Computers and Electronics in Agriculture 118 (2015) 38–46

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Moisture content prediction in tealeaf with near infrared hyperspectral imaging

Shuiguang Deng [a,c,*], Yifei Xu [a], Xiaoli Li [b], Yong He [b,c]

[a] College of Computer Science and Technology, Zhejiang University, PR China
[b] College of Biosystems Engineering and Food Science, Zhejiang University, PR China
[c] Cyrus Tang Center for Sensor Materials and Applications, Zhejiang University, PR China

A R T I C L E I N F O

Article history:
Received 23 October 2014
Received in revised form 11 June 2015
Accepted 16 August 2015

Keywords:
Hyperspectral image
Moisture content prediction
Three-Dimension Gabor Filter
Longjing tea
Partial-least-square regression

A B S T R A C T

Near infrared (NIR) hyperspectral imaging has been used as a rapid non-destructive technique to predict moisture content of tea. To improve the performance of predicting, we first find and validate the fact that the texture near the veins is continues and directional. And then we propose Three-Dimension Gabor Filter (TDGF) and its corresponding filterbank to describe the textures of tealeaf. After that we construct two types of models based on partial least squares (PLS) regression. Experiments are conducted to predict the moisture content of Longjing tea, and different regression models based on different types of features are built for comparison. The results show that the proposed filterbank is able to detect the optimal direction of water flow and the model combining the spectrum and TDGF textures outperform the other comparative models.

1. Introduction

Tea has remained popular in China more than 2000 years ago. Longjing tea is also known as Dragon Well tea; because it one of the top tea species, it is famous around the world. Currently, a complex of tea-frying procedure is required, and this process must occur by hand by experienced workers, increasing the cost of the product. Therefore, a rapid and non-destructive method for predicting moisture content is essential when developing equipment to automate this complex procedure (Román and Hensel, 2014; Yi et al., 2013).

Hyperspectral image analysis based on spectral data has been used for moisture content detection on living plants. In the NIR spectrum (Yin et al., 2013), the stretching and bending of the O–H bonds in water molecules determines whether electromagnetic radiation is absorbed or reflected at a specific wavelength (Lorente et al., 2012). In addition to the spectral information, hyperspectral images include additional information regarding texture. Compared to the spectral data, texture information is more similar to human visual perception, enabling the direct identification of complex features (Shrivastava and Tyagi, 2014); this method has proven effective for classifications in various fields, including remote sensing (Ma et al., 2014), biometrics (Deng et al., 2014; Noviyanto and Arymurthy, 2013) and food engineering (Deng et al., 2013; 2015). For classification problems, texture information has proven efficient tools at various tasks, such as identifying different teas (Li et al., 2011). For regression problems, texture information is usually discarded because texture features are irregular and discrete. Recently, although several works based on texture have been reported to analyze hyperspectral image (Li et al., 2011; Zhu et al., 2013), they are helpless in the prediction model for moisture content detection due to the fact that texture is continuous and directional, which have been proved at first in this work. Consequently, it is essential to design a texture filters that can reflect the water change effectively.

Therefore, in this study, we first propose TDGF and the filterbank of TDGF to describe the texture information, and then build two types of predictive models (PCPLS and SPAPLS) to predict the moisture content. Experiments on the Longjing tea are conducted to evaluate the performance of the predicting models. The results demonstrate that the models build on spectrum and TDGF textures outperform the comparative models, and confirm the hypothesis that the water transfer occurs primarily from the vein to the mesophyll. The study includes the following major contributions: (1) the property of continuous texture near the veins is found and validated; (2) three-dimensional Gabor filter (TDGF) algorithm and the filterbank covering different orientations and scales are proposed for hyperspectral image analysis; (3) two different

* Corresponding author at: College of Computer Science and Technology, Zhejiang University, PR China. Tel.: +86 571 88273422; fax: +86 571 87951647.
E-mail address: dengsg@zju.edu.cn (S. Deng).

http://dx.doi.org/10.1016/j.compag.2015.08.014
0168-1699/© 2015 Elsevier B.V. All rights reserved.

prediction models (PCAPLS and SPAPLS) combining the spectral and texture data are proposed for laboratory research and industrial applications, respectively. (4) The assumption of water transfer occurs primarily from the vein to the mesophyll has been proved.

The rest of this paper is organized as follows: Section 2 introduces the framework and methodology. The experiments and model settings are described in Section 3. Section 4 discusses the evaluation results. We conclude this paper and discuss the future work in Section 5.

2. Framework and methodology

2.1. Framework

The framework of predicting the moisture content is shown in Fig. 1. The input is tea sample; the output is the prediction models. It mainly includes the following three steps.

Data acquisition: this step is to obtain the original hypspectral image of the samples. The samples are (a) picked, (b) dried and weighed, and then (c) processed using a hyperspectral image system.

Image preprocessing: the goal of this step is to preprocess the hyperspectral images in order to improve the performance of the predicting models. (a) The hyperspectral images are calibrated first; (b) then, the regions of interest (ROI) are acquired; (c) finally, the denoise methods are applied to sharpen the textures.

Model establishment: the process establishes regression models based on texture features and spectrum, which plays a critical role in predicting the moisture content. (a) Texture extraction. It is to describe the texture informatively and can be implemented with different methods. (b) Prediction model. It focuses on the regression models based on the representative features to predict the moisture content.

Figs. 1 and 2 show the flow of Model Establishment. In the process of Texture Extraction (Fig. 2(b)), No texture imaging methods (GLCM and GLRLM), and texture imaging methods (TDGF) are executed to describe the informative textures, respectively. No texture imaging denotes that the textures are only presented in the statistic way but without any image, and the texture imaging always illustrates the textures in a visualization way with different orientations. In terms of no texture imaging, the 3D hyperspectral image first are projected into the 2D space of principle components (PCs) with the help of PCA, and then the resultant PCs are processed by GLCM and GLRLM to generate the texture features. With respect to texture image, the original hyperspectral image are convoluted by the designed filterbank of TDGF to form the texture features in a visualization way, which provides an applicable tool to detect the sensitive direction of water flow in the tealeaf. In the process of Prediction model, nine PCPLS and one SPAPLS are built based on different features to predict the moisture content.

In addition, Step 1 and Step 2 involves a series of experimental operations, and the details can be found in Sections 3.1 and 3.2, respectively. Step 3 requires a variety of methods to build regression models for predicting. Therefore, in the Methodology, we will present the methods handled in the process of Model Establishment.

2.2. Methodology

2.2.1. No texture imaging

In this paper, no texture imaging methods we use are the well-worked methods: the gray level co-occurrence method (GLCM) (Xian, 2010), the gray level run length method (GLRLM) (Galloway, 1975). GLCM and GLRLM both construct a new matrix to detail the relationship between pixels and their direct or indirect neighbors. GLCM reveals that the same gray level occurs in pairs at a given distance and direction. GLRLM reveals the number of times a gray level appears continuously along one straight line in a given direction. The GLCM measures the probability of whether a pixel at a particular gray level occurs at a specified direction and a distance from its neighboring pixels.

The GLRLM measures the maximum length that a pixel of a particular gray level occurs in a specific direction and distance continuously. Generally, four properties containing contrast, correlation, energy and homogeneity are calculated using GLCM (Haralick et al., 1973). Eleven properties including short run emphasis (SRE), long run emphasis (LRE), gray-level nonuniformity

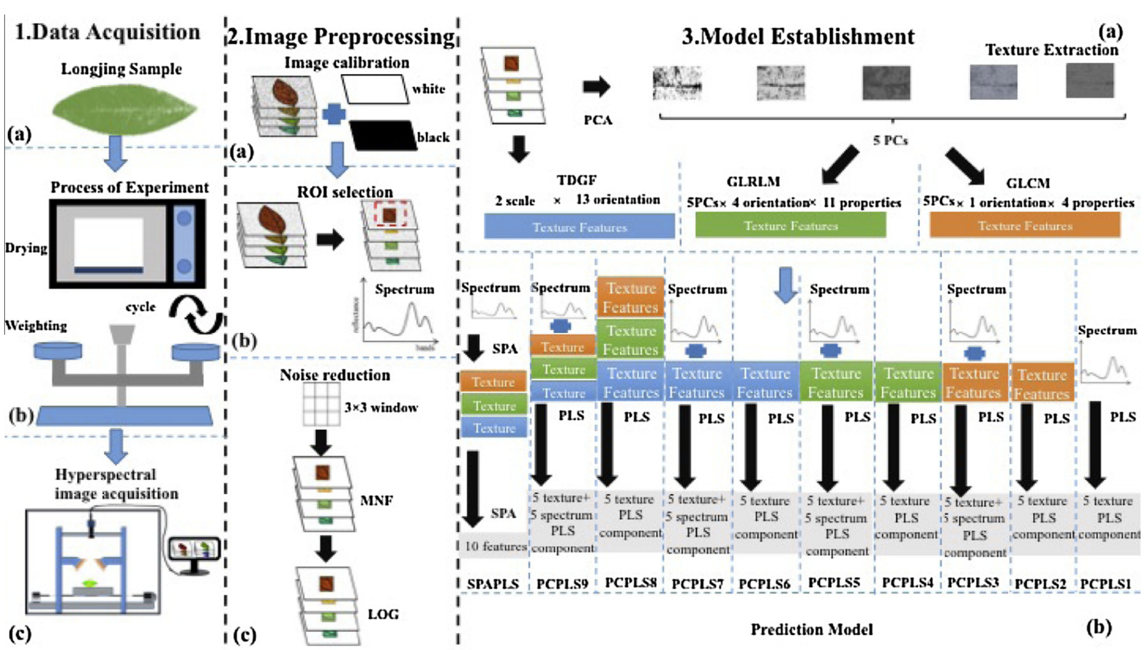

Fig. 1. The framework of predicting the moisture content of Longjing Tea.