Microchemical Journal 113 (2014) 10–16

Contents lists available at ScienceDirect

Microchemical Journal

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

Classification of edible vegetable oil using digital image and pattern recognition techniques

Karla Danielle Tavares Melo Milanez, Márcio José Coelho Pontes *

Universidade Federal da Paraíba, Departamento de Química, João Pessoa, PB, Brazil

ARTICLE INFO

Article history:
Received 5 August 2013
Received in revised form 26 September 2013
Accepted 28 October 2013
Available online 5 November 2013

Keywords:
Edible vegetable oil
Digital image
Linear discriminant analysis
Variable selection

ABSTRACT

This work presents a simple and non-expensive method based on digital image and pattern recognition techniques for the classification of edible vegetable oils with respect to the type (soybean, canola, sunflower and corn) and the conservation state (expired and non-expired shelf life). For this purpose, vegetable oil sample images were obtained from a webcam and the frequency distribution of color indexes in the red–green–blue (RGB), hue (H), saturation (S), intensity (I), and grayscale channels were obtained. Linear discriminant analysis (LDA) was employed in order to build classification models on the basis of a reduced subset of variables. For the purpose of variable selection, two techniques were utilized, namely the successive projection algorithm (SPA) and stepwise (SW) formulation. For the study evolving the classification with respect to oil type, LDA/SPA and LDA/SW models achieved a correct classification rate (CCR) of 95% and 90% respectively. For the identification of expired and non-expired samples, LDA/SPA models were found to be the best method for classifying sunflower, soybean and canola oils, achieving a CCR in the overall data set of 97%, 94% and 93%, respectively, while the LDA/SW correctly classified at 100% for corn oil data. These results suggest that the proposed method is a promising alternative for the inspection of authenticity and the conservation state of edible vegetable oils. As advantages, the method does not use reagents to carry out the analysis and laborious procedures for chemical characterization of the samples are not required.

1. Introduction

Edible vegetable oils are used as salad oils, cooking oils, liquid and solid shortenings, spreads and ingredients in several foods, including bakery products and fried foods [1]. Beneficial and adverse properties for human health depend on the raw material utilized, methods of processing, condition of storage and shelf life. In fact, lipid autoxidation and inadequate storage contribute significantly to the deterioration and reduction of the shelf life of vegetable oils causing changes in color, texture, odor and flavor, loss of vitamins, and damage to proteins [2]. Thus, appropriate quality control is required to ensure the distribution of products within the specification to final consumers.

Several works have been reported in literature exploring the use of gas chromatography [3], liquid chromatography [4], nuclear magnetic resonance [5], Fourier transform infrared [6], square wave voltammetry [7], and differential scanning calorimeter [8], among others [9] for quality control of vegetable oil. However, most of the methods are highly-sophisticated, expensive and laborious.

Nowadays, digital image is becoming more important because of its ability to perform fast and non-invasive low-cost analysis on foods. In fact, a wide variety of digital cameras and digitalization equipment has contributed to increase the number of papers published exploring the use of webcam [10], scanner [11], cell phones [12] and digital camera [13] to monitor the quality of several food samples. Moreover, another positive feature of using digital image to monitor the quality of foods is that it replaces the human visual system, often employed in these types of analyses [14]. Therefore, the use of digital image eliminates the subjective character of analyses as well as the dependence of the human visual system, which is substantially influenced by ambient conditions and subject to inconsistencies [15].

The color models' main objective is to specify color in a standard way [16]. The models most commonly used for classification are the RGB (R — red, G — green and B — blue), the HSI (H — hue, S — saturation and I — intensity) and the grayscale [17,18].

The RGB model is based on the mechanism of color formation in the human eye, where combinations, at different levels, of light radiation at red, green and blue wavelengths provide radiation of different colors [19]. In that mechanism phenomena of absorption and reflection of light are involved, since the colors perceived by human eyes in an object are related to the nature of the light reflected by it [16].

The HSI system is more representative of the way humans perceive colors, and sometimes it is also more convenient for image processing. In this alternative color space, hue represents the type

* Corresponding author at: Universidade Federal da Paraíba, Departamento de Química, Laboratório de Automação e Instrumentação em Química Analítica/Quimiometria (LAQA), CEP 58051-970, João Pessoa, PB, Brazil. Tel./fax: +55 83 3216 7438.
E-mail address: marciocoelho@quimica.ufpb.br (M.J.C. Pontes).

0026-265X/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.microc.2013.10.011

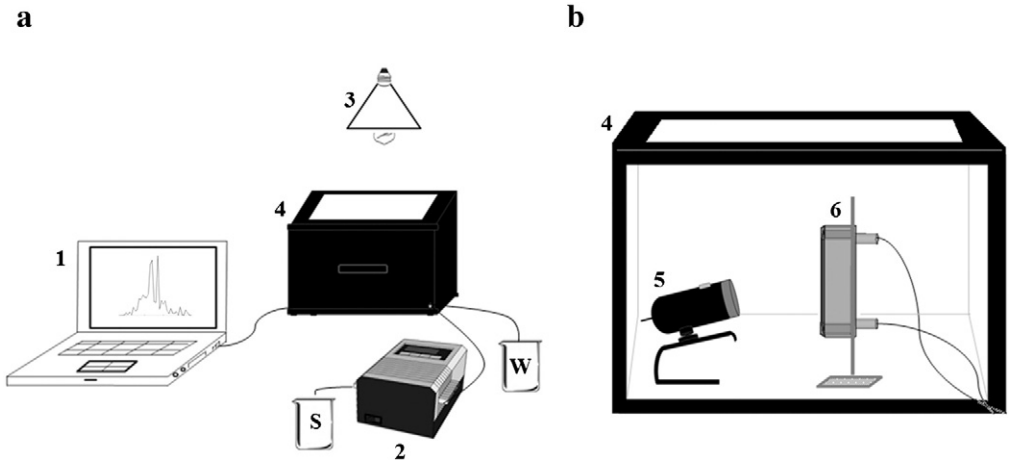

Fig. 1. (a) System used in the acquisition of images of the edible vegetable oil samples. 1: notebook; 2: peristaltic pump; 3: fluorescent lamp; 4: closed compartment. (b) Detailed compartment. 5: webcam and 6: flow cell.

of color (e.g. red or yellow), saturation refers to the relative purity or the amount of gray in a color, and intensity indicates brightness. The use of this color space in the field of color image analysis has already given quite satisfactory results [20–23].

Some works have shown the viability of the use of frequency distribution of color indexes in the different channels associated with multivariate analysis to monitor the quality of foods [24–27]. Diniz et al. [24] classified green and black tea samples according to geographical origin by using digital image and linear discriminant analysis. Godinho et al. [25] applied principal component analysis to mean histograms of image color channels in order to discriminate 29 brands of Guaraná, Cola, and orange flavors. Images of commercial potato chips were evaluated for various color and textural features to characterize and classify the appearance and to model the quality preferences of a group of consumers [26]. Mendoza et al. [27] classified commercial pork, turkey and chicken ham slices.

In all these works [24–27], the use of pattern recognition techniques were employed for discrimination and/or classification of different food samples. These techniques refer to the ability to assign an object to one of several possible categories according to the values of some measured parameters, and the classification is one of the principal goals of pattern recognition. Multivariate classification methods such as linear discriminant analysis (LDA) have been widely used in analytical applications [28].

The LDA classification method [29] employs hyperplanes, which are defined in order to maximize the ratio of between-class to within-class dispersion. In this method, the number of training samples must be larger than the number of variables to be included in the LDA model. In case of data set with high dimensionality, the use of appropriate variable selection procedures is required. Successive projections algorithm (SPA) [9,30] and stepwise (SW) [31] formulation have been adopted for this purpose in different classification problems [31–34].

The present work proposes a methodology based on digital image data and supervised pattern recognition techniques for the classification of edible vegetable oils with respect to the type (soybean, sunflower, corn and canola) and the conservation state (non-expired and expired

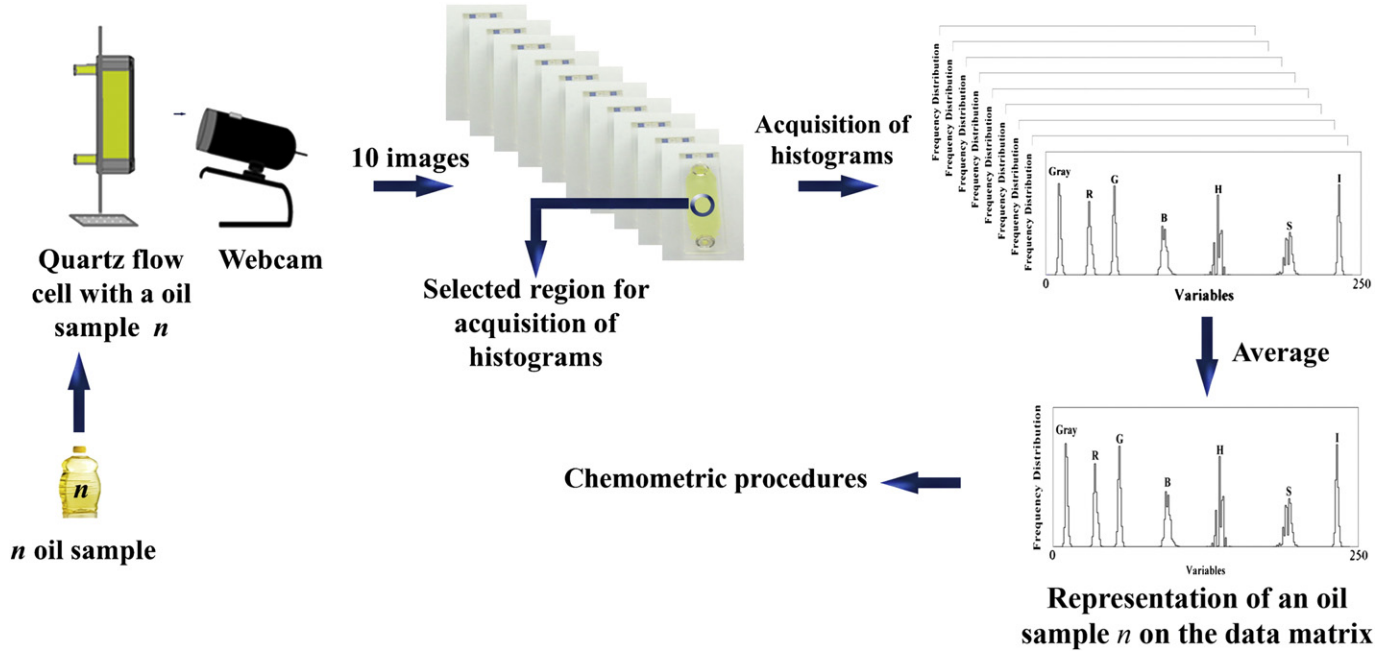

Fig. 2. Procedure of data set acquisition.