Tea classification and quality assessment using laser-induced fluorescence and chemometric evaluation

Mei, Liang; Lundin, Patrik; Brydegaard, Mikkel; Gong, Shuying; Tang, Desong; Somesfalean, Gabriel; He, Sailing; Svanberg, Sune

Published in:
Applied Optics

DOI:
10.1364/AO.51.000803

2012

Link to publication

Citation for published version (APA):

Total number of authors:
8
Laser-induced fluorescence was used to evaluate the classification and quality of Chinese oolong teas and jasmine teas. The fluorescence of four different types of Chinese oolong teas—Guangdong oolong, North Fujian oolong, South Fujian oolong, and Taiwan oolong was recorded and singular value decomposition was used to describe the autofluorescence of the tea samples. Linear discriminant analysis was used to train a predictive chemometric model and a leave-one-out methodology was used to classify the types and evaluate the quality of the tea samples. The predicted classification of the oolong teas and the grade of the jasmine teas were estimated using this method. The agreement between the grades evaluated by the tea experts and by the chemometric model shows the potential of this technique to be used for practical assessment of tea grades. © 2012 Optical Society of America

1 Introduction

Tea is a popular beverage used throughout the world, especially in China where people started drinking tea thousands of years ago [1]. There are numerous scientific reports about the beneficial effects of tea on human health [2–4]. The quality of tea is dependent on many factors, such as the growing environment [5] and the processing technique [6,7]. Repeated or constant quality controls for the manufacturing in the tea plant facility are also of highest importance to obtain a high standard tea. It is rather difficult for a nonspecialist to judge the quality of the tea on the market and there are also plenty of counterfeit teas which are of poor quality but still sold at high prices. Since a number of factors can influence the quality of tea, its assessment is a systematic and complicated process in which the integrated effect of color, aroma, taste as well as shape of the tea leaves should be considered. At present, the assessment of the tea quality is commonly made by sensory evaluation from tea experts, but this is unfortunately susceptible to influence of the environment and subjectivity. To make the evaluation more efficient and objective, several tea experts are usually summoned to assess the
the quality of the jasmine teas depends on the freshness and tenderness of the tea leaves and it is also affected by the white petals which are added into the tea during manufacturing. The jasmine tea samples used were provided by a specialized company. The tea qualities were determined by 2–3 company experts according to the Chinese National Standard for jasmine tea (GB/T 22292-2008).

B. Experimental Setup

The LIF experimental setup is depicted in Fig. 1. The third harmonic (355 nm) of a Nd:YAG laser with an energy of 3 mJ/pulse at 20 Hz and an approximate pulse duration of 10 ns was used to induce fluorescence in the tea samples. The samples were placed on an anodized aluminum plate with negligible fluorescence. The emission from the tea samples was filtered by a long pass filter (Edmund Optics, L38, 380 nm, 2.6 mm thickness) before it was recorded with an optical multichannel analyzer (OMA) composed by a crossed Czerny–Turner grating spectrometer (ORIEL instruments, Model 77400) connected to an image-intensified and 10 ns time-gateable charged coupled device (ANDOR Technology, Model DH501-25U-01). The fluorescence measurements for each sample were performed five times in a so-called pseudoreplication procedure, where different sections of the sample were illuminated in each recording. The tea samples were kept in movement while being irradiated by 100 laser pulses per measurement. A halogen filament lamp (Oriel, Model No. 63355, wavelength range, 250–2500 nm) was used to calibrate the response curve of the OMA system. The wavelength scale of the OMA system was calibrated by a cadmium lamp and a HeNe laser at 632.8 nm. The spectral resolution is 6.5 nm. The spectrometer does not have any second-order rejection filters, but the influence of the second-order contribution in the 700–800 nm region is minor.

Standard red-green-blue (RGB) photographs of the tea samples were acquired by a Panasonic digital camera (DMC-FX01) in a flash mode inside an optically isolated box. Crossed polarizers were put in front of the flash lamp and the objective in order to obtain a depolarized reflective image. This ensures

2 Materials and Methods

A. Tea

Oolong and jasmine teas were investigated in the present work. The oolong tea samples are from different regions of China, and are provided by different companies. The classification is assessed by tea experts through the color, shape, smell, taste, and the unique processing technique of the tea samples. The quality of the jasmine teas depends on the freshness and tenderness of the tea leaves and it is also affected by the white petals which are added into the tea during manufacturing. The jasmine tea samples used were provided by a specialized company. The tea qualities were determined by 2–3 company experts according to the Chinese National Standard for jasmine tea (GB/T 22292-2008).

B. Experimental Setup

The LIF experimental setup is depicted in Fig. 1. The third harmonic (355 nm) of a Nd:YAG laser with an energy of 3 mJ/pulse at 20 Hz and an approximate pulse duration of 10 ns was used to induce fluorescence in the tea samples. The samples were placed on an anodized aluminum plate with negligible fluorescence. The emission from the tea samples was filtered by a long pass filter (Edmund Optics, L38, 380 nm, 2.6 mm thickness) before it was recorded with an optical multichannel analyzer (OMA) composed by a crossed Czerny–Turner grating spectrometer (ORIEL instruments, Model 77400) connected to an image-intensified and 10 ns time-gateable charged coupled device (ANDOR Technology, Model DH501-25U-01). The fluorescence measurements for each sample were performed five times in a so-called pseudoreplication procedure, where different sections of the sample were illuminated in each recording. The tea samples were kept in movement while being irradiated by 100 laser pulses per measurement. A halogen filament lamp (Oriel, Model No. 63355, wavelength range, 250–2500 nm) was used to calibrate the response curve of the OMA system. The wavelength scale of the OMA system was calibrated by a cadmium lamp and a HeNe laser at 632.8 nm. The spectral resolution is 6.5 nm. The spectrometer does not have any second-order rejection filters, but the influence of the second-order contribution in the 700–800 nm region is minor.

Standard red-green-blue (RGB) photographs of the tea samples were acquired by a Panasonic digital camera (DMC-FX01) in a flash mode inside an optically isolated box. Crossed polarizers were put in front of the flash lamp and the objective in order to obtain a depolarized reflective image. This ensures
that specular reflexes are rejected and that image contrast is enhanced. The images are flat-field and white-calibrated by spatial interpolation, using a white paper included in parts of the photographs.

C. Singular Value Decomposition

Singular value decomposition (SVD) is an orthogonal linear transformation applied on the recorded data, which is widely used in chemometry. The original data matrix is transformed into a new coordinate system, where the original data can be represented by a few principal components. The procedure facilitates the extraction of the important information and leads to substantial reduction of redundant data features.

Considering the fluorescence spectra of \( N \) samples being recorded by a spectrometer which has \( P \) spectral bands for one measurement, the whole data set can form a \( N \) row, \( P \) column matrix, which is denoted by \( S \). Its SVD gives

\[
S = U\Sigma V^T
\]

where \( U \) is the normalized eigenvector of matrix \( SS^T \) which indicates the contribution of each principal component to the sample, \( \Sigma \) is the square root of the diagonal matrix of the eigenvalues of the matrices \( SS^T \) and \( S^TS \) which indicates the importance of the corresponding principal components, \( V \) is the normalized eigenvector of the matrix \( S^TS \) which gives the principal components of the decomposition.

If the number of samples (\( N \)) is much smaller than the number of data sampling points (\( P \)), most of the principal components contribute insignificantly to the reconstitution of the original data. To compress the information, we therefore select only the most important principal components to represent the original data. The important components can be selected by a scree or elbow test [31]. The eigenvalues of the principal components (\( \Sigma \)) are plotted according to their size and the point where the slope of the size of the eigenvalues goes from “steep” to “flat” is found (this is often called the elbow). A truncation before the elbow is determined and is denoted by \( \text{tr} \). The matrix \( S \) can be approximately described as follows:

\[
S = U_{1...N1...N}\Sigma_{1...N1...P}\Sigma_{1...P1...P}^T
\]

Here \( U_{1...N1...N} \) and \( \Sigma_{1...P1...P} \) are the first \( \text{tr} \) principal components of matrices \( U \) and \( \Sigma \), respectively.

\( \Sigma_{1...P1...P}^T \) is the transpose of the first \( \text{tr} \) principal components of matrix \( V \).

D. Linear Discriminant Analysis

In order to predict the classification or quality of the tea sample, we can establish a leave-one-out predictive model which describes how the principal components contribute to the classification or quality. In this predictive model, each sample is left out one by one and the rest are used to build the model. If the sample left out is denoted by \( i \) and the number of variables to be predicted is denoted by \( M \), the prediction model can be described as follows:

\[
Y_{k=1...N,k\neq i1...M} = \phi_{k=1...N,k\neq i0...\text{tr}1...M}^T \theta_{0...\text{tr}1...M}
\]

Here \( Y_{k=1...N,k\neq i1...M} \) is the predefined matrix of the classifications or qualities which are given by the tea experts, \( U_{k,j} \) is the contribution of the \( j \)th principal component to the sample \( k \), \( \theta_{0...\text{tr}1...M} \) is the linear coefficient matrix which needs to be solved by regression of Eq. (3). The matrix \( \phi \) is referred to as the regressor and the bias in its first column ensures that zero quality or a negative classification does not necessarily imply that the sample cannot fluoresce. In classification, this kind of linear equation is referred to as linear discriminant analysis (LDA). The classification or quality of the tea sample which has been left out can be predicted as

\[
\hat{Y}_{1...M} = \phi_{0...\text{tr}0...\text{tr}1...M}
\]

By performing the predictive model for each sample, we can obtain the predicted classifications and
qualities while exploiting the sample size in the best possible way.

3. Results

A. Classification for Different Types of Oolong Tea

Sixteen oolong tea samples originating from different locations in China—Guangdong (GD, Fig. 2), North Fujian (N-FJ, Fig. 3), South Fujian (S-FJ, Fig. 4), and Taiwan (TW, Fig. 5), were used for the measurements. The names, types, and corresponding abbreviations of the 16 oolong tea samples are given in Table 1.

The GD and N-FJ teas have similar shapes and the color of the tea leaves is basically black. The S-FJ and TW teas are granulated, the leaf color of the TW teas is close to black while the leaf color of the S-FJ teas is a bit green. Through visual inspection by nonspecialists it is very difficult to distinguish between these similar tea leaves, while the fluorescence of the four types of teas is very different, something that can be used to classify the tea samples. The average fluorescence spectra of the five recordings for each tea sample are shown in Fig. 6. The spectra are normalized by the fluorescence intensity of the chlorophyll around 690 nm. The peak around 355 nm is the elastic scattering of the excited wavelength.

Since each sample was measured five times, 80 pseudoreplicated tea sample recordings were obtained. All the pseudoreplicated tea sample recordings were numbered from 1 to 80. The types of the tea samples are denoted as \(m/0.0136\), which represents GD, N-FJ, S-FJ, and TW, respectively. The predefined matrix for the classifications can be given as

\[
Y_{n,m} = \begin{cases} 
1 & n \in \text{type } m \\
0 & n \notin \text{type } m 
\end{cases} \quad n = 1, 2, \ldots N. \tag{6}
\]

In order to describe the quality of the classification, we define a discrimination index \(Q\) for each type according to Eq. (7).

\[
Q_m = \frac{\left| \mu_{Y_{n,m}} - \mu_{Y_{n,\overline{m}}} \right|}{\sqrt{\sigma^2_{Y_{n,m}} - \sigma^2_{Y_{n,\overline{m}}}}}, \tag{7}
\]

Here \(\mu_{Y_{n,m}}\) and \(\mu_{Y_{n,\overline{m}}}\) are the mean output of the predictive classification algorithm for type \(m\), \(\sigma^2_{Y_{n,m}}\) and \(\sigma^2_{Y_{n,\overline{m}}}\) are the standard deviations corresponding to \(\mu_{Y_{n,m}}\) and \(\mu_{Y_{n,\overline{m}}}\), respectively, \(Q_m\) is the discrimination index for type \(m\).

The eigenvalues are presented in Fig. 7. According to the principle of an elbow test \([31]\), the truncation value is determined as 9 (\(\text{tr} = 9\)). As seen in this figure, the eigenvalues of the first three principal components are much higher than the others which mean that these components have much higher significance. The principal components used in the predictive model are given in Fig. 8. By evaluating Eq. (5) for each sample, we can obtain the predicted classifications and the discrimination indices, as shown in Fig. 9. The discrimination indices for the four different types of tea samples are 3.02, 7.37,
2.45, and 2.86, respectively. It shows that both the Guangdong and the north Fujian oolong teas are very easy to classify from the other tea samples, while the tea samples from south Fujian and Taiwan would be somewhat more difficult to distinguish from the other.

B. Grade Assessment for Jasmine Teas

Jasmine tea is made of jasmine flowers and thus a few white petals can be observed in jasmine tea samples (Fig. 10). The grade cannot easily be assessed by only observing the color and texture of the jasmine tea, as the samples look very similar to each other. In the fluorescence experiment, ten jasmine tea samples were measured and each sample was again measured five times by illuminating different sections of the tea samples with laser light, which gives 50 pseudo tea sample recordings. All the jasmine teas are denoted with different grades from 1 to 10 (1 being the best), which were assessed by the tea experts. The averaged experimental fluorescence spectra of the five recordings for each tea sample are shown in Fig. 11. According to the principle of the elbow test [31], the truncation can be determined as 3 (tr = 3). The eigenvalues and the first three principal components of the fluorescence spectra are shown in Fig. 12. Only one parameter is used to describe the quality corresponding to the so-called grade. The predefined matrix can be described as

\[ Y_{nt} = n \quad n = 1, 2, \ldots, 10, t = 1, 2, 3, 4, 5, \]  

where \( t \) is the measurement occasion. By evaluating Eq. (5), the predicted grade can be calculated. Since each sample has been measured five times, the predicted grades are also averaged five times, as shown in Fig. 13. The predicted grades do not have any upper or lower limitation and the deviation only means that the evaluated quality of the tea samples is lower or higher than the expert defined quality. As can be seen in Fig. 11, the fluorescence spectra of the three best samples (grade 1, 2, and 3) are almost overlapped in the near infrared region. This could be the main reason why these three samples are very difficult to discriminate, which also shows the limitation of the present fluorescence technique. However, it can be expected that by utilizing multiple excitation wavelengths, the quality of the three best tea samples could be better assessed.
samples could still be well separated. The prediction quality of the model can be described with the correlation between the predicted grade and the grade assessed by the experts. The corresponding correlation coefficient is in this case 0.986, which shows a very good prediction for tea grades. Because the experts do not evaluate the grade linearly, it is expected that the evaluated grades have some deviation from the expert grades. This could be overcome by introducing a sigmoidal link function in the predictive model as known from neurology, where the nonlinear expert grade can be considered by the link function. From this point of view, the predicted grades are very reliable. The experimental results show the potential for using this technique for grade assessment of jasmine tea.

Fig. 8. (Color online) Principal components: (a–i) correspond to the first through ninth components, respectively.

Fig. 9. (Color online) Predicted classifications for the oolong tea samples. The dotted lines are the mean predicted classification and the solid lines are the standard deviations corresponding to the mean predictive classification. $Q_1$ is the discrimination index for each type of tea.
4. Discussion

Our proof-of-principle measurements have demonstrated that the LIF technique in combination with the SVD and LDA evaluation methods can be used to accurately classify and evaluate tea samples. By setting a proper truncation, tea classification and quality assessment can be well performed using this analytical tool. Because of an insufficient amount of tea samples, pseudoreplication was used to increase the sample size. This procedure might not be the optimal choice for prediction, but it still reflects the situation if the method should be applied commercially. More samples can further increase the accuracy and make the predicted results more reliable. The investigation of more samples will be the target for future work. By setting a proper threshold, tea samples which are difficult for experts to classify could be easily identified using this technique.

As we can see from the classification results of oolong tea samples, some of them are not easy to distinguish since we only use one wavelength to induce fluorescence which limits the information obtained from the tea leaves. The best three jasmine tea samples are also very difficult to discriminate due to their similar fluorescence spectra. However, this could be improved by using multiwavelength excitation. We note that the evaluation variability between experts is generally lower than the system error of the present technique. However, human senses are prone to be less accurate due to fatigue or subjectivity, which is not the case for a spectroscopic-based instrument. We can also analyze the spatial chromatic variance of the tea image to understand the texture of the tea leaves which could be helpful for classifying the types and judging the grade. Further, the laser source could be replaced by several inexpensive and robust light emitting diodes (e.g., Roithner...

Fig. 10. (Color online) Jasmine tea pictures, (a–j) correspond to grade 1 to 10, respectively.

Fig. 11. (Color online) Fluorescence of jasmine teas. The fluorescence spectra have been averaged five times. The peak around 355 nm is the elastic scattering of the excited wavelength. The small peak around 710 nm is the second-order of the excitation wavelength. The weakness of this peak shows that the influence of the grating second-order spectrum in the 700–800 nm region is negligible. The fluorescence peak in the near infrared region is due to chlorophyll.

Fig. 12. (Color online) (a) Eigenvalues of the principal components, the first three eigenvalues represent most of the information; (b) the first three principal components.
References


