Computers and Electronics in Agriculture 144 (2018) 260–268

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

An evaluation of utilizing geometric features for wheat grain classification using X-ray images

Małgorzata Charytanowicz[a,c,*], Piotr Kulczycki[b,c], Piotr A. Kowalski[b,c], Szymon Łukasik[b,c], Róża Czabak-Garbacz[d]

[a] Institute of Mathematics and Computer Science, Faculty of Mathematics, Informatics and Landscape Architecture, The John Paul II Catholic University of Lublin, PL 20–708 Lublin, Poland
[b] Division for Information Technology and Systems Research, Faculty of Physics and Applied Computer Science, AGH University of Science and Technology, PL 30–059 Cracow, Poland
[c] Systems Research Institute, Centre of Information Technology for Data Analysis Methods, Polish Academy of Sciences, Newelska 6, PL 01–447 Warsaw, Poland
[d] Department of Physiopathology, Institute of Rural Health, PL 20-090 Lublin, Poland

ARTICLE INFO

Keywords:
Grain classification
Principal component analysis
Factor analysis
Correlations
Morphological features
Image processing
X-ray imaging
Object recognition

ABSTRACT

Nowadays, with the rapid development of digital image processing, there has been a notable increase in elaborating advanced tools for studying the internal structure of objects. This may be very helpful in characterizing certain morphological traits of grains, as well as in quantifying the differences between them. The current research was carried out to study the structure of the traits and to determine their importance in relation to grain classification and identification. To achieve better performance and deeper understanding of their usefulness, the investigation was done by means of both principal component analysis and multivariate factor analysis. Herein, the percentage of variation explained by the first three factors reached a high 89.97%. Thus, the presented methodology supported reliable discrimination of the wheat varieties as regards their shape descriptors. The conducted study confirmed the practical usefulness and effectiveness of the evolved method when applied to the many practical tasks wherein the image analysis commonly employed in multivariate statistical methods is recommended.

1. Introduction

Recent advances in information technology and in digital image processing have resulted in the development and practical usage of computer-aided techniques in data analysis. The three-dimensional nature of computed tomography scanning allows the same object to be scanned multiple times and provides an opportunity to investigate its particle at any location within a sample. In the last few years, research studies indicate that X-ray computed tomography provides an alternate approach for the nondestructive measuring technique (Papadopoulos et al., 2009; Peth et al., 2010). This is very useful in characterizing the internal structure of objects and in quantifying their geometric features (Charytanowicz, 2014; Charytanowicz and Kulczycki, 2014; Czachor et al., 2015). In our research, the feature extraction method based on the utilization of X-ray images is proposed to measure several grain traits, hence, enabling their classification.

The identification of wheat grain requires some knowledge of their characteristics. High classification accuracy can be obtained by using kernel shape, color, length, and texture (Wiwart et al., 2012; Zapotoczny, 2011). They are considered as major distinctions and can be combined to construct the feature vector, which represents wheat grain. In the previous studies on wheat classification morphology, color, and texture were exploited for wheat variety recognition (Utku, 2000). Indeed, Majumdar and Jayas have suggested several different approaches for classification cereal grains using different types of features and their combinations (Majumdar and Jayas, 2000a, 2000b, 2000c, 2000d).

Various computer-aided systems based on morphological features for the classification have been reported in literature (Guevara-Hernandez and Gomez-Gil, 2011; Niewczas et al., 1995). The majority of different features have involved the identification of grain varieties. The key problem encountered in practice is a very large number of variables. Still, when the dimensionality of the data increases, classification problems become significantly harder. A high number of features can lead to lower classification accuracy. Moreover, the amount of computations required for classification increases exponentially with

* Corresponding author at: Systems Research Institute, Centre of Information Technology for Data Analysis Methods, Polish Academy of Sciences, Newelska 6, PL 01–447 Warsaw, Poland.
E-mail address: mchmat@ibspan.waw.pl (M. Charytanowicz).

https://doi.org/10.1016/j.compag.2017.12.004
Received 9 February 2017; Received in revised form 17 October 2017; Accepted 5 December 2017

the growth of data dimension. On the other hand, a reduction of the space dimensionality leads to a better understanding of a model and simplifies the usage of different visualization techniques (Camastra, 2003). A main problem is identifying a representative set of features from which a classification model for a particular task will be constructed. This addresses the problem of feature selection through a correlation based approach.

A logical attitude for categorizing the traits requires the use of multivariate statistical methods such as principal component analysis and factor analysis. These methods basically reduce a high number of variables to several components without a significant loss in classification accuracy. These approaches also enable the user to detect and explain correlations among variables. In addition, dimensionality reduction aims to reveal meaningful structures and guarantees to show the genuine properties of the original data (Janecek et al., 2008; Tian et al., 2010).

Vahid et al. (2011) has employed factor analysis to research the relationship in totality of some quantitative traits with regard to wheat grain yield under end drought stress via factorial split plot and on the basis of completely randomized block design in three replications. In this work, according to factor analysis, through decomposition to main components, four factors altogether explained 83.51% of all variations. The first factor was called the effective factor to yield. The second factor was that of effective traits to spike characteristics. The third factor was called the effective factor to plant height. Finally, the fourth factor was called the effective factor to plant growth. Leilah and Al-Khateeb (2005) have harnessed factor analysis and principal components analysis to study the relationship between wheat grain yield and its components under the drought conditions. Herein, three main factors accounted for 74.4% of the total variability in the dependent structure. The results showed that biological yield, harvest index, weight of grains/spike, spike length and number of spikes/m^2 had the highest communality, and, consequently, the high relative contribution in wheat grain yield.

Furthermore, the accomplished studies showed that the digital image processing techniques commonly applied in multivariate statistical analysis give reliable results in recognizing wheat varieties. In the previous paper (Charytanowicz et al., 2010), conducted in the earlier stage of the research study, an effective gradient clustering algorithm was proposed for wheat variety classification. This had resource to the basic geometric features, including the kernel area, kernel perimeter, compactness, kernel length, kernel width, asymmetry coefficient and length of kernel groove, common to particular wheat grain varieties. The presented procedure achieved an accuracy of about 92%. The utility of the investigated methodology in the context of discrimination by way of using the geometric features of wheat grain has been also confirmed by the results of discriminant analysis. In the work (Charytanowicz et al., 2016), the selected combination of geometric features, including the kernel perimeter, compactness, asymmetry coefficient, the ratio between the germ length and the kernel length, the ratio between the germ area and the kernel area, and the ratio between the kernel width and the kernel length, had significant contribution to the discrimination, and such features have permitted discriminant analysis to achieve a recognition rate of 89–96%. Moreover, in the paper (Kulczycki and Łukasik, 2014), the problems of reducing data set dimension and size were investigated. After the reduction process, the number of wheat grains assigned to the right class was very high – achieving almost 90%. Finally, the data set of wheat grain was used to verify classification quality results for the probabilistic neural network with simplified structure (Kowalski and Kusy, in press). The obtained results confirmed the practical usefulness of the proposed methodology.

Thus, this paper demonstrate the utilization of grain geometric traits in wheat variety recognition. The main objective of this work is to determine a basic set of these parameters with respect to wheat grain morphology. Both principal component analysis and multivariate factor analysis are exploited to better comprehend the relations between traits, as well as to identify effective factors in wheat grain classification.

2. Materials and methods

The research was conducted at the Institute of Agrophysics of the Polish Academy of Sciences in Lublin. For this work, we chose combine harvested wheat grain of three varieties: Canadian, Kama, and Rosa. Herein, the relationship between the different grain traits and their effectiveness in grain classification were studied by way of utilizing two multivariate statistical procedures which can be employed for structure detection: principal component analysis and factor analysis (Basilevsky, 1994; Jolliffe, 2002; Morrison, 2005). Furthermore, the Pearson correlation analysis was carried out for all analyzed variables (Draper and Smith, 1981). The significant differences between mean values were tested by analysis of variance and the Tukey's test.

2.1. Data acquisition

In our work, image processing methods were used to acquire the data. In order to evaluate the quantitative traits of wheat grains, a high quality visualization of the internal kernel structure was done through the application of a soft X-ray technique. This is an objective, precise and nondestructive method that is considerably cheaper than other more sophisticated techniques such as magnetic resonance imaging, scanning microscopy or laser technology. For each X-ray exposure, grain kernels were evenly positioned groove down. The images were obtained in the form of photograms at the scale of 5:1.

The photograms were then scanned by way of an Epson Perfection V700 table photo-scanner that was equipped with a transparency adapter, at 600 dpi resolution and 8 bit gray scale levels. This produced bitmap graphics files with a sufficient resolution for reflecting distinct features important for the proper characterization of objects.

Fig. 1 presents the exemplary X-ray images of these kernels for each studied variety: Canadian, Kama, and Rosa.

However, sole visualization of the kernels did not provide quantitative measures of shape parameters and their relations. In order to carry out accurate grain traits measurements, the specialized image processing package Grains (Niewczas and Wozniak, 1999; Strumillo et al., 1999), which incorporates image processing algorithms, was employed for measuring the particular characteristics of any selected grain. Using the commands available in the program menu, automatic boundary detection and diverse measurements relevant to the study were enabled for each individual kernel. In our research, to evaluate the factors affecting the grain differentiation, the following traits were measured: the kernel area, kernel perimeter, compactness, asymmetry coefficient, kernel length, kernel width, length of kernel groove, germ area, germ length, and additionally the ratio between the germ length and the kernel length, the ratio between the germ area and the kernel area, as well the ratio between the kernel width and the kernel length.

Compactness is a shape descriptor computed according to the formula:

$$C = 4\pi \times \frac{A}{P^2} \quad (1)$$

where A denotes the kernel area and P denotes the kernel perimeter. The maximum value of the compactness is equal to one and is taken for a circle. For values close to zero, the shape is increasingly elongated.

The asymmetry coefficient given by the formula

$$AC = \frac{|A_{left} - A_{right}|}{A} \quad (2)$$

is the ratio of two quantities: the absolute value of the difference between areas of the left and right part of a kernel, and the total area of that kernel.

Finally, the obtained data incorporated twelve real-valued