Journal of Dairy Research

cambridge.org/dar

Somatic cell count in buffalo milk using fuzzy clustering and image processing techniques

Aline Silva Ramos[1], Cristiano Hora Fontes[1], Adonias Magdiel Ferreira[1], Camila Costa Baccili[2], Karen Nascimento da Silva[2], Viviani Gomes[2] and Gabriel Jesus Alves de Melo[3]

[1]Graduate Program in Industrial Engineering, Polytechnic Institute, Federal University of Bahia, Salvador, Brazil; [2]College of Veterinary Medicine and Animal Science, University of São Paulo, São Paulo, Brazil and [3]Federal Institute of Bahia, Ilhéus, Brazil

Research Article

Cite this article: Silva Ramos A, Hora Fontes C, Magdiel Ferreira A, Costa Baccili C, da Silva KN, Gomes V and de Melo GJA (2021). Somatic cell count in buffalo milk using fuzzy clustering and image processing techniques. *Journal of Dairy Research* **88**, 69–72. https://doi.org/10.1017/S0022029921000042

Received: 5 May 2020
Revised: 30 November 2020
Accepted: 7 December 2020
First published online: 17 February 2021

Keywords:
Biomarker; fuzzy clustering; image processing; mastitis; milk quality

Author for correspondence:
Cristiano Hora Fontes, Email: cfontes@ufba.br

Abstract

This research communication presents an automatic method for the counting of somatic cells in buffalo milk, which includes the application of a fuzzy clustering method and image processing techniques (somatic cell count with fuzzy clustering and image processing|, SCCFCI). Somatic cell count (SCC) in milk is the main biomarker for assessing milk quality and it is traditionally performed by exhaustive methods consisting of the visual observation of cells in milk smears through a microscope, which generates uncertainties associated with human interpretation. Unlike other similar works, the proposed method applies the Fuzzy C-Means (FCM) method as a preprocessing step in order to separate the images (objects) of the cells into clusters according to the color intensity. This contributes signficantly to the performance of the subsequent processing steps (thresholding, segmentation and recognition/identification). Two methods of thresholding were evaluated and the Watershed Transform was used for the identification and separation of nearby cells. A detailed statistical analysis of the results showed that the SCCFCI method is able to provide results which are consistent with those obtained by conventional counting. This method therefore represents a viable alternative for quality control in buffalo milk production.

Mastitis represents a great challenge to the buffalo milk production chain due to its impact on the volume and quality of milk produced. Somatic cell count (SCC) has been used to monitor subclinical mastitis in herds during lactation and can be performed by direct and indirect methods. The most commonly used indirect methods are the California mastitis test and the Wisconsin mastitis test. The direct methods are based on simple counting of somatic cells through the microscope or using electronic equipment. Electronic counting is fast and accurate, but the equipment is expensive. The count performed using a microscope is called direct microscopy somatic cell counter (DMSCC). Although the method is used for the calibration of electronic equipment, it is quite tedious and subject to uncertainties associated with human visualization and interpretation. This method also requires considerable work in the assembly of the slides and visual identification through the microscope which, depending on the number of images, can take a long time to complete. Studies have proposed the application of various techniques for the recognition of somatic cells in bovine and buffalo milk. These have focused on the segmentation step (image processing) and, in general, involve the application of unsupervised classification methods (*k*-means, Fuzzy C-means – FCM, neural networks) and multivariate statistical techniques (principal components analysis, PCA). The use of non-hierarchical clustering methods (k-means and FCM) to categorize each pixel into one of all identified classes has already been performed in several applications involving the processing of colored images through information such as boundaries, texture and distribution of intensity of color (Ramaraj and Niraimathi, 2017). Melo *et al.* (2015) present an image segmentation method for counting somatic cells in bovine milk. The original RGB (red, blue, green) image is converted into Lab color space and the k-means algorithm is applied to recognize two clusters of pixels. Melo *et al.* (2015) also proposed a new thresholding method that is adopted in this work. Bai *et al.* (2015) also focus on the segmentation of images for counting somatic cells in bovine milk and present the application of the *k*-means to segment the image previously converted into a smaller dimension space (RG channel and RB channel). Gao *et al.* (2017) present a method for the classification of four different types of somatic cells in bovine milk. The method is based on a feature extraction filter (Gabor wavelet) and PCA to reduce the dimension of future space. Xue *et al.* (2009) also apply PCA to reduce the dimensionality of images whose features were extracted in different color spaces.

© The Author(s), 2021. Published by Cambridge University Press on behalf of Hannah Dairy Research Foundation

This paper presents an innovative method for SCC in samples of buffalo milk based on fuzzy clustering and image processing techniques (somatic cell count with fuzzy clustering and image processing, SCCFCI). A detailed statistical analysis of the results showed that the SCCFCI method is able to provide results which are consistent with those obtained by conventional counting.

Materials and methods

Five different buffaloes were selected and from these milk samples were harvested. They presented a high somatic cell count which scored 2+ and 3+ in the California mastitis test. The preparation of the milk smears was based on the method of Prescott and Breed (1910) in which a quantity of milk equivalent to 0.01 ml was homogeneously distributed over an area of 1 cm^2 using a calibrated pipette. The slides were dried for 24 h at room temperature and stained. Each slide was placed under a microscope and scanned by at least 100 different fields in order to capture the images (online Supplementary Fig. S1).

100, 103, 100, 100 and 108 images were obtained from the slide of each sample (5 animals, total of 511 images), respectively. The images were captured by a digital camera (CoolSnap Color, Media CybernéticsTM, USA) coupled to the optical microscope (Eclipse E800, Nikon®, Japan) with a 1000-fold increase. The 511 images in the RGB color space had an original size of 1392×1040 pixels. The size of each image was reduced to 535×400 pixels through an editing program (Adobe Photoshop CC 2015), without loss of information regarding the identification of cells.

After the acquisition of the images in the RGB color space, the experimental procedure comprised the following steps:

- Conversion of each image from the RGB pattern to the monochrome pattern to obtain the grayscale histogram
- Application of the FCM method for clustering the images of the whole sample according to the gray shade distribution of each image. In this case, each image (object) is represented by a single vector with 256 (0, 1, 2, …, 255, grayscale) components. Each component is obtained by the product between the gray intensity and its respective frequency of occurrence
- Obtaining the thresholds related to each of the clusters recognized by the FCM method using the methods of Rosin and Melo
- Segmentation, application of the WT and counting of cells.

The application of the FCM method is justified by the heterogeneity of the whole sample which, in turn, suggested that the use of a single threshold for all images would not be appropriate (online Supplementary Fig. S2 shows four images for which the same threshold was considered). After applying the FCM method, a threshold specific value was obtained for each image belonging to a given cluster. Since the images belonging to the same cluster have similar gray shades (similar histograms), the threshold of each cluster was simply obtained by the arithmetic mean of the thresholds of all the images belonging to this same cluster.

Statistical analyses used included a normality test (Quantile−Quantile (Q−Q) plot test) applied to the numbers of cells identified in the images. The similarity between the results obtained by the specialists and the proposed method was verified through the *Kruskal Wallis* nonparametric test, Spearman's correlation test and the Bland−Altman plot. All tests were performed using Matlab.

Results

Since the FCM is a non-hierarquical clustering method, the initial definition of the number of clusters is required to classify the 511 images (objects) according to the similarity. Clustering tests with 2 to 6 groups were performed and the best result (lowest misclassification rates according to experts) was obtained with 5 clusters. The number of images belonging to each of the clusters is presented in online Supplementary Table S1. The direct validation of the quality of the clustering was performed by verifying the visual similarity between the images of each cluster and, subsequently, indirectly through the results of final counting of the cells. The application of FCM was able to recognize patterns of similarity between the several images in the sample (online Supplementary Fig. S3). Some images of each of the recognized clusters are presented in online Supplementary Fig. S4.

Automatic cell counting of all animal samples was performed using the proposed method (SCCFCI) and compared to cell counts performed by three different specialists who had no contact with each other. The three experts applied the traditional DMSCC method. Online Supplementary Table S2 presents the counting results considering both thresholding methods (Melo and Rosin). The Melo thresholding method provided better results (quite close to the count of all specialists) while the Rosin method tended to overestimate the amount of cells (which is also illustrated in online Supplementary Fig. S5).

Online Supplementary Table S2 also presents the results of counting with the proposed method but without the previous clustering of the images and, therefore, with the definition of a single threshold for the whole sample.

Figure 1 presents a box-plot analysis considering the cell count obtained by the proposed method (SCCFCI) and conventional countin by specialists, for each of the 511 sample images. Regarding the dispersion, all the medians (SCCFCI and specialists) were equal to 2 somatic cells and the third quartile (75% of values above the median) was also the same in all counts. The first quartile (25% of values below the median) obtained by the SCCFCI method was lower than the results obtained by the conventional counting.

The analysis of the medians obtained by the proposed method (SCCFCI) and conventional counting requires verification to see if these are statistically similar. The Quantile-Quantile (Q-Q) plot test (Das and Imon, 2016; Xu *et al.*, 2017) was applied to verify if the number of cells identified in the images (Fig. 1) are distributed as a standard normal. Online Supplementary Fig. S6 shows that the distribution of cell counts by the SCCFCI method does not follow expected behavior. The same was verified with the cell counts obtained by each of the specialists performing conventional counting (1, 2 and 3).

Considering the normality test performed on the distributions and the similarity between them, the *Kruskal Wallis* nonparametric statistical test (Chaloupková *et al.*, 2018; Khan and Khan, 2018) was applied in order to confirm whether the medians of the distributions are statistically similar. The level of significance obtained (0.54 > 5%) confirms the similarity between the medians and, therefore, the consistency of the SCCFCI method (using the Melo thresholding method).

The Spearman's correlation test (online Supplementary Table S3) and the Bland−Altman plot (confidence leve equal to 95%) were applied to jointly assess the statistical similarity between the methods (SCCFCI and DMSCC, online Supplementary Fig. S7). The results show that the vast majority

https://doi.org/10.1017/S0022029921000042 Published online by Cambridge University Press

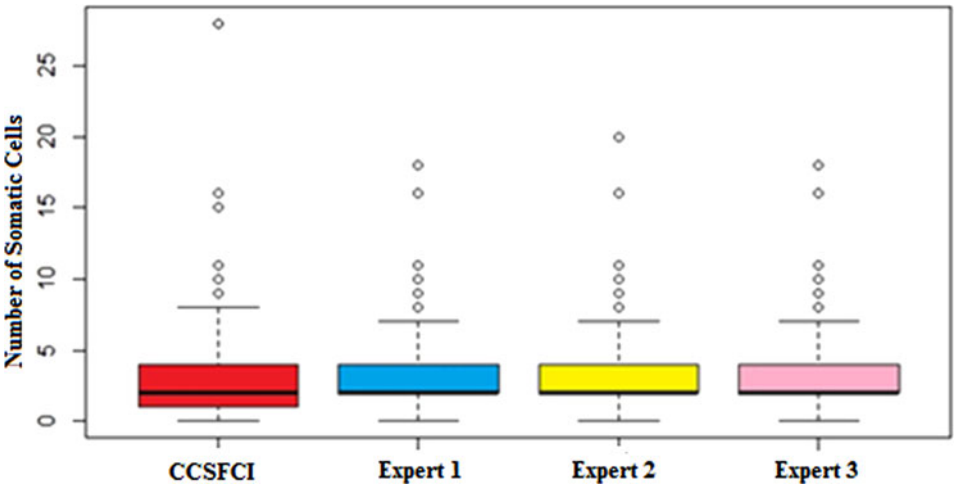


Fig. 1. Boxplot analysis. Counting cells in the 511 images.

Table 1. False positive (FP) and false negative (FN) results

FP Esp1	FN Esp1	Percentual FP Esp1	Percentual FN Esp1
207	206	(207/1476) × 100 = 14.02%	(206/1475) × 100 = 13.97%
FP Esp2	FN Esp2	Percentual FP Esp2	Percentual FN Esp2
197	228	(197/1476) × 100 = 13.35%	(228/1507) × 100 = 15.13%
FP Esp3	FN Esp3	Percentual FP Esp3	Percentual FN Esp3
204	238	(204/1476) × 100 = 13.82%	(238/1510) × 100 = 15.76%

of the differences between the counts (considering the three experts) are not significant and the correlations are moderate, which suggests statistical similarity between the counting methods.

An additional analysis (Table 1) consisted of verifying the total of false positives (FP, debris erroneously identified as cells) and false negatives (FN, cells that have not been identified) obtained by the SCCFCI method with reference to the cells identified by the conventional counting. In this case, the percentage of FP is calculated based on the total number of cells identified by the SCCFCI method whereas the percentage of FN is calculated based on the total number of cells identified by the specialist. The percentages of the false positives and percentages of the false negatives were very close which shows that the SCCFCI method is able to maintain a balance without overestimating or underestimating the presence of cells in the images. Two phenomena may justify the occurrence of false positives and false negatives and these are related to the existence of imperfections in the images (online Supplementary Fig. S8). The first case concerns the image generation process with improper lighting. In addition, the fact that buffalo milk has a high amount of fat can cause difficulty in the preparation of some smears due to the fact that the dye used does not have the necessary adhesion to the cell nucleus, making it not very clear in the image.

Discussion

The key point of the method comprises the application of a clustering method based on optimization (FCM) to previously recognize patterns of similarity among several images of the sample. Unlike other related studies, in our proposed SCCFCI method, the Fuzzy C-Means (FCM) method was applied in the preprocessing, enabling the recognition of image patterns which directly contribute to the accuracy of the final result.

We have demonstrated that this prior clustering/classification step contributes decisively to the success of the subsequent steps of cell processing and counting. The gray intensity distribution obtained after conversion from RGB to monochromatic was able to provide the necessary features for the recognition of existing (but not previously labeled or known) clusters and patterns in the sample.

In addition to providing an automatic counting alternative, not subject to individual expert interpretation, the proposed method has the potential to increase productivity by significantly reducing cell counting time. In the traditional approach (DMSCC), although slide preparation is relatively fast, it takes an expert around 2.5 h to read the entire sample (511 images) while automatic counting (SCCFCI) of the same sample can be performed in 5 min. Additionally, the proposed method requires a system for image capture and processing, similar to the configuration used in this work. The hardware is not highly complex nor high in cost and a standard configuration involving an Intel Core I3 Processor, 4 GB RAM and 500 TB HD would be appropriate to process the captured images from the milk samples. It would require specialized software capable of processing the images and clustering them based on the Fuzzy C-Means algorithm.

The sample size used in this work supports the consistency of the results obtained. The sample size of 511 provides a 5.7% error for a 99% confidence level (Beleites *et al.*, 2013). In addition, considering the sample correlation coefficient (r) obtained in the results (0.72, Spearman's correlation test), a population

https://doi.org/10.1017/S0022029921000042 Published online by Cambridge University Press

correlation coefficient (R) of around 0.66 is estimated. This is considered a good result (strong linear correlation between the proposed model and the measurement of specialists expected for the population) according to Schober *et al.* (2018).

In conclusion, we have developed and validated an improved automated counting method (somatic cell count with fuzzy clustering and image processing|, SCCFCI). suitable for SCC measurement in buffalo milk.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0022029921000042

Acknowledgments. The research work has been supported by the Ministry of Education of Brazil under the program for improvement of higher education personnel scholarship. The authors would like to thank to CNPq Productivity of Research Funds Processes 301478/2018-0.

References

Bai J, Xue H and Zhou Y (2015) The milk somatic cell image segmentation method based on dimension reduction and fusion. In Li D and Li Z (eds) *Computer and Computing Technologies in Agriculture IX. CCTA 2015. IFIP Advances in Information and Communication Technology*. Cham: Springer, p. 478.

Beleites C, Neugebauer U, Bocklitz T, Krafft C and Popp J (2013) Sample size planning for classification models. *Analytica Chimica Acta* **760**, 25–33.

Chaloupková V, Ivanova T, Ekrt O, Kabutey A and Herák D (2018) Determination of particle size and distribution through image-based macroscopic analysis of the structure of biomass briquettes. *Energies* **11**, 331.

Das K R and Imon AHMR (2016) A brief review of tests for normality. *American Journal of Theoretical and Applied Statistics* **5**, 5–12.

Gao X, Xue H, Pan X, Jiang X, Zhou Y and Luo X (2017) Somatic cells recognition by application of Gabor. *International Journal of Pattern Recognition and Artificial Intelligence* **31**, 1757009.

Khan MF and Khan MA (2018) Optik information preserving histogram segmentation of low contrast images using fuzzy measures. *Optik – International Journal for Light and Electron Optics* **157**, 1397–1404.

Melo GJA, Gomes V, Baccili CC, Almeida LAL and Lima AC (2015) A robust segmentation method for counting bovine milk somatic cells in microscope slide images. *Computers and Electronics in Agriculture* **115**, 142–149.

Prescott SC and Breed RS (1910) The determination of the number of body cells. *American Journal of Public Hygiene* **20**, 663–664.

Ramaraj M and Niraimathi S (2017) Application of color based image segmentation paradigm on rgb color pixels using fuzzy c-means and k means algorithms. *International Journal of Computer Science and Mobile Computing* **6**, 430–440.

Schober P, Boer C and Schwarte LA (2018) Correlation coefficients: appropriate use and interpretation. *Anaesthesia and Analgesia* **126**, 1763–1768.

Xu R, Xu L and Xu B (2017) Assessing CO2 emissions in China's iron and steel industry: evidence from quantile regression approach. *Journal of Cleaner Production* **152**, 259–270.

Xue H, Li H, Wang Y and Zhao T (2009) The segmentation of color milk somatic cells images. *2nd International Congress on Image and Signal Processing*, Tianjin, pp. 1–4.

https://doi.org/10.1017/S0022029921000042 Published online by Cambridge University Press