Food Control 94 (2018) 233–240

Contents lists available at ScienceDirect

Food Control

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

Development of an automated method for the identification of defective hazelnuts based on RGB image analysis and colourgrams

A. Giraudo[a], R. Calvini[b], G. Orlandi[b], A. Ulrici[b], F. Geobaldo[a], F. Savorani[a,*]

[a] Department of Applied Science and Technology, Polytechnic of Turin, Corso Duca degli Abruzzi 24, 10129, Torino, Italy
[b] Department of Life Sciences, University of Modena and Reggio Emilia, Padiglione Besta c/o Campus San Lazzaro - Via Amendola 2, 42122, Reggio Emilia, Italy

ARTICLE INFO

Keywords:
Hazelnut
Defect detection
Multivariate image analysis
Colourgrams
Classification
Variable selection

ABSTRACT

Over the past decades, Red-Green-Blue (RGB) image analysis has gained increasing importance in industrial applications, since it has widely proved to be a suitable tool for food quality and process control.

This article describes the development of a fast and objective method for the automated identification of defective hazelnut kernels based on multivariate analysis of RGB images.

To this aim, an overall sample set of 2000 half-cut hazelnut kernels, previously assigned by industrial expert assessors as sound or defective (i.e. rotten or pest-affected), was collected and imaged using a digital camera. The colour-related information of the images was converted into one-dimensional signals, named colourgrams, which were firstly explored through the Principal Component Analysis and subsequently used to build classification models, based on both Partial Least Square-Discriminant Analysis (PLS-DA) and interval-PLS-DA (iPLS-DA) algorithms.

A tree-structure hierarchical classification approach has been considered, i.e. the discrimination between sound and defective kernels as a first rule, and the discrimination between the two types of defect as a second rule. The best sound vs defective classification model was able to correctly recognize approximately the 97% of the test set defective samples, while the best rotten vs pest-affected model allowed classifying correctly more than 92% of the test set samples.

Moreover, the image reconstruction performed using the selected colourgram features led to an exhaustive interpretation of the decision-making criteria adopted by the classification algorithms and further confirmed the reliability of the proposed method.

1. Introduction

Hazelnut (*Corylus avellana* L.) is one of the most important crops in the world, with a yearly world's production of approximately 835900 tonnes of in-shell hazelnuts. Overall, 66.2% of the whole production (553000 tonnes) is provided by Turkey, while Italy is the second world's producer, covering about 10.9% of the global nut production (91113 tonnes). Minor but not negligible hazelnut producers are United States of America (36327 tonnes), Georgia (33933 tonnes), Azerbaijan (30207 tonnes), China (23173 tonnes) and Iran (20756 tonnes), (average production 2012–2014, FAOSTAT, 2017).

Over years, hazelnuts have achieved great importance due to their high organoleptic characteristics and rich nutritional composition. In addition, hazelnuts are widely used as ingredients in the confectionery industry for the preparation of baked and chocolate-based products (Moscetti, Frangipane, Monarca, Cecchini, & Massantini, 2012).

However, the perishable nature of this raw material entails potential problems in terms of quality degradation, which may occur directly in field during growing or later, during harvesting and storage (Ozay, Seyhan, Pembeci, Saklar, & Yilmaz, 2008). For these reasons, an appropriate quality evaluation of incoming batches of hazelnut kernels plays a paramount role among the different stages of hazelnut processing performed at the industrial level. Possible defects may include nut kernels with abnormal shape features (shrivelled or double-seeded) and primarily anomalous colour features. The most common defects of hazelnuts, affecting the kernels' colour among other quality features, are represented by two main categories: *rotten* and *pest-affected* kernels.

The first kind of defect is due to the contamination by some species of fungi and bacteria that, using specific enzymes, are able to break down the chemical bonds of carbohydrates and promote the hydrolysis of lipids in free fatty acids, as well as the autocatalytic oxidation of unsaturated fatty acids, while promoting the growth of the contaminant

* Corresponding author.
E-mail address: francesco.savorani@polito.it (F. Savorani).

https://doi.org/10.1016/j.foodcont.2018.07.018
Received 28 April 2018; Received in revised form 11 July 2018; Accepted 12 July 2018
Available online 17 July 2018
0956-7135/

organism itself. These chemical reactions often result in an extended browning of the surface of hazelnut kernels and in the development of rottenness, moulds and off-flavours.

On the other hand, the presence of pest-affected kernels is caused directly in field by the bite of plant pests, which become a potential vehicle for micro-fungal contamination (Moscetti et al., 2014). In addition, the pest's saliva contains enzymes responsible for proteinase, lipase, esterase and amylase activity. The bite of the insect usually results in localized brown spots that develop on the boundary regions of the hazelnut kernel, together with the development of intense off-flavours (Vaccinio, Guidone, Corbellini, & Tavella, 2008).

At the industrial level, control procedures of hazelnut quality are currently based on visual inspection and evaluation, which is performed by expert assessors on a representative sample set collected from each incoming batch of hazelnut kernels in order to take a reliable decision about its acceptance or rejection. Unfortunately, this control procedure has several drawbacks since it is extremely time consuming and labour intensive. In addition, the evaluation must be performed by experienced assessors and it is operator-dependent, which implies that this method is not easily transferable between different laboratories or industries. Another big problem arises from the inconsistency of the human eye, which can lead operators to provide a different evaluation for the same sample over time (Paulus, Busscher, & Schrevens, 1997).

In this context, the development of cheap and automated systems able to manage aspect-related data attracts a great deal of the industrial interest, since it can provide fast, objective and reproducible information about the appearance of a specific product (Lo Fiego et al., 2007). For these reasons, over the past decades Red-Green-Blue (RGB) image analysis has gained an increasing importance in industrial applications. In fact, RGB imaging allows to perform a detailed analysis of each pixel of an image, giving information about both local and global colour features of a sample. In this manner, RGB image analysis is particularly suitable for the evaluation of colour-related aspects of highly heterogeneous matrices, like most of the food products (Cubero, Aleixos, Moltó, Gómez-Sanchis, & Blasco, 2011; Prats-Montalbán, de Juan, & Ferrer, 2011; Wu & Sun, 2013). Some research works reported the use of RGB image-based methods for the detection of external defects in food products, like fruits (López-García, Andreu-García, Blasco, Aleixos, & Valiente, 2010), vegetables (Barnes, Duckett, Cielniak, Stroud, & Harper, 2010; Riquelme, Barreiro, Ruiz-Altisent, & Valero, 2008) and meat (Barni, Cappellini, & Mecocci, 1997; Marty-Mahe, Loisel, & Brossard, 2003). All the above-mentioned research works proposed ambitious and, in some cases, rather complex algorithms to extract, from RGB images, the features related to the external defects of the food product under investigation.

However, when dealing with practical applications of image-based quality control, it has to be considered that a great number of samples must be acquired in order to develop a robust and reliable prediction model. In this situation, a massive computational effort is required to simultaneously elaborate the information included in each pixel of all the acquired images. Furthermore, when each image refers to a specific sample, it is extremely important to achieve a so-called "object-based" or "image-based" prediction able to take a decision by considering the sample as a whole, rather than focusing on the single pixels (Calvini, Foca, & Ulrici, 2016; Kucheryavskiy, 2013).

In this context, Antonelli et al. (2004) developed an image-based analytical approach able to handle large datasets of RGB images by converting each image into a one-dimensional signal, named *colourgram*. Basically, this approach consists in describing the colour content of each analysed image through the frequency distribution curves of the R, G and B channels, and of some colour-related quantities derived from the R, G and B values. In the original proposal by Antonelli et al., the resulting colourgram is a one-dimensional signal with length equal to 4900 points.

The versatility of this approach arises from the possibility to simultaneously investigate many different colour-related properties without *a priori* choosing the features of interest based on a specific analytical problem. Moreover, it has to be underlined that the 4900 points of the colourgram are anyway a small number if compared to the millions of data included in a typical RGB image. Therefore, the conversion of the image into a colourgram leads to a sizeable data compression, which can be further improved using feature selection approaches, without losing the colour-related information of interest contained in the original image. As a consequence, also all the subsequent steps of data elaboration will require lower computational efforts compared to the analysis of the entire RGB image.

The colourgram-based approach has been already applied to different case studies in the frame of food quality control, leading to very promising results. In Antonelli et al. (2004), the combined use of colourgrams and feature selection/classification algorithms (Cocchi, Seeber, & Ulrici, 2001) allowed classifying different brands of 'Pesto' Italian sauce. In Foca, Masino, Antonelli, and Ulrici (2011), multivariate calibration models were developed using colourgrams and a feature selection/calibration algorithm (Cocchi, Seeber, & Ulrici, 2003) in order to predict the content of pigments, such as chlorophylls, β-carotene and pheophytins, once again in pesto sauce. Ulrici, Foca, Ielo, Volpelli, and Lo Fiego (2012) proposed a flexible approach, based on colourgrams and multivariate data analysis, to detect the red skin defect of raw hams. The most recent application of this approach was presented by Orlandi, Calvini, Foca, and Ulrici (2018) and refers to the quantification of defective maize kernels.

In this context, the present study aims at developing a fast and automated colourgram-based system for the detection of defective hazelnuts. Digital RGB images were acquired on a representative set of hazelnut kernels previously classified by industrial expert assessors into the three reference categories, i.e. 'sound', 'rotten' and 'pest-affected'. The RGB images were converted into colourgrams, which were first explored by means of PCA and then used to develop Partial Least Square Discriminant Analysis (PLS-DA) (Wold, Ruhe, Wold, & Dunn, 1984) classification models. Moreover, interval Partial Least Square Discriminant Analysis (iPLS-DA) (Nørgaard et al., 2000) feature selection algorithm was used to select the most informative regions of the colourgram signal aimed at better discriminating the defined classes. Finally, some RGB images, representative of the three reference hazelnut categories, were reconstructed using only the selected colourgram features, in order to investigate the criteria that underline the choices performed by the iPLS-DA algorithm.

2. Materials and methods

2.1. Hazelnut samples

The samples used in this study were kindly provided by Soremartec Italia s.r.l. and presented the same features of the hazelnut kernels delivered, in bags, to the confectionery industry, i.e. without the external shell, with the episperm and dried until reaching the reference commercial moisture of < 6% w/w. Following the standard procedures traditionally adopted for quality control at the industrial level, the nut kernels were half-cut lengthwise and carefully inspected one by one by expert assessors, who assigned each one to the right category. An overall sample set of 2000 representative half-cut kernels was collected, paying attention to span as much as possible the intrinsic variability associated to each reference category.

2.2. Image acquisition

The RGB images were acquired using a DigiEye System (VeriVide Limited, UK), which include a digital camera Nikon D7000 (Nikon Corporation, Tokyo, Japan), a cubic box with a 40 × 50 cm field of view, where the sample is introduced to be imaged, and a software for image acquisition and management. The digital camera was equipped with a 35 mm focal length and the images were acquired using 1/5 s