ilmedia

TECHNISCHE UNIVERSITÄT
ILMENAU

Lemanzyk, Thomas; Anding, Katharina; Linß, Gerhard; Rodriguez Hernández, Jorge; Theska, René

Food safety by using machine learning for automatic classification of seeds of the South-American Incanut plant

Original published in:	2014 Joint IMEKO TC1-TC7-TC13 Symposium: Measurement Science Behind Safety and Security / Sousa, João A.. - Bristol : IOP Publ.. - (2015), art. 012036, 5 pp. ISBN 978-1-5108-0291-9 (Journal of physics. Conference Series ; 588)
Conference:	Joint IMEKO TC1-TC7-TC13 Symposium: Measurement Science Behind Safety and Security ; (Madeira) : 2014.09.03-06
Original published:	2015-02-16
ISSN:	1742-6596
DOI:	10.1088/1742-6596/588/1/012036
[*Visited:*	2024-02-01]

This work is licensed under a Creative Commons Attribution 3.0 Unported license. To view a copy of this license, visit https://creativecommons.org/licenses/by/3.0/

TU Ilmenau | Universitätsbibliothek | ilmedia, 2024
http://www.tu-ilmenau.de/ilmedia

Food Safety by Using Machine Learning for Automatic Classification of Seeds of the South-American Incanut Plant

Thomas Lemanzyk[1], Katharina Anding[1], Gerhard Linss[1], Jorge Rodriguez Hernández[2] and René Theska[3]

[1]Faculty of Mechanical Engineering, Department of Quality Assurance and Industrial Image Processing, Ilmenau University of Technology, Ilmenau, Germany

[2]Departamento Académico de Ingeniería, Pontificia Universidad Católica del Perú (PUCP), San Miguel, Lima, Peru

[3]Faculty of Mechanical Engineering, Department of Precision Engineering, Ilmenau University of Technology, Ilmenau, Germany

E-mail: thomas.lemanzyk@tu-ilmenau.de

Abstract. The following paper deals with the classification of seeds and seed components of the South-American Incanut plant and the modification of a machine to handle this task. Initially the state of the art is being illustrated. The research was executed in Germany and with a relevant part in Peru and Ecuador. Theoretical considerations for the solution of an automatically analysis of the Incanut seeds were specified. The optimization of the analyzing software and the separation unit of the mechanical hardware are carried out with recognition results. In a final step the practical application of the analysis of the Incanut seeds is held on a trial basis and rated on the bases of statistic values.

1. Introduction

Through increasing globalization, the compliance with international quality standards is necessary. The quality demand of food is related for example to avoidance of uncleanness inside or on food products, contamination through environmental influences, foreign material or degeneration of the product itself. To avoid this inadvertent decrease of quality, control- and counteractive measures in the handling process of the end-product are necessary.

This paper particular handles the quality control in the production process of seeds of the so called Sacha Inchi, also known as Incanut because of its South-American origin.

As a natural product, the seeds of the Incanut are open to different climatic conditions. These nature given influences have a direct effect on the end-product, which can appear for example in discolouration, rests of shells, rottenness or missing seed-parts. All the unwanted parts in an amount of seeds are called impurities, as it is also done with other grocery, especially in the case of grain. Seeds which fit to in design and colour to the demand of the producer are named as proper seeds or seeds in a proper condition. To avoid common contamination as good as possible, they have to be recognized and identified initially. In general there are two different ways of identification. On the one hand there

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

is the manual identification of unwanted impurities by human sighting (the state of the art) and on the other hand there is the automatic identification with the help of industrial image data processing through the mechanical systems and production engineering. The automatic identification of the impurities means the automatic classification of objects in the given sample into their related classes by using supervised machine learning algorithms. The main attention in this paper is spotted on the identification and the classification of the appearing seeds through machine help under the regard of the environmental conditions in the manufacture processing of the Incanut [1].

With an automatic control and handling of the Incanut, health risks and quality degradation of the end-product should be avoided, especially in comparison to manual methods of quality control.

2. Classes of the Incanut

The evaluation to relate the seeds to their fitting classes is usually made manually by sighting and sorting. These manual classified sorted samples build the a priori knowledge for training the classifiers.

Figure 1. Constituent parts of the Incanut seeds

	Class of seed / Besatz	Image
Directly usable	Flawless whole seed	
	Flawless seed-fraction	
Recyclable	Seed with coat	
	Seed with coat-rests	
Waste	Seed-coat	
	Rotten seed	

Figure 2. The 6 subclasses and the 3 superordinated classes of the Incanut

Figure 1 illustrates the constituent parts of the Incanut seeds and figure 2 illustrates the different six subclasses and the three superordinated classes of the Incanut, which are important for production process and food safety. Four of the six different classes of seeds can be utilized or rather be leaded back to the process. Just two classes of the main class “waste” are classified as useless and are being disposed. Therefore in the daily routine of production, the main classes “directly usable”, “recyclable” and “waste” are mainly distinguished and accordingly sorted.

3. Image analyses - object recognition in colour images

The automatic classification of Incanut seeds based on the image processing of colour images and the using of an adapted trained classifier. The manual classification and sorting of the seeds, this means their fitting in the defined classes, builds the basis of the developing of an automatic recognition routine. The investigations are based on the grainspector, an analysing software and a mechanical hardware device for the automatic analysis of grain, which were developed in a prior project of the department of Quality Assurance and Industrial Image Processing, Ilmenau University of Technology (see [2], [3], [4], [5]). So the task was the optimizing of the classifier and the separation unit of the mechanical hardware device. The colour images are captured on a conveyor belt with a 3-CCD colour line scan camera with divisor prism. A combination of incident light and transmitted light was chosen. A wide sample dataset is generated, which contains all object classes in a large number of object samples. Our research is based on a dataset of the 6 object classes (seen in figure 2) with totally nearly 6500 objects.

Randomly chosen two-thirds of the images of the image-dataset are used for the classifier training and randomly chosen one-third of the images are used for the testing. The classifier is being learned through the training process with the attribute-vectors of the learning-dataset.

Thereby altogether 227 contour, colour and texture attributes of the software Halcon are used [6]. Three of the used features should described in detail. As for example the contour feature *area_center* computes the area and the center of an input region. The contour feature *contlength* calculates the total contour-length of the input region and the contour feature *roundness*, which is a specific shape factor from the contour and is determined as the relation between the mean value of the distance (*Distance*, see equation (1)) and the standard deviation of the mean distance (*Sigma*, see equation (2)) [6]. The Roundness is calculated in equation (3) with p is the center of the area, p_i the pixels and F the area of the contour.

$$Distance = \frac{1}{F}\sum \left\| p - p_i \right\| \tag{1}$$

$$Sigma^2 = \frac{1}{F}\sum \left(\left\| p - p_i \right\| - Distance \right)^2 \tag{2}$$

$$Roundness = 1 - \frac{Sigma}{Distance} \tag{3}$$

4. Classification with Support Vector Machines (SVM)

The used analysing software [2], [3], [4], [5] allows the training of several SVM classifiers after feature extraction with different kernel types (for example Radial Basis Function (rbf) or homogeneous and inhomogeneous polynomial kernel functions) and different internal parameters. In these investigations the implemented SVM of [6] was optimized on the given recognition task by optimising the parameters of the SVM classifier. The classification with SVM needs an optimal parameter selection for a good classification performance.

The SVM was first introduced by [7] and is mentioned as being one of the most powerful classifiers today. It is derived from the statistical learning theory [8]. The algorithm is motivated by the structural risk minimization which says that not only the training error but also the complexity of the model influences its generalization ability. The SVM was designed to solve binary classification problems but there are different strategies to solve multi-class problems, too. The SVM executes a non-linear projection of data in a higher dimensional feature space. The classes are separable in a linear way in this higher dimensional feature space. During the training process an optimal hyperplane is constructed. Optimal means that it leaves a maximal margin between the hyperplane and the closest training point on both sides.The kernel function $k(x,x_i)$ extends the linear discriminant SVM to a nonlinear machine. The given decision function [9] is given in equation (4) and the used radial basis function kernel (rbf) as one of the most popular kernel functions is defined in (5).

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{m} \alpha_i y_i k(x, x_i) + b \right). \tag{4}$$

$$k(x, x') = e^{-\gamma \left\| x - x' \right\|^2} \tag{5}$$

After optimizing the given hard- and software and after elimination of images with foreign-particles as wrong apriority-knowledge in manual expert-sorted samples or images with artefacts caused by belt-defects the individual recognition rates (RR) for the 6 subclasses obtain are for flawless seed-fractions 95.8 %, flawless whole seeds 94.4 %, seeds with coat 98.2 %, seeds with coat-rests 89.2 %, seed-coats 93.4 % and rotten seeds 94.3 %. The total recognition rates over all 6 classes obtain 94.4 % (see figure 3).

Under application of the three main classes (superordinated classes), which make sense for the production process, significantly higher recognition rates are achieved (see figure 4). Thereby the individual recognition rates for the main classes are for directly usable 97.3 %, Recyclable 97.8 % and Waste 96.1 % with a total recognition rate of 97.0 %.

		Classified as							
	Classes	[1]	[2]	[3]	[4]	[5]	[6]	Individual RR [%]	Total RR [%]
Real	Flawless seed fractions [1]	680	9	0	0	17	4	95.8	94.4
	Flawless whole seeds [2]	10	252	0	0	2	3	94.4	
	Seeds with coat [3]	0	0	321	1	3	2	98.2	
	Seeds with coat rests [4]	1	0	24	273	8	0	89.2	
	Seed coats [5]	10	0	4	1	470	18	93.4	
	Rotten seeds [6]	18	1	0	0	2	348	94.3	

Figure 3. Optimised SVM – recognition performance for 6 subclasses

		Classified as				
	Main class	Directly usable	Recycable	Waste	Individual RR [%]	Total RR [%]
Real	Directly usable	951	0	26	97,3	97,0
	Recycable	1	619	13	97,8	
	Waste	29	5	838	96,1	

Figure 4. Optimised SVM – recognition performance for 3 superordinated classes

5. Results

An informative basic research on the plant and plant-components of the South-American Incanut and their use was carried out through initial extensive research. Although the plant is generally not well-known in Europe, a definition of the classes of the Incanut-seeds was done on basics of experiences with the plant in South-America.

With the help of these information it was possible to create the basics for the realization of the classification of the Incanut-seeds for the production of Incanut-products. The software and hardware of the grainspector [2], [3], [4], [5] were modified and optimized to the Incanut-seeds. With this configuration, initial practical tests with samples of material of Incanut-seeds from Ecuador could already be analyzed. The test results and the function of the hard- and software are very satisfying. After optimizing the hard-and software and after the elimination of the pictures with belt-defects and foreign-particles (wrong apriority-knowledge in the actual correctly sorted samples) the individual recognition rates for the classes are: flawless seed-fractions 95.8 %, flawless whole seeds 94.4 %, seeds with coat 98.2 %, seeds with coat-rests 89.2 %, seed-coats 93.4 % and rotten seeds 94.3 %. The total recognition rate of all classes was 94.4 %.

Under application of the three main classes, which make sense for the production process, significantly higher recognition rates are achieved. Thereby the total recognition rates for the main classes are: directly usable 97.3 %, recyclable 97.8 % and waste 96.1 % with a total recognition rate of 97.0 %.

The next steps for these investigations will be the increasing of datasets and the further optimizing of the hardware device for decreasing of image artefacts and in-contact objects.

References

[1] Lemanzyk T 2013 Basic Research on automatic optical Classification of Sacha Inchi Seeds *master thesis* Ilmenau University of Technology Ilmenau Germany

[2] Brückner P, Anding K, Dambon M and Garten D 2010 Measuring Wheat Quality, *Vision Systems Design Magazine – Vision and Automation Solutions for Engineers and Integrators Worldwide* **15** Issue 6 Nashua USA

[3] Anding K and Garten D 2008 Comparison of Different Classification Algorithms at the Application of Automatical Quality Assurance of Grain *Proceedings of the 53rd International Scientific Colloquium (IWK)* of Ilmenau University of Technology, Faculty of Mechanical Engineering p 349-350

[4] Anding K, Brückner P and Linß G 2010 Influence of Dataset Character on Classification Performance of Support Vector Machines for Grain Analysis, *Artificial Intelligence and Applications* Innsbruck Austria p. 75 – 81

[5] Garten D, Anding K, Lerm S, Linß G and Brückner P 2013 Image Analysis of Natural Products, *Conference on Optical Characterization of Materials* OCM 2013 Karlsruhe

[6] Halcon, *Reference Manual* - Halcon 8.0.3, MVTec Software GmbH, www.mvtec.com 2009

[7] Boser B, Guyon I and Vapnik V 1992 A Training Algorithm for Optimal Margin Classifiers *Proceedings of the Fifth Annual Workshop on Computational Learning*

[8] Vapnik V 1995 The Nature of Statistical Learning Theory (New York: Springer)

[9] Guyon I, Elisseeff A 2006 Feature extraction: foundations and applications (Berlin: Springer)