BIOSYSTEMS ENGINEERING 156 (2017) 51–60

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Bag-of-Feature model for sweet and bitter almond classification

Abozar Nasirahmadi [a,*], *Seyed-Hassan Miraei Ashtiani* [b]

[a] School of Agriculture, Food and Rural Development, Newcastle University, Newcastle upon Tyne NE1 7RU, UK
[b] Department of Biosystems Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

ARTICLE INFO

Article history:
Received 26 October 2016
Received in revised form
23 January 2017
Accepted 24 January 2017

Keywords:
Almond
Bag-of-Feature
Classification
Keypoint
Machine vision

The use of computer vision techniques in post-harvest processing of agricultural products has increased considerably in recent years due to their non-destructive and rapid monitoring abilities. Image processing, combined with pattern recognition, has been applied in fruit sorting and classification. In this study, a Bag-of-Feature (BoF) model is used for the classification of 20 sweet and bitter almond varieties. Harris, Harris–Laplace, Hessian, Hessian–Laplace and Maximally Stable Extremal Regions (MSER) keypoint detectors along with a Scale Invariant Feature Transform (SIFT) descriptor are used in the BoF model. The k-means clustering method is applied for building a codebook from keypoint descriptors. The performance of 3 classifiers, which were k-Nearest Neighbour (k-NN), linear and chi-square Support Vector Machine (L-SVM and Chi-SVM, respectively) were compared using classification results in the model. It was observed that the Chi-SVM classifier outperformed the k-NN and L-SVM classifiers. Using the BoF model, it was possible to detect and classify sweet and bitter varieties with high overall accuracy.

1. Introduction

Almonds (*Prunus dulcis*) contain numerous bioactive and nutritional ingredients (e.g. vitamin E, fibre, arginine, polyunsaturated and monounsaturated fatty acids). In 2014, worldwide almond production was about 2.69 million tonnes and the five top producing countries were the USA, Australia, Spain, Iran and Morocco (FAOSTAT, 2014).

Based on taste, almonds can be divided into two major species: bitter and sweet almonds (Borràs, Amigo, van den Berg, Boqué, & Busto, 2014). Usually, sweet almonds are used in the food industries or for human consumption without any pre-processing (Loghavi, Souri, & Khorsandi, 2011), whereas bitter almonds are toxic due to their cyanogenic glycosides and cannot be consumed without removal of toxic contents. However, the essential oil of bitter almonds is used in the pharmaceutical and cosmetic industries (Loghavi et al., 2011). Sometimes, bitter and sweet almonds are unintentionally mixed during post-harvest processing, which gives an undesirable experience to consumers. Furthermore, a bitter almond contains 4–9 mg of highly toxic hydrogen cyanide, and consumption in large amounts could lead to death (Volpi, 2016). It is therefore important to distinguish bitter and sweet almonds due to health risks of cyanide intake and likely economic losses from reduced value (Toomey, Nickum, & Flurer, 2012).

There are several techniques for recognition and analysis of cyanogenic glycosides, which can serve as criteria for separation of sweet and bitter almonds, including high-

* *Corresponding author.*
E-mail addresses: a.nasirahmadi@gmail.com, abozar.nasirahmadi@ncl.ac.uk (A. Nasirahmadi).

http://dx.doi.org/10.1016/j.biosystemseng.2017.01.008
1537-5110/

performance liquid chromatography (Dicenta et al., 2002), micellar capillary electrophoresis (Campa et al., 2000) and thin layer chromatography (Zhao, 2012). Although these methods are highly accurate in measurement, and therefore sweet/bitter almond recognition, they are complex, expensive, time-consuming and destructive (Borràs et al., 2014). Two expert systems, i.e. human vision and machine vision, are used for recognition and classification of agricultural products. Human vision-based inspection and classification methods have received less attention due to serious weaknesses, e.g. high costs, low speed, need for experienced work force, low efficiency and accuracy, particularly for crops with highly similar cultivars like almonds. In contrast, machine vision has been an important technique for a wide variety of applications in agriculture, post-harvest processing and food engineering. This technique is an alternative, cheap and non-contact method that can replace human observation and other destructive methods of food assessment.

There are several studies in the literature where computer vision, along with pattern recognition (expert systems) methods, has been applied for recognition and classification of agricultural products in recent years. For instance, shape, colour and textural features have been used for wheat kernel detection (Shrestha, Kang, Yu, & Baik, 2016), rice and paddy classification (Chaugule & Mali, 2016), classification of beans and asparagus (Donis-González & Guyer, 2016; Nasirahmadi & Behroozi-Khazaei, 2013) and barley variety detection (Szczypiński, Klepaczko, & Zapotoczny, 2015). However, no studies have yet been carried out on the topic of detection of sweet and bitter almond varieties by means of image processing as a non-destructive method. To differentiate sweet and bitter almonds, Micklander, Brimer, and Engelsen (2002) and Borràs et al. (2014) used Raman and near-infrared spectroscopy, respectively. They used data on different molecular structures of samples to separate bitter and sweet almonds. Despite their strengths, these techniques have drawbacks. Fluorescent compounds usually interfere with Raman spectroscopy, resulting in no suitable Raman spectra. In addition, during the laser measurement process, the sample might be altered or destroyed (Thygesen, Løkke, Micklander, & Engelsen, 2003). The weaknesses of the near-infrared spectroscopy method are that it is time-consuming and needs a calibration procedure, in addition to complex analysis and data interpretation (Giokas, Thanasoulias, & Vlessidis, 2011).

In addition to the nutritional values of almond and widespread use in various industries (i.e. food, medical and cosmetics) the economic importance of this product is another reason to choose this crop for study. The world trade export value of almonds has reached US$ 4.7 billion by 2014 (FAOSTAT, 2014). In this study, an algorithm based on a Bag-of-Feature (BoF) model for sweet and bitter almond variety detection is proposed. Over the last years, in order to carry out object recognition and classification, researchers in the machine vision field introduced the BoF model (Yoo & Kim, 2013). The model has been used due to its low computational cost and its robustness (Tamaki et al., 2013). The BoF model is based on image keypoint detectors, an image descriptor, visual dictionary and classifiers, which are addressed in the following paragraphs. Recently, some keypoint detectors and descriptors have been applied in agricultural-based research for detection and classification of products/objects using different classifiers. Thanks to the capability of high dimensional datasets classification and ability to well address the over-fitting problem (Lei et al., 2017) Support Vector Machine (SVM) classifier has been used in most of the studies recently. Five orchard insect species were classified by means of local features, a scale invariant feature transform (SIFT) descriptor and six classifiers by Wen, Guyer and Li (2009). Similarly, in order to automatic fruit recognition in images, a Bag-of-Words along with a SVM classifier model was proposed by Song et al. (2014). In another study by applying dense SIFT features and SVM, 40 different wheat grain varieties were classified with a satisfactory accuracy rate (Olgun et al., 2016). Pires et al. (2016) introduced a high accuracy method based on image local descriptors and SVM classifier for detecting soybean disease.

Although these studies have concentrated on recognition and classification of products by image analysis, no specific study for the detection of almond varieties has been reported.

Addressing the challenges of categorising sweet and bitter almonds with no need for sample pre-processing is a key factor in introducing this model into design of automatic sorting systems. Hence, we attempt here to employ a BoF model followed by k-Nearest Neighbour (k-NN) and SVM classifiers, for classification of 20 bitter and sweet almonds. The remaining sections are organised to gives a description of the different steps in the BoF model, the almond varieties used in this work, the results and discussion of the experiment.

2. Material and methods

2.1. Image acquisition

Twenty varieties comprising, 8 bitter almonds: Ahmadabad, Talkhe-Kaghazi, Mehrizi, ME-34, Koohi, HC-9, Dehshir, Bahadoran and 12 sweet almonds: Sangi-12, Najafabadi, Sangi-05, MT-101, Shahrood-13, Harir, PA-C-9, Arzhan-Y, Mostofi, Dirgol-A-2, Badamak, Sefid were randomly selected from an almond orchard in a commercial farm from Yazd, Iran (Fig. 1). A fluorescent lamp with 36 W and of around 500 mm diameter was mounted above the samples to provide light during the capturing and colour (RGB) images were acquired using a commercial CCD camera (Nikon D3200). The camera was fixed at 300 mm above the samples with its lens pointing downward to acquire top view images with an original resolution of 4320×3240 pixels which were then cropped to 1024×1024 pixels resolution. It is crucial to detect background and foreground of an image because the background usually contains disturbing visual information that can affect the performance of models (Feng, Bhanu, & Heraty, 2016). Therefore, multi-threshold methods, as described by (Xiao-bo, Jie-wen, Yan-xiao, & Holmes, 2010) were applied for background removal using the Image Processing Toolbox of MATLAB® (the Mathworks Inc., Natick, MA, USA).

2.2. Bag-of-Features model

In order to detect the appearance of the different almond classes, a BoF model was applied. The BoF framework represents an image as a histogram of representative local