Talanta 121 (2014) 144–152

Contents lists available at ScienceDirect

Talanta

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

Determination of fat and total protein content in milk using conventional digital imaging

Sergey Kucheryavskiy [a,*], Anastasiia Melenteva [b], Andrey Bogomolov [b,c]

[a] Section of Chemical Engineering, Aalborg University, Campus Esbjerg, Niels Bohrs vej, 8, Esbjerg 6700, Denmark
[b] Samara State Technical University, Samara, Russia
[c] J&M Analytik AG, Essingen, Germany

ARTICLE INFO

Article history:
Received 27 July 2013
Received in revised form
3 December 2013
Accepted 24 December 2013
Available online 3 January 2014

Keywords:
Milk quality
Digital imaging
Image analysis
Light scatter

ABSTRACT

The applicability of conventional digital imaging to quantitative determination of fat and total protein in cow's milk, based on the phenomenon of light scatter, has been proved. A new algorithm for extracting features from digital images of milk samples has been developed. The algorithm takes into account spatial distribution of light, diffusely transmitted through a sample.

The proposed method has been tested on two sample sets prepared from industrial raw milk standards, with variable fat and protein content. Partial Least-Squares (PLS) regression on the features calculated from images of monochromatically illuminated milk samples resulted in models with high prediction performance when analysed the sets separately (best models with cross-validated $R^2=0.974$ for protein and $R^2=0.973$ for fat content).

However when analysed the sets jointly with the obtained results were significantly worse (best models with cross-validated $R^2=0.890$ for fat content and $R^2=0.720$ for protein content). The results have been compared with previously published Vis/SW-NIR spectroscopic study of similar samples.

1. Introduction

Efficient routine analysis of milk quality is of critical importance for any dairy production. Fat and protein content are two particularly important milk quality parameters, characterising its nutritional value. Nowadays, traditional physicochemical analysis of milk tends to be replaced by modern optical spectroscopic techniques combined with multivariate data analysis. Thus, mid-infrared spectroscopy has been widely accepted as a laboratory standard for the milk nutrient analysis [1]. At the same time, constantly growing demand for real-time milk analysis stimulates the development of alternative techniques capable of performing in-line or field measurements. An effective real-time technique should provide high throughput and reliability of analysis at a reasonable price.

The present work gives further development to the idea of exploiting the phenomenon of light scattering by fat and protein particles for their quantitative analysis. Early turbidimetric analysis was based on the observed correlation between the fat content and the detected extinction of light dispersed by a milk sample at individual wavelengths [2,3]. This method, however, is highly susceptible to the size variability of colloidal milk particles, even in homogenised milk, and thus, is now considered obsolete. There are only a few later works making use of the scatter for milk fat and protein analysis. They typically require an intensive pretreatment of milk samples, i.e. deep homogenisation and protein dispersion [4], and thus, are impracticable in the case of raw natural milk. The light propagation in the raw milk also stays too complex for a direct theory-based solution, due to the presence of two species having complex and varying size distributions under the conditions of multiple scattering. As a consequence, optical spectroscopic methods of fat and protein determination are mainly based on the components' absorption, provided that the scatter is possibly avoided or suppressed [5]. The visible (Vis) light region (360–780 nm), where the scatter strongly dominates, is rarely used in quantitative milk analysis [6,7]. At the same time, the Vis region is very attractive for the analysis, because of a wide choice of available equipment, including light sources and guides, optics and detectors.

The feasibility of scatter-based quantitative analysis of fat and total protein in the raw milk using Vis and short-wave near infrared (SW-NIR) spectroscopy has been recently proved by Bogomolov et al. [8–10], the difference of individual spectral patterns (i.e. wavelength dependencies) of scatter by differently sized protein and fat particles was shown to be sufficient for their quantitative analysis using formal multivariate modelling, e.g. PLS regression. The method successfully handles an artificially introduced variation of fat globule sizes [8].

* Corresponding author. Tel.: +45 2787 9818.
E-mail address: svk@bio.aau.dk (S. Kucheryavskiy).

0039-9140/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.talanta.2013.12.055

The present study develops further the ideas of using light scatter effect for analysis of milk quality published in [8,9] and aims at the elaboration of a technologically simple approach to the quantitative analysis of raw milk fat and protein content, using light emitting diodes (LED) as monochromatic light sources and conventional digital RGB imaging as a detection technique. This combination, in fact, is an inexpensive alternative to optical spectroscopy. In this novel approach, an essential loss of spectral resolution is compensated by the detection area extension, thus, giving an advantage of detecting spatial intensity distribution of the scattered light.

Conventional digital imaging nowadays is increasingly used for quantitative analysis in industrial applications [11–13] but stays an uncommon tool for the milk analysis. The published work has been mainly devoted to a microscopic investigation of milk particles and their aggregates, e.g. [14–16]. To our knowledge, digital imaging has not been used for the quantitative analysis of milk constituents before. General feasibility of this approach was suggested in [17], where the opposite problem had been solved: rendering images of different media by their scatter and absorption properties using the Lorenz–Mie scattering theory generalisation.

This study presents the evaluation results of using conventional RGB digital imaging and light emitting diode (LED) illumination for quantitative determination of fat and total protein in raw milk. The modelling and validation is based on a designed experiment including the variability of fat globule sizes. Initially, intensity histograms, first-order statistics and Angle Measure Technique (AMT) have been tried as image features. Finally, a simple but efficient feature extraction algorithm, which takes into account spatial intensity distribution on the milk images, was developed. The PLS regression on features, calculated using the developed algorithm, gave models with practically relevant prediction performances confirming the feasibility of suggested approach. The results were compared with a recent Vis/SW-NIR spectroscopic study of the same sample set [8].

2. Materials and methods

2.1. Samples

The experimental samples were prepared from two sets of raw milk standards (QSE GmbH, Wolnzach, Germany) with predominantly varying fat or protein content – *F*- and *P*-set, respectively (Table 1). Sixteen samples were prepared from each set: four initial standards (with known fat and total protein content) and twelve their pair-wise mixtures in proportions 1:2 and 2:1. Every sample was analysed three times: in its original state and after two subsequent homogenisations: for 10 and 20 s, using an ultrasound homogeniser. The homogenisation was applied to introduce gentle variation of particle size distribution occurring in the natural milk and significantly affecting its scattering properties [10,18]. Sample homogenisation degree was qualitatively characterised by optical microscopy and Vis/SW-NIR spectroscopy [8]. Spectral changes caused by applied homogenisation times were comparable in magnitude with the effects fat and protein content differences in the chosen range, and therefore, presented an essential factor of sample variability.

Thereby each of the two sets was represented by three measurements (one for each homogenisation degree) of 16 samples, which gave 48 measurements per set (96 in total). Further information about the samples can be found in [8].

Table 1
Fat and total protein content in raw milk standards.

Sample	Fat, % w/w	Protein, % w/w
P1	3.63	2.99
P2	4.27	3.30
P3	4.03	3.71
P4	4.33	4.05
F1	1.99	3.45
F2	3.23	3.47
F3	4.22	3.61
F4	5.47	3.21

2.2. Image acquisition and preprocessing

Images were acquired with DSLR camera Canon 400D fixed on a tripod. For every measurement, 4 ml of milk was put into a Petri dish (inner diameter of 30 mm) placed in front of the camera so that the centre of the dish was coaxial with the lens optical axis. The thickness of milk layer in the dish was about 4 mm. Three powerful LEDs emitting blue (maximum intensity at 465 nm) green (526 nm), and red (630 nm) light, were used for sample illumination. The emission spectra of the diodes are shown in Fig. 1. The incident light was delivered through a fibre optical guide with 1.2 mm diameter, which was coaxially mounted at 90° to the to the Petri dish bottom. Thereby every image captured a LED light spot diffusely transmitted through the milk sample. The image acquisition was performed in a dark room at 24 ± 1 °C.

Five photos for each light source were taken using bracketing with exposition times of 1, 1/2, 1/4, 1/8, and 1/15 s. Therefore, every measurement was represented by 15 images. The acquired images were cropped automatically to remove the dish walls.

The cropping algorithm worked with grayscale representation of the images. For each image it found a light spot with maximum intensity using threshold segmentation, estimated the centre of the spot and cropped an image using equal distances from the centre, resulting in symmetric and easily comparable pictures. The final images had size of 1024×1024 pixels. Fig. 2 shows a full set of preprocessed images taken from one of the standards after 10 s of homogenisation.

The high dynamic range (HDR) images were also made for every bracketing series using an algorithm described in [19] and implemented in Matlab Image Processing Toolbox function *makehdr*. In this case every measurement was represented by three HDR images – one for each light source.

The purpose of using HDR images was twofold: to decrease the number of variables taking advantage of enhanced image dynamic range provided by the bracketing technique. HDR technique allows

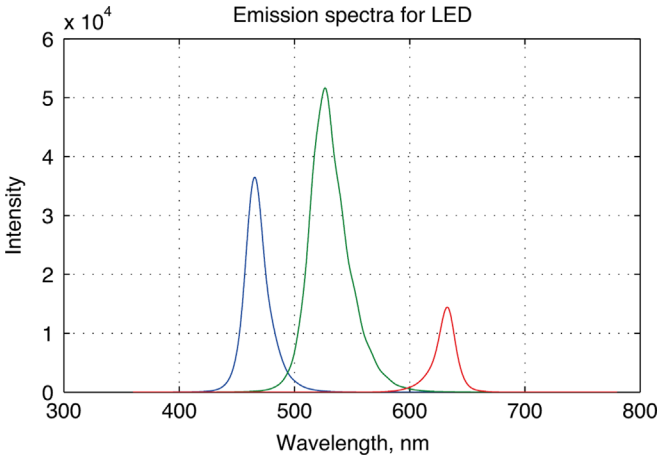

Fig. 1. Emission spectra of the used LEDs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)