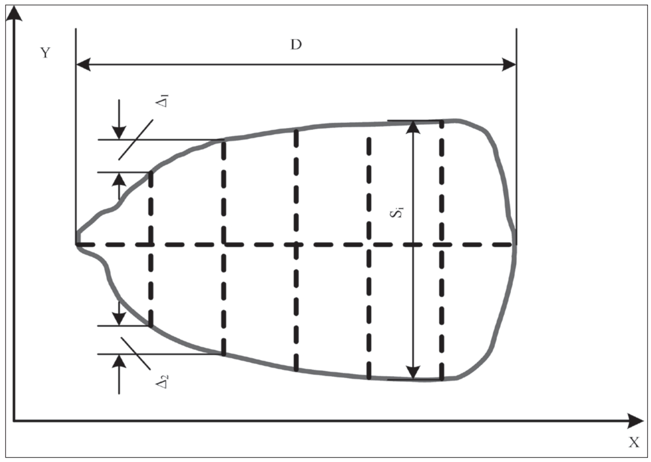

Fig. 1. Object shape description: D – length of the bisection line; $h_i = s_i/D$ – length of a cross–section; $\Delta_i = \Delta_1 - \Delta_2$ - difference between neighbor cross–sections

Table 2
Corn grain sample classes and subclasses

Normative (quality) classes	Color classes	Shape classes
1cst - standard kernel (whole grains and broken grains bigger than the half of the whole grain,) with appearance, shape and color typical of the variety	1cc- grains with color typical of the variety, back side	1csh- with typical of the whole grains variety shape
	2cc- grains with color typical of the variety, germ side	2csh- broken grains bigger than the half of the whole grain
2cst-grain impurities: broken grains smaller than the half of the whole grain, heat-damaged grains, small grains, shrivelled grains, green grains, sprouted grains, infected (with Fusarium) grains, smutty grains.	3cc- heat-damaged grains	3csh- broken grains smaller than the half of the whole grain and small and shrivelled grains
	4cc- green grains	
	5cc- mouldy grains	
	6cc- smutty grains	
	7cc- infected (with Fusarium) grains	
	8cc- sprouted grains	
3cst-non-grain impurities: corn-cob particles, leaf and stem fractions, pebbles, soil and sand, as well as harmful elements	9cc- non – grain impurities	4csh- non – grain impurities

it) the contour descriptions could be sufficiently different. It is necessary to define two types of descriptions and models for 2csh and 3csh classes (for its corresponding subclasses): for broken grain where the germ exists in the remaining part of the grain and where it does not exist. The all shape groups are divided into 18 subgroups (subclasses).

Spectra analysis. Grain groups and subgroups

The spectra analysis is used to evaluate the color and texture features of grain groups like Fusarium infected grains, shriveled grains, sprouted grains, smutty grains and non grain impurities. The color image analyses doesn't enable to obtained sufficiently precise assessment of such features.

The groups and subgroups in which the corn grain sample elements are distributed based on the spectra analysis are shown in Table 2.

Features extraction from spectra and data dimensionality reduction

Different methods like Principal Component Regression, Partial Least Squares Regression, Principal Component Analysis, Hierarchical Cluster Analysis and other methods are applied for developing a model to predict a property of interest, as well as for feature extraction and large and complex spectra data reduction. Methods like K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Cluster Analysis (CA), Support Vector Machines (SVM), Neural Networks (ANN), and Soft Independent Modeling of Class Analogy (SIMCA) are mainly used for assessment of different grain features using data from grain spectra.

The spectral characteristics are obtained using QE65000 spectrophotometer. Each characteristic is a vector with about 1500 components. Principle Component Analysis (PCA) and combination of Wavelet descriptions and PCA are applied for extracting typical features from object spectra and for spectral data dimensionality reduction. The Wavelet1(detail coefficients) and Wavelet2 (approximation coefficients) and the Haar, Daubechies2, Coiflet2, Symlet2 wavelet functions are used in this investigation. The level of decomposition is varied from m = 1 to m = 4. The most informative wavelet coefficients are chosen using PCA method.

Grain quality assessment fusing data from image and spectra analyses

Because the color and shape features are extracted and described in a different manner, the assessment of these characteristics is separately done. After that the results from the two assessments are fused in order to obtain the object's final categorization to one of the normative classes.

Different variants of data fusion schemes are developed at different stages of the study (Mladenov et al., 2011c), the schemes developed could be associated with hierarchical clustering algorithms. Their typical feature is that different criteria for class merging are used at different levels of data fusion.

***Variant 1* -** The First scheme uses a simplified fusion scheme. It is presented in Figure 2.

The input data (input classes) are separated in two groups – data about object color characteristics and object shape data. The first group consist of 10 color zones typical of the corn grains. These are regions of pixels in the grain image with similar color characteristics. The color zones were extracted from kernel images within the framework of a preliminary investigation.

The second group consists of 18 shape subclasses (1scsh, 2scsh,…, 18scsh). The first six of them correspond to different shape models of whole grains, the next six – to models of broken grains bigger than the half of whole grain and the last six – to models of broken grains smaller than the half of whole grain.

The color class (1cc, 2cc, …, 8cc) is determined on the basis of preliminary defined combinations of color zones at the first stage of fusing the results from the color analysis. The shape subclasses are merged into one of the three main shape classes (1csh, 2csh and 3csh).

At the second stage of the analysis the fusion of color and shape classes is made in order to form the final decision of object classification in one of the three normative classes (1cst, 2cst and 3cst). The assessment whether the shape of the object is typical of one of the three classes or not, is used as a fusion criterion for color classes 1cc to 5cc. For 6cc, 7cc and 8cc classes the shape is not important at all.

***Variant 2* -** The second scheme (Figure 3) uses color and combined topological models of typical color zones. The

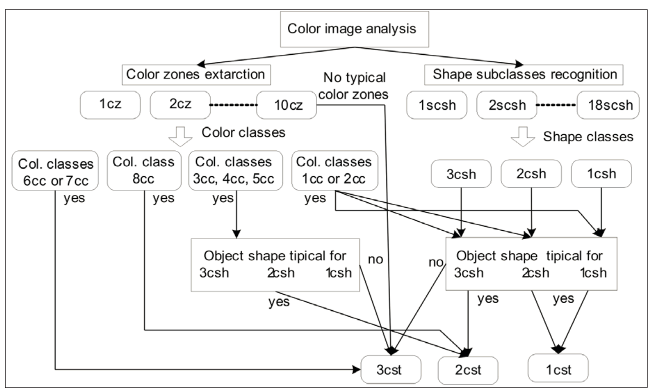

Fig. 2. Data fusion of the color and shape characteristics based on an image analysis only

topological models represent the plane distribution of the color zones within the object area. A set of color topology models (when 3 or more typical for the kernels color zones are found Figure 4a) and combined topology models (when only 2 typical color zones are found Figure 4b) is preliminary defined. The combined topological models represent the plane distribution of some shape element (kernel tipcap or crown region) and the color zones found.

The final categorization when such topology is found is performed on the base of the object area only. The object shape and the object area are important for the final categorization, when one typical color zone is found. For 6cc, 7cc and 8cc classes the shape is not important at all.

***Variant 3* -** The third scheme (Figure 5) fuses color characteristics extracted from spectra and shape characteristics extracted from images. This is because the color class recognition is more precise when we use spectra analysis instead of image analysis. The main criterion for final categorization is the object color class. The correspondence of the object shape and/or the object area to the typical for the grain sample elements shapes is an additional criterion.

The variants 1 and 2 of color and shape data fusion could be associated with the first level of multisensory data fusion (Direct fusion of sensor data (Liggins et al., 2008), we can consider the results from color and shape analyses as signals obtained from two different sensors (the shape and the color are extracted and presented in a different way).

The third variant of data fusion is a typical example of third level of multisensory data fusion (Decision level fusion). The color and shape data are obtained from two different sensors – spectrophotometer and RGB camera. This data is separately processed and after that the results are combined to get the final decision.

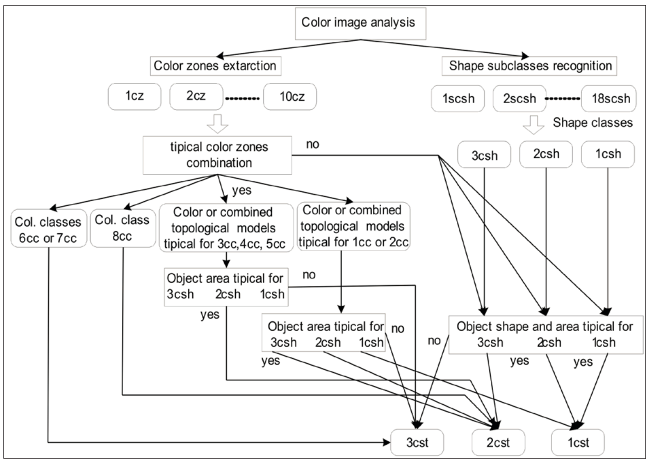

Fig. 3. Data fusion of color and shape characteristics based on an image analysis and grain topological models

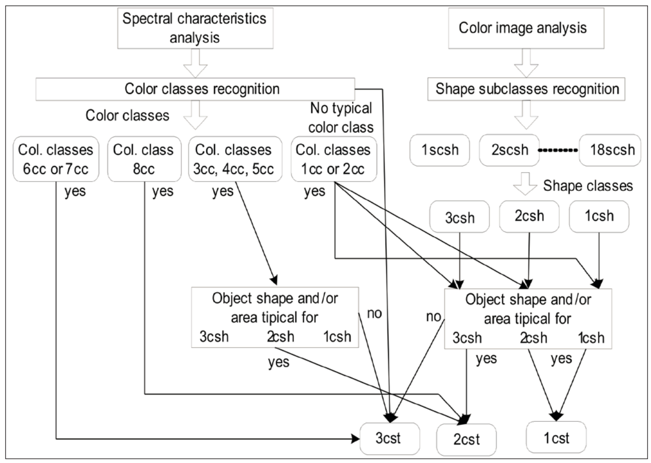

Fig. 5. Data fusion of color characteristics extracted from spectra and shape characteristics extracted from images

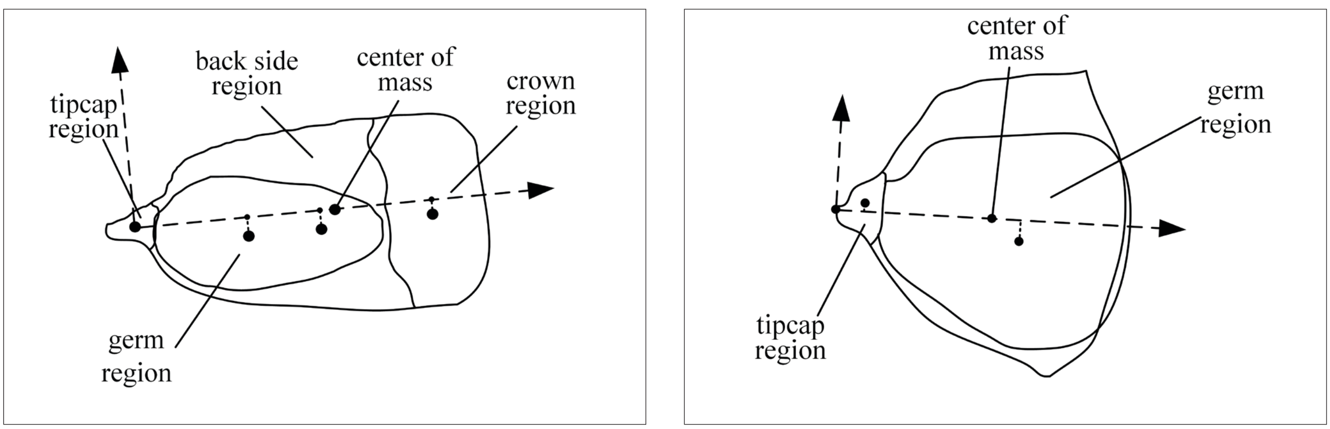

Fig. 4. Color (a) and combined (b) topology

It is expected that the data fusion procedures can improve the final classification in normative classes in comparison with the variants, when the categorization is made using object color data or object shape data only. This expectation is based on the assumption, that the fusion procedure can ignore or decrease some of the factors, which determine big errors of the final classification, when we use object color or object shape data only. The effectiveness of the proposed methods for data fusion is confirmed with the results presented in the tables.

Classification of the grain sample elements

Specific classification strategy, classifiers, and validation approach (Mladenov, 2011a,b) are applied for the categorization of the grain sample elements, which are conditioned by the specificity of the classification tasks.

Classification approach

If the classes (related with the color, shape, PCA and Wavelet + PCA descriptions) are presented in the feature space, a part of them (1cc, 2cc,…8cc) will form comparatively compact class regions. The sets of descriptions extracted from the grain sample training sets are used for developing the models of these grain sample groups. Each class model is presented by the class centre (the average value of the class training data) and the class boundary surface. The boundary surface is determined through a threshold value of the covariance of the class training data. A correct model for the 9cc class (non-grain impurities) could not be created because the characteristics of the elements of this class could be sufficiently different in each subsequent grain sample.

As a correct model for the 9th grain group could not be created, a part of the descriptions of such objects of the testing set could get into the boundaries of the other eight classes. A big part of them would get outside the class regions and could be located in a random place in the feature space. These descriptions could be considered as noisy vectors. It could be assumed that the comparatively compact class regions of the objects from the first eight groups are submerged in a noisy environment. Therefore the task for categorization of the grain sample elements can be interpreted as a task for classification in classes, whose boundaries have definite shapes, dimensions and location in the feature space, and they are situated in a noisy environment (Mladenov, 2011a).

Under this formulation, the use of popular strategies like LDA, CA, SVM, KNN and some other methods, which build boundaries between class regions, is obviously not a good choice. This is due to the fact that for the class 9cc a correct model cannot be created.

Furthermore if there are too big deviations of the actual values of the object characteristics and intensive measurement noise, the class areas can be overlapped. Very often correct information about prior probabilities of the classes is missing. This makes the classification problem more complex. If we use a classifier, which demands the prior probabilities to be known (for example Bayesian classifier), the training procedure has to be implemented using the prior probabilities obtained from the number of elements in training sets. When we assess quality of an unknown sample, the ratio of the number of elements from different classes could be sufficiently different from this ratio in the training sets. The classifier decision can be sufficiently different from the optimal decision under these circumstances. In this case the classification task is reduced to a task for approximation of overlapping class areas when the classes are situated in noisy environment and correct information for class a priori probabilities is missing.

Classifiers. The task for grain class modelling is reduced to a task for approximation of the boundaries of the grain class regions. For this purpose classifiers based on Radial Basis Elements (RBEs) could be used. Such classifiers will easily perform the approximation of the class regions and will simplify the classification procedure.

The following classifiers (Mladenov et al., 2011c) are used for class area approximation: CSRBE, CDRBE and CRBEP.

Classification accuracy

The accuracy of classification procedure is evaluated on the bases of the following classification errors:

$$e_i = \frac{FN_i}{(TP_i + FN_i)} \quad , \tag{2}$$

where e_i gives the relative part of objects from some class i, which are assigned incorrectly to other classes k = 1...N, where FN_i is the number of elements from the i^{th} class classified incorrectly to other classes, TP_i is the number of correctly classified elements from the i^{th} class;

$$g_i = \frac{FP_i}{(TP_i + FP_i)} \quad , \tag{3}$$

where g_i gives the relative part of objects from other classes, which are assigned to class i, where FP_i is the number of elements from other classes assigned to the i^{th} class;

$$e_o = \frac{\sum_{i=1}^{N} FN_i}{\left(\sum_{i=1}^{N} TP_i + \sum_{i=1}^{N} FN_i\right)} \quad , \tag{4}$$

where e_o (classification error rate) gives the relative part of all incorrectly classified objects, were N is the number of classes.

Test Setup

The hardware system consists of the following main components: computer vision system (CVS) (5) and spectrophotometer (4) (Figure 6). The CVS includes two color CCD cameras (1) which give a possibility to obtain color images of investigated object (2) in two planes (horizontal and vertical). The illuminant system (3) is used for direct object illumination. The reflectance spectral characteristics are obtained using spectrophotometer type QE65000 (4). The specifications of the camera, lenses, spectrophotometer and computer are presented in Table 3.

All investigations are carried out in laboratory conditions with a constant artificial lighting. Investigations in industrial environments are not made. The object color images are obtained in a dark room. The objects are separately placed on a color pad with a color different from the grain sample elements. It allows the object zone to be accurate extracted from the background.

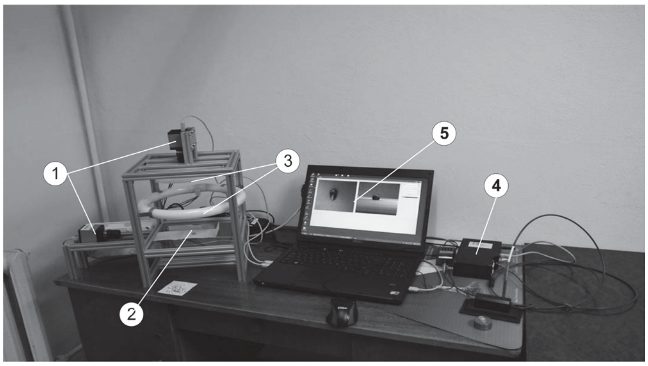

Fig. 6. Test Setup. 1- color CCD camera; 2- investigated object; 3- illuminant system; 4- spectrophotometer; 5- computer vision system (CVS)

Results and Discussions

Training and testing sets

The developed procedures for grain sample quality assessment are trained, validated and tested with sets presented in Table 4. The training and testing sets include a number of different elements (whole grains, broken grains, etc.)

The results from objects classification in shape and color classes are presented in Table 5. All classifiers described in the section "Classifiers" are used in the investigation. The "selected classifier" is the classifier with the best performance.

The results from objects classification in normative classes using the three variants of data fusion are presented in Table 6. Two classifiers are placed in the field "Selected classifiers". The first classifier is applied in color class recognition and the second classifier is used in shape class recognition.

- The test results in shape class recognition show that the rate of objects from class 1csh assigned to other classes is comparatively small (4.9%). On the other hand, the rate of objects assigned to this class which actually belong to other classes is sufficiently bigger (35.2%). The rate of objects from class 3csh assigned to 2csh and 4csh is big too.
- The classification error rate of objects from class 3csh (parts of kernels) is large. This is an expected result because it is impossible to define some standard shape for objects from

Table 3
Technical specifications of camera, lens, spectrometer, computer and lighting system

Name	Information
Color camera DFK 31AU03	Color digital video camera with USB interface, 1/3" Sony CCD sensor with progressive scanning, resolution – 1024x768 pixels;
Camera lens T2Z 3514 CS	Lens with variable focal length – 3.5 to 8 mm, diaphragm – from 1.4 to infinity, CS – assembly, MOD 0.3m;
Computer system Dell Vostro 1720	CPU - Intel Core 2 Duo P8700 (2.53 GHz, 3MB L2 Cache, 1066 MHz FSB) RAM - 4 GB (2x2048 GB) DDR2, 800 MHz Video card - NVIDIA GeForce 9600M GS 512MB
Spectrometer QE65000	Detector: Hamamatsu S7031-1006; Range: 350-1000 nm; Resolution: 1024 x 58 pixels; Optical resolution: ~0.14-7.7 nm FWHM; S/N ratio: 1000:1; ADC: 16 bits; Dynamic range: 7.5 x 109, 25000:1 for single measurement; Integral time: 8 ms to 15 min ; Adjusted linearity: >99.8%;
Lighting system Fluorescent lamps	The lighting system is compound of two ring-shape fluorescence sources with different diameters. It is used for direct illumination of investigated objects. They are placed so that the light can uniformly illuminate the object.

this class. In many cases even a qualified expert will not recognize such objects if no color characteristics but only shape is taken into consideration. During the classifiers training models of broken kernels are created on the basis of whole kernels models and that is why the training sample classification error rates for classes 2csh and 3csh are small. This explains the big difference between training and testing classification results for these two classes.

- The testing error in color class recognition using spectra analysis (7.3%) is acceptable bearing in mind the specific investigation conditions and the diversity of grain sample elements.
- The comparative analysis of the results obtained using different variants of classifier validation, training and testing confirms the effectiveness of the classification strategy, classifiers, validation approach and data models. For example, if we use the three data models: PCA, Wavelet1 + PCA and Wavelet2 + PCA the training errors are 6.8%, 6.3% and 10.3% respectively using the CDRBE classifier. The validation approach (when the non – grain impurities are included in validation procedure, but are excluded from training sets) decreases the testing error 3.8 times (from 27.6% to 7.3%) in comparison with the traditional validation approach (when the non – grain impurities are simultaneously excluded or included in validation and training sets). The choice of an appropriate classifier for specific classification task has an influence over the classification accuracy too.

Table 4
Training and testing sets

Color classes recognition using CVS							
Classes	1cc	2cc	3cc	5cc	7cc	8cc	9cc
Training sets	10	10	12	15	18	19	
Testing sets	47	81	44	24	74	39	168
Object shape recognition							
Classes	1csh	2csh	3csh	4csh			
Training sets	120	135	135				
Testing sets	122	63	11	256			
Color classes recognition using spectra analysis							
Classes	1cc	2cc	3cc	5cc	7cc	8cc	9cc
Training sets	120	120	80	53	192	42	536
Testing sets	30	30	20	13	48	11	134

Table 5
Classification errors in shape and color classes recognition

Color and shape class recognition						
	Shape class recognition		Color class recognition using CVS		Color class recognition using spectra analysis	
Selected classifiers	CRBEP		CDRBE		CDRBE Wavelet1+PCA model	
Errors	Test. errors		Test. errors		Test. Errors	
Class	g_i,%	e_i,%	g_i,%	e_i,%	g_i,%	e_i,%
1	39.5	5.7	8.9	8	0	15.2
2	62.0	69.8	8.3	4.9	0	5.1
3	87.5	77.8	0	15.9	0	0
4	21.9	43.1				
5			79.4	8.3	4.5	12.5
6						
7			35	13.5	19.5	5.4
8			0	5.3	0	10.3
9			23.2	68.4	12.0	6.6
	e_0=35.6%		e_0=30.6%		e_0=7.3 %	

For example, the training errors obtained using the CDRBE, CSRBE and CRBEP classifiers and PCA data model are 6.8%, 72%, and 7.3% respectively.

- The classification errors in color class recognition using spectra analysis are sufficiently smaller than the errors using image analysis. For example, the testing errors are 1.3% and 10% respectively using the two approaches when the non grain impurities are excluded from the validation and testing sets. When we include the non grain impurities in validation and testing sets these two errors are 7.3% and 42%. The big difference between the two errors can be explained by the fact that the object spectral characteristics contain not only information for objects color characteristics, but for their surface texture too. Although typical for some grain groups color zones are found in a big part of non – grain impurities, the surface texture of these elements is sufficiently different from the typical for the grains.
- Object classification in normative classes (1cst, 2cst and 3cst) includes complex assessment of color and shape characteristics of the investigated objects. For this purpose color data and shape data are fused. The data fusion procedure improves sufficiently the final classification results. The classification error rate e_0 in normative classes using CVS (Selected variant CDRBE–CRBEP) is 15.3% when data fusion Variant 1 is used and 8.6% when Variant 2 is used, while the errors of object color zones extraction and object shape recognition are 30.6% and 35.6% respectively. When we use Variant 3 for classes' recognition the classification error rate e_0 decreases about 1.6 times in comparison with the better result obtained using CVS. This is due to the fact that the spectra analysis gives the best result in color classes' recognition.

Conclusions

The results from the investigation at this stage of the INTECHN project implementation concerning grain sample quality assessment using complex assessment on the basis of color image and spectra analyses can be summarized as follow:

- The developed approaches, methods and tools for grain samples quality assessment based on the complex analysis of object color, object surface texture and object shape give an acceptable accuracy under specific experimental circumstances. The error rate e_0 = 5.3% of the final categorization in the normative classes can be accepted as a good result at this stage of project implementation.
- The data fusion procedure improves sufficiently the final classification results. The classification error rate e_0 using CVS is 15.3% when Variant 1 is used and 8.6% when Variant 2 is used, while the errors of object color zones extraction and object shape recognition are 30.6% and 35.6% respectively.
- According to the minimal quality requirements of grain samples defined in Bulgarian Standard Regulation 1272/2009 the impurities (grain class 2cs and non-grain 3cs) in a grain sample cannot be more than 12%. About 28% of standard grains are recognized as impurities using Variant 1. This means that the error is bigger than the permissible percentage of the impurities and this variant is not applicable for the analyses of real grain samples. The errors of the Variant 2 and 3 are permissible from the point of view of the Regulation 1272/2009. Variant 3 is preferred because the error (these are objects from third and second class recognized as objects from first class) is about 2 times smaller than the same error of the Variant 2.

The results obtained show that the choice of an appropriate procedure for fusion the results from color characteristics and objects shape analysis has a significant influence over the final classification accuracy. When we use the second algorithm (Variant 2) which is based on color or combined topology assessment the classification error rate decreases about 1.8 times compared to the first algorithm (Variant 1) in

Table 6
Classification errors in normative classes recognition

Color and shape data fusion						
Fusion variant	Variant 1		Variant 2		Variant 3	
Selected classifiers	CDRBE-CRBEP		CDRBE-CRBEP		CRBEP-CDRBE Wavelet1+PCA model	
Errors	Test. errors		Test. errors		Test. Errors	
Class	g_i,%	e_i,%	g_i,%	e_i,%	g_i,%	e_i,%
1cst	7.7	28.2	4.2	1.7	2.7	6.8
2cst	27.9	32.2	17.2	14.4	0.8	12.7
3cst	12.7	0.8	6.2	9.1	8.4	0.4
	e_0=15.3%		e_0=8.6%		e_0=5.3 %	

which color class assessment is based on the registration of the typical color zones combinations only. When we fuse the results from color classes recognition obtained on the basis of spectra analysis and shape classes recognition obtained on the basis of image analysis (Variant 3) the final classification accuracy is increased 2.9 and 1.6 times in comparison with Variant 1 and Variant 2 respectively.

Acknowledgements

This investigation is a part of implementation of the research project “Intelligent Technologies for Assessment of Quality and Safety of Food Agricultural Products”, funded by the Bulgarian National Science Fund.

References

Aitkenhead, M. J., I. A. Dalgetty and C. E. Mullins, 2003. Weed and crop discrimination using image analysis and artificial intelligence methods. *Computers and Electronics in Agriculture*, **39**: 157-171.

Brosnan, T. and D. W. Sun, 2003. Inspection and grading of agricultural and food products by computer vision systems - a review. *Computers and Electronics in Agriculture*, **36**: 193-213.

Bulgarian Government Standard, Grain maize for purchase and marketing. Pp. 607-73.

Damyanov, Ch. I., 2006. Non desatructive recognition of quality In system for food products automatic sorting. Monography. *Academic Publishing of the University of Food Technologies*, Plovdiv.

DO 02-143/16.12.2008. Project INTECHN - Intelligent Technologies for Assessment of Quality and Safety of Food Agricultural Products, 2008-2012, funded by the *Bulgarian National Science Fund*.

Dowell, F. E. E., B. Maghirang, F. Xie, G. L. Lookhart, R. O. Pierce, B. W. S. Seabourn, R. Bean, J. D. Wilson and O. K. Chung, 2006. Predicting Wheat Quality Characteristics and Functionality Using Near-Infrared Spectroscopy. *Cereal Chem.*, **83** (5): 529-536.

Dowell, F., T. Pearson, E. Maghirang, F. Xie and D. Wicklow, 2002. Reflectance and Transmittance Spectroscopy Applied to Detecting Fumonisin in Single Corn Kernels Infected with Fusarium verticillioides. *Cereal Chemistry*, **79** (2): 222-226.

Emes, M. J., C. G. Bowsher, C. Hedley, M. M. Burrell, E. S. F. Scrase-Field and I. J. Tetlow, 2003. Starch synthesis and carbon partitioning in developing endosperm. *J. Exp. Bot.*, **54** (382): 569-575.

Girolamo, A. D., V. Lippolis, E. Nordkvist and A. Visconti, 2009. Rapid and noninvasive analysis of deoxinivalenol in durum and common wheat by Fourier – Transform infrared spectroscopy. *Food Additives and Contaminants*, **26** (6): 907-917.

Huang, H., H. Yu, H. Xu and Y. Ying, 2008. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *Journal of Food Engineering*, **87** (3): 303-313.

Kos, G., H. Lohninger and R. Krska, 2003. Development of a Method for the Determination of Fusarium Fungi on Corn Using Mid-Infrared Spectroscopy with Attenuated Total Reflection and Chemometrics, *Anal. Chem.*, **75** (5): 1211-1217.

Liao, K., M. R. Paulsen, J. F. Reid, B. C. Ni and B. E. P. Maghirang, 1993. Corn kernel breakage classification by machine vision using a neural network classifier. *Transactions of the ASAE*, **36** (6): 1949-1953.

Liggins, II M. , D. Hall and J. Llinas, 2008. Handbook of Multisensor Data Fusion: Theory and Practice, Second Edition, *CRC Press*, 856 pp.

Mahesh, S., D. S. Jayas, J. Paliwal and N. D. G. White., 2008a. Identification of western Canadian wheat classes at different moisture levels using near-infrared (NIR) hyperspectral imaging. *CSBE Paper*, (08): 196.

Mahesh, S., D. S. Jayas, J. Paliwal and N. D. G. White, 2010b. Identification of wheat classes at different moisture levels using near-infrared hyperspectral images of bulk samples. *Sensing and Instrumentation for Food Quality and Safety*, **2** (3-4): 1007-1015.

Mahesh, S., A. Manickavasagan, D. S. Jayas, J. Paliwal and N. D. G. White, 2008a. Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes. *Biosystems Engineering*, **101** (1): 50-57.

Majumdar, S. and D. S. Jayas, 2000a. Classification of cereal grains using machine vision, Part 3: Texture models. *Transactions of the ASAE*, **43** (6): 1681-1687.

Majumdar, S. and D. S. Jayas, 2000b. Classification of cereal grains using machine vision: Color Models. *Transactions of ASAE*, **43** (6): 1677-1680.

Miralbés, C., 2003. Prediction Chemical Composition and Alveograph Parameters on Wheat by Near-Infrared Transmittance Spectroscopy. *J. Agric. Food Chem*, **51** (21): 6335-6339.

Mladenov, M., 2011a. Pattern classification in a noisy environment. *Information Technologies and Control*, (1): 23-33.

Mladenov, M. I., 2011b. Grain quality analysis and assessment. Monograph, *Academic publishing of the University of Ruse*, Ruse.

Mladenov, M. and M. Dejanov, 2008. Application of neural networks for seed germination assessment. Proceedings of the 9th WSEAS international conference on Neural Networks. pp. 67–72.

Mladenov, M., S. Penchev, M. Dejanov and M. Mustafa, 2011. Quality assessment of grain samples using colour image analysis. Proc. of тhe 8th IASTED International conference on Signal Processing, Pattern Recognition and Applications.

Mladenov, M., Ts. Draganova, P. Daskalov, R. Tzonev and M. Dejanov, 2008. Application of color and texture models and wavelet transformations in seed sowing properties assessment using computer vision. *Agricultural Engineering*, **6:** 2-7.

Nakakimura, Y., M. Vassileva, T. Stoyanchev, K. Nakai, R. Osawa, J. Kawano and R. Tsenkova, 2012. Extracellular metabolites play a dominant role in near-infrared spectroscopic quantification of bacteria at food-safety level concentrations. *Analytical Methods*, **4**: 1389-1394.

Ng, H. F. and W. F. Wilcke, 1998. Machine vision evaluation of corn kernel mechanical and mold damage. *Transactions of the ASAE*, **41** (2): 415-420.

Ning, S., R. Ruan, L., Luo, X. Chen, P. L. Chen and R. K. Jones, 1988. Automation of a machine vision and neural network based system for determination of scabby rate of wheat samples. *ASAE* Annual International Meeting, Orlando, Florida.

Paliwal, J., N. S. Visen and D. S. Jayas, 2003a. Comparison of a neural network and non-parametric classifier for grain kernel identification. *Biosystems Engineering*, **85** (4): 405-413.

Paliwal, J., N. S. Visen and D. S. Jayas, 2001b. Evaluation of Neural Network Architectures for cereal grain classification using morphological features. *Journal of Agricultural Engineering Research*, **79** (4): 361-370.

Paliwal, J., N. S. Visen, D. S. Jayas and N. D. G. White, 2003. Cereal grain and dockage identification using machine vision. *Biosystems Engineering*, **85** (1): 51-57.

Paulsen, M. R. and M. Singh, 2004. Calibration of a near-infrared transmission grain analyzer for extractable starch in corn. *Biosystems Engineering*, **89** (1): 79-83.

Paulsen, M. R., L. O. Pordesimo, M. Singh, S. W. Mbuvi and B. Ye, 2003. Corn starch yield calibrations with near-infrared reflectance. *Biosystems Engineering*, **85** (4): 455-460.

Pearson, T. C., D. T. Wicklow, E. B. Maghirang, F. Xie and F. E. Dowell, 2001. Detecting aflatoxin in single corn kernels by transmittance and reflectance spectroscopy. *Transactions of the American Society of Agricultural Engineers*, **44:** 1247-1254.

Peiris, K. H. S., M. O. Pumphrey, Y. Dong, E. B. Maghirang, W. Berzonsky and F. E. Dowell, 2010. Near-Infrared Spectroscopic Method for Identification of Fusarium Head Blight Damage and Prediction of Deoxynivalenol in Single Wheat Kernels, *Cereal Chem*., **87** (6): 511-517.

Ruan, R., Y. Li, X. Lin and P. Chen, 2002. Non-destructive determination of deoxynivalenol levels in barley using nea-infrared spectroscopy. *Applied Engineering in Agriculture*, **18** (50): 549-553.

Schneider, H. and H. Kutzbach, 1999. Determination of broken kernels in threshing process using image analysis and machine vision. In: Proceedings of 27th Symposium, Actual Tasks on Agricultural Engineering, Croatia.

Tsenkova, R., H. Meilina, S. Kuroki and D. Burns, 2010. Near infrared spectroscopy using short wavelengths and leave-one-cow-out cross-validation for quantification of somatic cells in milk. *Journal of Near Infrared Spectroscopy*, **17** (6): 345-352. 38

Wang, Z. J., J. H. Wang, L. Y. Liu, W. J. Huang, C. J. Zhao and C. Z. Wang, 2004. Prediction of grain protein content in winter wheat (*Triticum aestivum* L.) using plant pigment ratio (PPR). *Field Crops Research*, **90** (2-3): 311-321.

Wesley, I. J., O. Larroque, B. G. Osborne, N. Azudin, H. Allen and J. H. Skerritt, 2001. Measurement of gliadin and glutenin content of flour by NIR spectroscopy. *Journal of Cereal Science*, **34:** 125-133.

Received July, 8, 2014; accepted for printing December, 2, 2014.