Computers and Electronics in Agriculture 115 (2015) 142–149

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

A robust segmentation method for counting bovine milk somatic cells in microscope slide images

G.J.A. de Melo [a], V. Gomes [b], C.C. Baccili [b], L.A.L. de Almeida [c], A.C. de C. Lima [a,*]

[a] Universidade Federal da Bahia, Rua Aristides Novis, nº 2, Bairro Federação, 40.210-630 Bahia, Brazil
[b] Universidade de São Paulo, Av. Dr. Orlando Marques de Paiva, nº 87, 05508 270 São Paulo, Brazil
[c] Universidade Federal do ABC, Av. dos Estados, nº 5001, Bairro Bangu, 09210-580 São Paulo, Brazil

ARTICLE INFO

Article history:
Received 19 June 2014
Received in revised form 24 April 2015
Accepted 29 May 2015
Available online 12 June 2015

Keywords:
Image processing
Watershed transform
k-means clustering

ABSTRACT

Mastitis is an infectious disease associated with the increased number of somatic cells in cow's milk, and it is one of the most relevant cause of economic losses in dairy farming industries. In this paper, we propose a method capable of determine, with 99.7% accuracy, the number of these cells in microscope slide images. This level of accuracy is achieved by changing the image original RGB format to Lab color space and applying *k*-means clustering algorithm to remove debris and other background features. A new gray level thresholding is proposed, and the remaining bound cells are separated in the final segmentation step applying Watershed transform. Many microscope slide images with debris, contrast, and hue variation were used to validate the experimental results. Comparison between the proposed method and manual counting indicates that this new approach is a robust and promising solution to be incorporated in a future automated somatic cell counting system using optical microscopy.

1. Introduction

Somatic cells, such as leukocytes and epithelial cells, are always present in bovine milk. Inflamed mammary glands, also known as bovine mastitis, lead to considerable growth in the number of such cells. This biologic process causes economic losses to the dairy industry, once it affects the milk quality (Cavero et al., 2007). The National Mastitis Council (Bramely et al., 1996) estimated annual losses per cow in the U.S. of US$ 185.00 due to mastitis, and a total annual expenditure of US$ 1.8 billion. This is largely due to subclinical form of the disease (Wu et al., 2005).

Mastitis can be detected using various diagnostic techniques based on milk samples, and they are divided into direct and indirect methods. As examples of indirect methods one can cite the California Mastitis Test (CMT), the Wisconsin Mastitis Test (WMT) and the measurement of Electrical Conductivity (EC). Whereas examples of direct methods include counting of somatic cells by Direct Optical Microscopy (DMSCC) and electronic analyzers, such as flow cytometry or particle detection electronics (Moon et al., 2007). The Somatic Cell Count (SCC) is historically the most commonly used indicator for identifying mastitis. However, other indicators, such as electrical conductivity (EC), Neutrophil Count based on detection of O_2, Lactate Dehydrogenase (LDH) and NAGase have been investigated in the last decades (Chagunda et al., 2006; Okada et al., 2009; Viguier et al., 2009; Ankinakatte et al., 2013). The EC of milk was introduced in 1940 as an indicator of mastitis (Davis, 1947; Norberg, 2005). Since then, several studies have been conducted to assess the accuracy of the EC and predict the state of infection (Hillerton and Semmens, 1999; Shoshani et al., 2000; Mele et al., 2001; Norberg et al., 2004; Norberg, 2005; Janzekovic et al., 2009; Ferrero et al., 2014).

Other studies have been conducted to develop indirect sensing systems for mastitis detection. For example, Eriksson et al. (2005) proposed a system based on gas-sensor and Wu et al. (2005) suggests another one using a sensor capable of detecting the Deoxyribonucleic Acid (DNA). Systems like these, together with artificial intelligence algorithms (AI), provide data that allow early detection of mastitis. For instance, Cavero et al. (2007) and Miekley et al. (2012) implemented a detection system based on univariate indicator variables, whereas Kamphuis et al. (2010) applied decision trees to the problem. Cavero et al. (2008) and Ankinakatte et al. (2013) have also developed systems based on neural networks, whereas De Mol and Woldt (2001), Cavero et al. (2006) and Kramer et al. (2009) devise a system based on fuzzy logic, and finally, Miekley et al. (2013) proposed another one based on PCA (Principal Component Analysis). However, beside all attempts to use AI and other algorithms, such systems based on detectors

* Corresponding author. Tel.: +55 (71) 3283 9478.
E-mail addresses: gabrielmelo@ifba.edu.br (G.J.A. de Melo), viviani.gomes@usp.br (V. Gomes), camila.rcosta@usp.br (C.C. Baccili), luiz.almeida@ufabc.edu.br (L.A.L. de Almeida), acdcl@ufba.br (A.C. de C. Lima).

http://dx.doi.org/10.1016/j.compag.2015.05.018
0168-1699/

are vulnerable to noise, which has a considerable influence on the characteristics of the signal to be processed and the detection technique to be applied (Kamphuis et al., 2010). Therefore, there is still a strong need for improved detection methods that translate sensor data into reliable information to be processed.

Contrary to indirect methods, which provide only an indication of potential problems associated with high somatic cell count (Ferrero et al., 2014), direct methods are more accurate and recommended by the Food and Drug Administration (FDA). In the direct approaches, the levels of mastitis are determined by SCC per milliliter of milk (Janzekovic et al., 2009), and flow cytometers are one of the most widely used test systems for this purpose, due to the high throughput and accuracy. Their use has been standardized and recommended by international dairy organizations (Gonzalo et al., 2004; Garcia-Cordero et al., 2010). However, flow cytometers are expensive to purchase (>US$ 50,000), operate and maintain, precluding small farmers from adopting them for health routine monitoring of their herds. Typically, small farmers send milk samples to government agencies or centralized private laboratories to test them by SCC, along with other tests (Garcia-Cordero et al., 2010). On the other hand, it is possible to detect mastitis and the severity of it in milk samples by SCC performed via DMSCC. This is the reference methodology adopted by the International Dairy Federation and used even for electronic equipment calibration (Orlandini and van den Bijgaart, 2011).

Unfortunately, the manual DMSCC is a tedious task that requires slide preparation and two or more hours for visual counting by a human expert. Nevertheless, a considerable number of producers in many regions around the world cannot afford electronic counting equipment, and employing DMSCC is still necessary. Thus, the introduction of image processing techniques for automatic counting in DMSCC is of major importance for SCC.

Image processing is widely employed for recognition of different cells in the fields of biology and medicine, by using techniques or algorithms already standardized. There are several software available in the market that use image processing for counting of different kind of objects, particles and cells (Abràmoff et al., 2004). But, this kind of software has specific functions for very particular processing purposes, and each practical application requires the development of new algorithms or customizing the existing ones. For example, those software usually have specific functions for segmentation and counting of biologic cells in a general sense, but the expected limitations led Kong et al. (2011), Pan et al. (2012) and Dorini et al. (2013) to develop new algorithms to deal with human blood cells. In the same way, Baro et al. (2005) presents a detailed description of the hardware and preparation methods for an automated DMSCC system (referred to as Video Microscopy – VM). This VM system is intended for SCC in cows' milk, but no details of the employed algorithms are presented in Baro's work. Moreover, recent works for somatic cell targeting have been proposed by Xue et al. (2008), Na and Heru (2009), Xue et al. (2009), Na and Heru (2010) and Wang and Xue (2010), but none of them has the necessary investigation for image variations.

The manual DMSCC technique is subjected to different manipulation procedures, depending on where it is practiced. Thus, the slide images obtained suffers many variations in their features, such as shade, hue, colorfulness, saturation, lightness and brightness, as well as the presence of debris. These variations are inherent to the process that precedes the capture of images and depend on the lactation period of the animal, hygiene conditions during the sample acquisition, staining and dyes types in the slide preparation, microscope illumination and slide region where images were taken from. Additionally, the variations in morphological patterns of somatic cells may deteriorate the performance of traditional segmentation techniques, since they can be efficient for a group of images but not for another one.

In this work, we propose a robust image processing algorithm for image segmentation to be used for SCC in automated VM system. The proposed technique is a combination of *k*-means clustering algorithm (Hartigan and Wong, 1979), Watershed transform (Roerdink and Meijster, 2000) and a new proposal for histogram thresholding. A set of one thousand DMSCC images was analyzed by a human expert and by the proposed algorithm. Promising results indicate that this technique may be used to obtain robust automated video SCC systems.

2. Materials and methods

2.1. Slide Image Features

The slide preparation for microscopy of bovine milk consists of spreading 10 μL of milk in an area of 1 cm^2, and left to dry out at room temperature to be colored with the Rosenfeld technique (Gondim et al., 1998). After completing this procedure, the slide is visualized and photographed in a microscope, using the proper magnification.

Fig. 1 exemplifies different aspects in the microcopy of somatic cells. As said before, cell morphology, debris and image variations are of major concern. According to Tripaldi et al. (2010), the somatic cells have cytoplasm and nucleus with sizes that varies from 4 to 8 μm. The nucleus and cytoplasm are colored in "dark" and light blue, respectively. The intensity of the "dark" blue in the nuclei can vary according to the used coloring technique (Fig. 1a and e).

The nuclear mass generally does not have a fixed pre-defined shape and can appear partially destroyed (Fig. 1f, cells 2 and 9). Usually, the nuclear body appears as a blue solid unit or in a granular shape (Fig. 1f, cell 8), and sometimes it can be seen in other peculiar forms. The use of different coloring techniques can lead to distinct colors for nucleus and cytoplasm, with the first one darker than the last one. On the other hand, images may have different contrast and shade variations (Fig. 1a and b), and also the presence of debris (Fig. 1c and f).

2.2. Proposed methodology for image processing

Fig. 2 illustrates schematically the procedure for segmentation and counting of somatic cells.

Initially, the RGB image is converted into Lab color space (Wyszecki and Stiles, 1982; de Melo et al., 2014) and then clustered using *k*-means algorithm (Dubey et al., 2013). This algorithm is set to separate the image content into two clusters. In one of them, cluster 2, all undesired image information of the background and debris are agglomerated. The other one, cluster 1, retains information of disjoint and connected cells. Cluster 2 image is discarded whereas cluster 1 image is converted into grayscale and then processed into Black and White (B&W) format through the application of a gray level threshold. The threshold is obtained directly from the original image histogram in grayscale format, as it is explained in Section 2.3. The connected cells are segmented by the use of Watershed transform (Roerdink and Meijster, 2000), which has been largely applied in many fields of image processing for separation of connected objects, such as presented by Tulsani (2013) and Pan et al. (2012) for human blood segmentation, and Grau et al. (2004) for brain image segmentation. Following that, the remaining forms in the image are labeled and classified into two groups: somatic cells and fragments, following the procedure presented by Chang et al. (2004).