Computers and Electronics in Agriculture 139 (2017) 138–152

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Robust classification approach for segmentation of blood defects in cod fillets based on deep convolutional neural networks and support vector machines and calculation of gripper vectors for robotic processing

Ekrem Misimi *, Elling Ruud Øye, Øystein Sture, John Reidar Mathiassen

SINTEF Ocean AS, NO-7465 Trondheim, Norway

A R T I C L E I N F O

Article history:
Received 30 November 2016
Received in revised form 10 March 2017
Accepted 18 May 2017

Keywords:
Image segmentation
Industrial application
RGB-D image
Robotics
Support vector machines
Deep convolutional neural networks

A B S T R A C T

Despite advances in computer vision and segmentation techniques, the segmentation of food defects such as blood spots, exhibiting a high degree of randomness and biological variation in size and coloration degree, has proven to be extremely challenging and it is not successfully resolved. Therefore, in this paper, we propose an approach for robust automated pixel-wise classification for segmentation of blood spots, focusing specifically on challenging texture-uniform cod fish fillets. A multimodal vision system, described in this paper, enables perfectly aligned RGB and D-depth images for localization of segmented blood spots in 3D. Classification models based on (1) Convolutional Neural Networks - CNN and (2) Support Vector Machines - SVM for the classification of defective fillets were developed. A colour-based, pixel-wise and SVM-based model was developed for accurate segmentation and localisation of blood spots resulting in 96% overall accuracy when tested on whole fillet images. Classification between normal and defective fillets based on GPU (Graphical Processing Unit) - accelerated CNN classification model achieved 100% accuracy, versus the SVM-based model achieving 99%. We present a novel data augmentation approach that desensitizes the CNN towards shape features and makes the CNN to focus more on colour. We show how pixel-wise classification is used for an accurate localization of blood spots in 3D space and calculation of resulting 3D gripper vectors, as an input to robotic processing.

1. Introduction

Blood spots and discolouration resulting from inappropriate bleeding are detrimental to fillet flesh quality (Erikson et al., 2010). The visual effect of residual blood in fillets reduces consumer acceptance and the market value of the product. Currently, fillets with blood spots are manually sorted and trimmed to remove parts that are discoloured due to the presence of blood. The industry requires a robust, rapid, non-invasive and cost-efficient method for the effective discrimination of normal and defective fillets, which automatically segments and localises blood spots using image technologies. Blood spot segmentation is a scientific challenge that remains unresolved despite recent developments in image-based segmentation techniques. Image-based segmentation continues to be a very challenging problem, and is highly application dependent (Yang et al., 2014). The segmentation of blood spots in fillet muscle tissue falls into the category of hard-to-solve challenges due to high levels of randomness, high variation in colour, spectral similarity of blood spots with other similar defects and inherent biological variation encountered in biological raw materials. For this reason, addressing this challenge with a cost-efficient multimodal imaging system has great value, both generically and in terms of practical application. While recent imaging techniques, combined with machine learning (Misimi et al., 2007; Jackman et al., 2009; Balaban et al., 2011; González-Rufino et al., 2013; Cernadas et al., 2005), have been shown to be efficient tools for food quality assurance, it is also shown that food applications and recognition are challenging topics in computer vision (Martinel et al., 2016).

Blood spot detection and segmentation in raw material has proved to be very challenging to automate. Mertens et al. (2011) performed a spectral characterisation of egg shells to detect blood spots and concluded that brown pigments and other discolouration of the shells interfere with the peak detection of blood (at 577 nm) and thus makes the detection of blood spots challenging. Balaban et al. (2011) developed an image analysis method to quantify gaping and bruising, and the presence of blood spots, observed on salmon fillets, by adaptively applying an L (Lightness) threshold value. The authors suggested that a robust blood spot detection

* Corresponding author.
E-mail address: ekrem.misimi@sintef.no (E. Misimi).

http://dx.doi.org/10.1016/j.compag.2017.05.021
0168-1699/

system was required, based on a specifically tailored classification algorithm. Although popular due to their simplicity (Wang et al., 2011), image thresholding segmentation methods using traditional histogram-based thresholding cannot separate areas exhibiting high similarity in grey scales not belonging to same regions.

Image segmentation still remains an important area of research in the field of computer vision (Yang et al., 2014), and several approaches and methods have been proposed to solve this generic problem. These approaches are categorised according to methodology: histogram thresholding methods, clustering methods, edge detection, region methods and graph methods (Yang et al., 2014; Wang et al., 2011; Unnikrishan et al., 2007). For biological raw materials, image segmentation approaches are extremely application dependent. Image structure and information exhibit high levels of variation and randomness, making the segmentation operation even more challenging. Sometime, as in the case of cod fish fillets, the high degree of uniformity of texture of the fillet muscle image is a disadvantage and disabling factor in, for example, including texture alongside the colour as features to be used for development of robust segmentation approaches.

A prerequisite for an effective automated system is that following trained learning, it must be able to classify objects into respective class categories based on features detected on images. The selection of the appropriate classification algorithm is, therefore, key to this process. The most commonly used classification approaches for generic non-food and food related applications are (a) statistics-based, and (b) those based on Neural Networks (NN). In recent years, Support Vector Machines (SVMs) (Vapnik, 1995) have emerged as powerful classification algorithms for food applications due to their excellent performance in a variety of quality inspection tasks (Deng et al., 2013) as it can be used to solve both classification and regression problems. SVM classification algorithms has already been successfully used in several food applications such as prediction of product quality in industrial bakery processes, prediction of beef tenderness using image colour and texture features (Rousu et al., 2003; Sun et al., 2012). Du and Sun (2005) used low dimensional colour features and support vector machine algorithm to perform an automated classification of pizza sauce spread achieving 96.6% classification accuracy on the test set. The concept of deep learning is also emerging as a powerful machine learning method that allows computational models composed of multiple processing layers to learn representations of data containing multiple levels of abstraction and has dramatically improved the state-of-the-art of visual recognition applications (LeCun et al., 2015). Kagaya et al. (2014) used a convolutional neural network for recognizing food images and they observed that the network achieved significantly better performance accuracy (93.8%) than the baseline method (89.7%). In deep learning, data augmentation (Goodfellow et al., 2016) is important since in practice the amount of available data for training the network is limited. Therefore, data augmentation procedure must be performed correctly so that transformations performed in the image does not change the image class.

3D image information is valuable in applications involving robotic processing of food and calculation of respective gripper vectors, containing the pose information for the gripper, is necessary in such applications (Misimi et al., 2016). Misimi et al. (2016) demonstrate how 3D information from the Kinect v2 RGB-D camera is used to calculate the correct grasping point for 3D vision based robotic harvesting of chicken fillets.

The main research objectives of this study were: (a) to develop a robust, colour-based pixel-wise classification algorithm for blood spot segmentation in fillets as an example of objects with high intra-class variance when it comes to size, colour and localization of blood spots, (b) to develop a model for accurate classification of normal and defective fillets, (c) to develop an approach for perfectly aligned RGB-D images that can make use of pixel-wise classification for accurate localization of blood spots in 3D and calculation of gripper vectors for robotic processing; (d) to acquire a deeper understanding and visualisation of how changes in SVM hyperparameters influence pixel-wise classification in general and blood segmentation in particular, and (e) to exploit the capabilities and acquire deeper understanding of CNN for classification of cod fillets as an example of food objects and appropriateness of current data augmentation techniques for such applications.

To the best of our knowledge, no work has been published on the automated segmentation of blood spots or similar defects in food objects based on perfectly per-pixel aligned RGB-D images and robust pixel-wise classification and localisation in 3D space. Contribution on visualising the effects of change of SVM parameters in resulting classification and segmentation accuracy is also original. This paper investigates the application of CNN-based deep learning classification in food sorting applications. For this reason, the knowledge obtained by means of this study on the use and understanding of deep learning for raw food material classification is original. The data augmentation approach used to reduce the sensitivity of the CNN approach in terms of shape and increased colour sensitivity is also novel.

The rest of the paper is organized as follows: in materials and methods section we describe the collected datasets, multimodal vision system overview, and the approach for classification and segmentation of blood spots. In results and discussion section, we show in detail our results and discussion regarding the CNN and SVM classifications model, we visualise and discuss the effect of SVM hyperparameters in actual pixel-wise classification and segmentation of blood spots and we calculate the 3D gripper vectors for robotic processing. In future work section are given some solid future research directions, and finally in conclusion section, we draw some final conclusions.

2. Materials and methods

2.1. Sample preparation

Fish fillets taken from farmed Atlantic cod (Gadus morhua) were differentiated by a qualified human inspector into two categories: (a) normal (n = 33), and (b) defective (d = 32), with mean length 48.5 cm ± 5.8 cm. They were subsequently shipped from the Norway Seafoods (Melbu, Norway) fish processing company to SINTEF SeaLab in Trondheim where they were stored at 4 °C prior to imaging.

2.2. Computer vision system used to acquire the image dataset

Currently existing vision cameras such as 3D SICK IVP ColorRanger, or RGB-D Kinect v2 don't generate aligned RGB and D-depth images. This is very often a drawback when combination of both RGB and D-depth information is needed for accurate localization of regions of interest and defects in 3D space for robotic applications (Balaban et al., 2016). The multimodal vision system in this paper consisted of a colour imaging line scan CMOS camera, Grasshopper 3 (GS3-U3-23S6CC, Point Grey, Canada), with a USB 3.0 interface, enabling aligned RGB and 3D images. The Region of Interest (ROI) used for imaging the fillets was 1376×64, with an exposure time of 500 μs. The working distance to the camera was 54 cm and the tilt angle was 17 degrees. Each fillet was placed on a conveyer belt for image acquisition. A laser emitting a 100 mW red uniform laser line at 660 nm wavelength, with a fan angle of 30 degrees, was used in triangulation mode to acquire 3D and reflectance images of the cod fillets. White illumination used to acquire RGB images was provided by a flexible white LED