RAPID COMMUNICATION

Simultaneous Quantitative Analysis of Three Compounds Using Three-Dimensional Fluorescence Spectra Based on Digital Image Techniques

Hong Lin Zhai · Zhi Jie Shan · Rui Na Li · E Yu

Abstract Digital image processing has been applied on various fields such as classification and qualitative analysis. In this work, a very simple quantitative approach was proposed for the first time. Based on the digital grayscale images of three-dimensional fluorescence spectra, several wavelet moment invariants were calculated, and used to establish the linear models for the quantitative analysis. This approach was applied to the quantitative analysis of Tryptophan, Tyrosine and Phenylalanine in mixture samples, and the correlation coefficients R^2 of the obtained linear models were more than 0.99, which were supported by the strict statistical parameters as well as leave-one-out and Jackknife cross-validations. Our study indicates that the selected wavelet moment invariants are immune from the noise and background signals, and the quantitative analysis can be performed accurately based on the overlapping peaks of compounds in mixture. This proposed approach provides a novel pathway for the analysis of three-dimensional spectra.

Keywords Quantitative analysis · Three-dimensional fluorescence spectra · Digital image processing · Wavelet moment invariant · Cross-validation

H. L. Zhai
State Key Laboratory of Applied Organic Chemistry,
Lanzhou University,
Lanzhou 730000, People's Republic of China

H. L. Zhai (✉) · Z. J. Shan · R. N. Li · E. Yu
College of Chemistry & Chemical Engineering,
Lanzhou University,
Lanzhou 730000, People's Republic of China
e-mail: zhaihl@163.com

Introduction

With the increasing of environmental pollution, the quality and safety of foods or medicaments become more and more important. The sensitive and rapid analytical technologies for their monitoring and surveillance are indispensable. The routine methods, such as chromatographic separation, extraction, mass spectroscopy and atomic absorption spectroscopy, are expensive and time consuming due to the fussy pretreatments of samples and the exploration of strict experimental conditions. Therefore, the rapid, robust and inexpensive analytical methodologies are preferred for routine test. Many foods or medicaments contain a wide range of naturally occurring fluorescent compounds such as aromatic amino acids, vitamins, secondary metabolites, pigments, toxins and flavoring compounds, which is very useful in the detections of the quality and safety [1, 2].

Fluorescence spectroscopy is a versatile analytical technique and widely used in many fields such as food, biological sciences and environmental protection owing to its higher sensitivity and specificity, rapidity, nondestructive detection, better representation and lower limit of detection. Traditionally, the fluorescence spectra are obtained under a single excitation wavelength, which can't describe the characteristics of fluorescent compounds completely because that the intensities in fluorescence spectra are variable with the different excitation and emission wavelengths. Three-dimensional fluorescence spectrum, the complete fluorescence intensities under the excitation-emission matrix (EEM), is composed of the serial emission spectra under different excitation wavelengths, which provides the special fluorescence signal for one compound, and could be used in quantitative analysis. However, in multi-component mixtures, the fluorescence signals usually overlap each other that bring on the difficulty of quantitative analysis.

For the purpose of the quantitative analysis to the target compounds in complex mixtures without a previous physical or/and chemical separation, several analytical techniques of mathematical separation have been developed and widely accepted in the wake of the applications of high-order analytical instrumentation and multi-way data collection. The trilinear parallel factor analysis (PARAFAC) model, introduced in 1970 by Harshman [3], and Carroll and Chang [4], is one of the multi-way models, which refers to multivariate data analysis on the data arrays. A comparison of several algorithms for fitting the PARAFAC model was reported [5], in which the limits and advantages of the available methods were compared. Subsequently, the various algorithms using second-order advantage were published such as alternating trilinear decomposition (ATLD) [6], alternating penalty trilinear decomposition (APTLD) [7], multivariate curve resolution-alternating least-squares (MCR-ALS) [8]. Other techniques had been proposed such as N-way principal component analysis (N-PCA) [9] and N-way partial least squares (N-PLS) [10]. Several second-order multivariate calibration algorithms and their applications had been reviewed in references [11, 12].

It is well known that a trilinear model can provide reasonable results when the dataset is trilinear and the component number is selected correctly. However, there is not perfect situation in practice. Although the three-dimensional fluorescence intensities of chemical components possess the trilinearity, the signals derived from the Rayleigh and Raman scatterings, diffraction harmonics and background interferences are not trilinear. Usually, these problems are avoided by the selection of suitable wavelengths or the model improved with a specific algorithm.

Digital image techniques have been applied in various research fields including chemistry. The visual appearance of cereal flakes was employed for the quantitative quality control measurement [13]. An approach to the quantification of Lactobacillus in fermented milk was reported using digital color images [14]. Wavelet texture analysis was applied on the classification of paper surface quality [15]. In our previous work, based on the fingerprints obtained from high performance liquid chromatography coupled with diode array detector, digital image techniques were applied to the clustering analysis of different samples, and the densities in grayscale image were used for the quantitative analysis of three compounds [16].

In this present paper, based on digital image processing, another novel approach to the quantitative analysis of compounds in mixtures was proposed. Three-dimensional fluorescence spectrum was regarded as grayscale image, and then several wavelet moment invariants were calculated. The linear models, which present the relationship between the selected wavelet moment invariants and the concentration of compound, were established for three compounds in mixtures, respectively. The statistical evaluation indicates that the linear models have high reliability and accuracy.

Materials and Methodology

Samples and Data

The dataset employed in this work was obtained from the public database available in the website (http://www.models.kvl.dk/amino_acid_fluo). The samples were generated and measured by Claus A. Andersson, and its description can be referred to Ref. [17]. This dataset consisted of five simple laboratory-made samples, and each sample contained different amounts of Tryptophan (Trp), Tyrosine (Tyr) and Phenylalanine (Phe) dissolved in phosphate buffered water. The composition of these samples is listed in Table 1 as Exp. columns. The samples were measured by fluorescence (excitation 250–300 nm, emission 250–450 nm, 1 nm intervals) on a PE LS50B spectrofluorometer. Therefore, the structure of this dataset was provided as the matrix of 5 samples × 201 emissions × 61 excitations. The three-dimensional fluorescence spectrum of one sample is shown in Fig. 1.

As can be seen from Fig. 1, the fluorescence spectrum derived from the mixture sample including three compounds presents the overlapping of three peaks each other, and there were disturbances including noise and Rayleigh scattering signals. The quantitative analysis will be a hard task without special methods.

Analytical Method

The Image of Three-Dimensional Fluorescence Spectrum

The intensities in a three-dimensional fluorescence spectrum were recorded as an n-by-m matrix, in which the rows (n) corresponded to emission wavelengths and the columns (m) to excitation wavelengths. The two-dimensional data table (matrix) could be regarded as a "grayscale image", and this image can represent the intensities in the three-dimensional fluorescence spectrum completely. In other words, the densities in grayscale image agree well with the intensities in the three-dimensional fluorescence spectrum, which are proportional to the concentrations of the compounds in mixture. As shown in Fig. 2, the peaks in a three-dimensional fluorescence spectrum correspond to the different gray regions in the image.

Wavelet Moment Invariants

In image processing techniques, the characterization of image can be drawn with many features including color, textures, shape features, and so on. Image moment invariant is one kind of important features computed from measurements. Since Hu proposed the concept of moment invariant

Table 1 The composition and the calculated results of samples

Sample No.	Tryptophan(Trp)		Tyrosine(Tyr)		Phenylalanine(Phe)	
	Exp.	Calc.	Exp.	Calc.	Exp.	Calc.
1	2.67e–06	2.67E–06	0	4.14E–07	0	3.96E–06
2	0	7.96E–09	1.33e–05	1.33E–05	0	−2.68E–06
3	0	7.99E–10	0	−1.59E–08	9.00e–4	9.01E–04
4	1.58e–06	1.59E–06	5.44e–06	5.89E–06	3.55e–4	3.49E–04
5	8.79e–07	8.59E–07	4.40e–06	3.53E–06	2.97e–4	3.01E–04

in 1962 [18], several types of moment invariants have been developed and applied to pattern recognition such as Zernike moments [19], Geometric moments [20], Legendre moments [21] and Fourier-Mellin moments [22]. Wavelet moment methods have been attracted more attentions and gained extensive applications owing to their multi-resolution property [23–26].

In this study, wavelet moment invariants were employed to describe the features in the grayscale image of three-dimensional fluorescence spectrum. Wavelet moment invariants have been well described in many monographs and articles. Here a brief description is given as fellows:

Let $f(x,y)$ represents the density distribution function of a image in Cartesian coordinates, and the geometric moments of $(p + q)$ order are defined as:

$$M_{pq} = \iint f(x,y)x^p y^q dx dy \quad (p, q = 0, 1, 2...) \quad (1)$$

In polar coordinates, there are $x = r\cos(\theta)$ and $y = r\sin(\theta)$. Substitute them into Eq. (1), and the general expression of moments can be obtained as follows:

$$F_{p,q} = \iint f(r,\theta)g_p(r)e^{jq\theta}r dr d\theta \quad (|r| \leq 1, 0 \leq \theta \leq 2\pi) \quad (2)$$

With the different expression of $g_p(r)$, the Eq. (2) gives the different moments:

$g_p(r) = r^p$ Hu moments
$g_p(r) = R_{m,n}(r)$ Zernike moments
$g_p(r) = \Psi_{m,n}(r)$ Wavelet moments

A set of wavelet basis functions can be generated by translation and scaling the mother wavelet, written as:

$$\psi_{a,b}(r) = \frac{1}{\sqrt{a}}\psi\left(\frac{r-b}{a}\right) \quad (3)$$

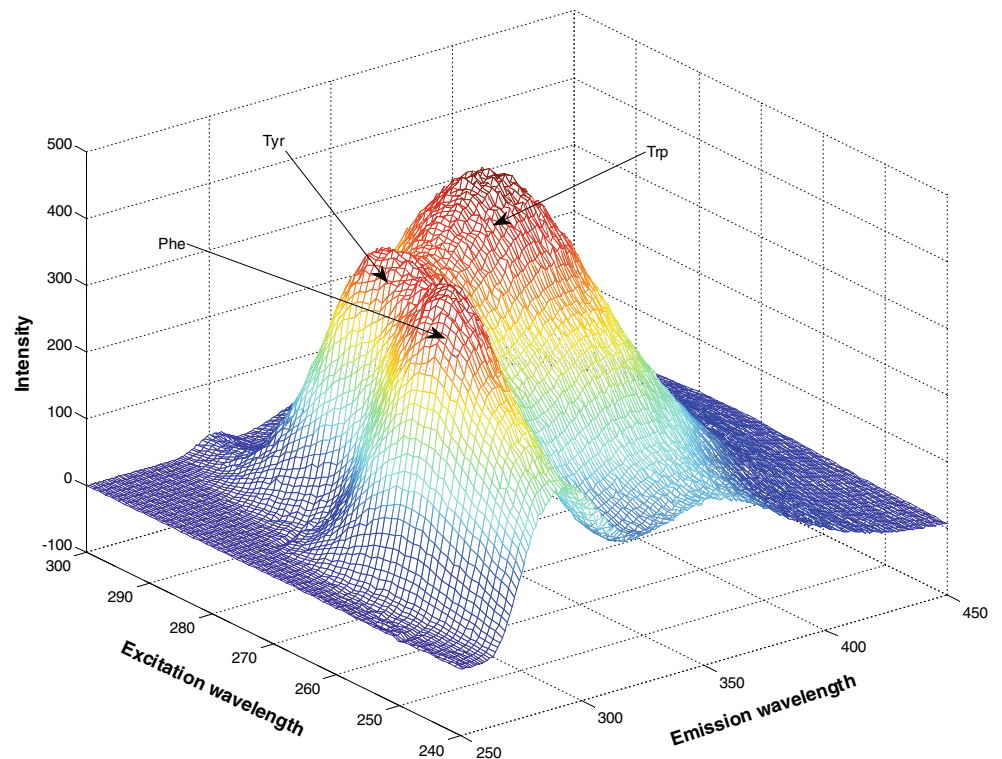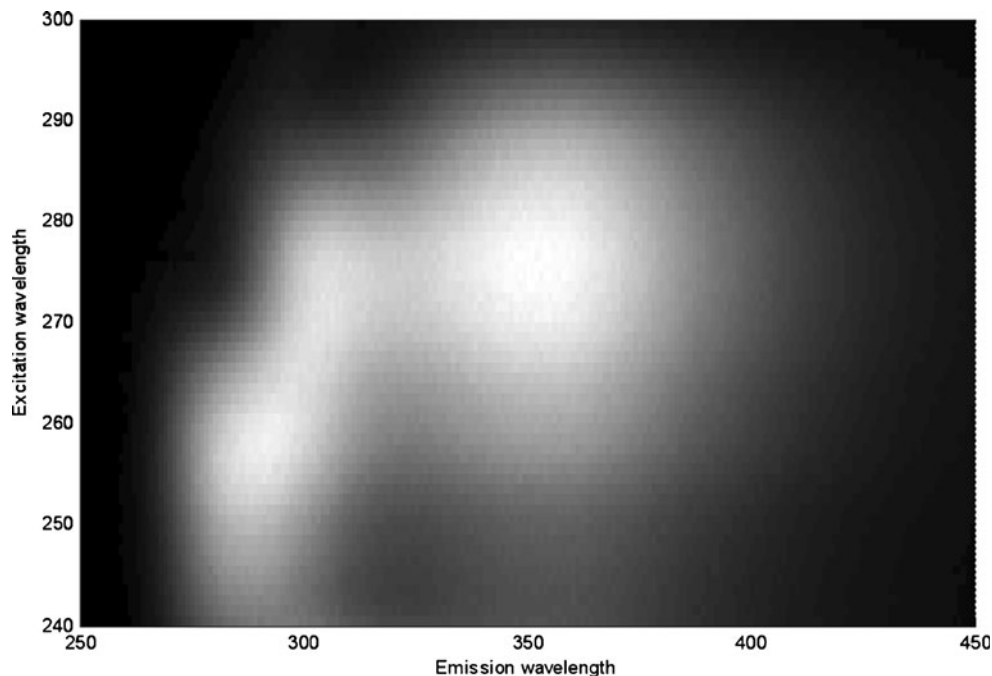**Fig. 1** Three-dimensional fluorescence spectrum of one sample

Fig. 2 The digital image of a three-dimensional fluorescence spectrum

where, a is the scaling factor, b is the shift factor, and $\Psi(r)$ is mother wavelet function. For the digital image, parameters a and b in Eq. (3) take only discrete values.

Suppose $a=2^{-m}$ ($m=0, 1, 2, …$) and $b=n2^{-m}$ ($n=0, 1, 2, …, 2^{m+1}$), the wavelet function (3) is changed into:

$$\psi_{m,n}(r) = 2^{\frac{m}{2}} \psi(2^m r - n) \qquad (4)$$
$$(m = 0, 1, 2, …; n = 0, 1, 2, …, 2^{m+1})$$

In this work, cubic B-spline function was adopted as mother wavelet function. Here it is presented:

$$\psi(r) = \frac{4a^{n+1}}{\sqrt{2\pi(n+1)}} \sigma_w \cos(2\pi f_0(2r-1)) \exp\left(-\frac{(2r-1)^2}{2\sigma_w^2(n+1)}\right) \qquad (5)$$

in which, $n=3$, $a=0.697066$, $f_0=0.409177$, and $\sigma_w^2 = 0.561145$.

From Eqs. (2) and (4), the expression of wavelet moments is obtained:

$$W_{mnq} = \iint f(r,\theta)\psi_{mn}(r)e^{jq\theta}rdrd\theta$$
$$= \int \psi_{mn}(r)r\left[\int f(r,\theta)e^{jq\theta}d\theta\right]dr \qquad (6)$$
$$= \int S_q(r)\psi_{mn}(r)rdr$$

where, $S_q(r) = \int f(r,\theta)e^{jq\theta}d\theta$

The wavelet moment invariants (F_{mnq}) are denoted from Eq. (6):

$$F_{mnq} = \|W_{mnq}\| = \left\|\int S_q(r)\psi_{mn}(r)rdr\right\|$$
$$\left(\begin{array}{l} m = 0, 1, 2, 3, …, n = 0, 1, …, 2^{m+1}, \\ \qquad \qquad \text{and } q = 0, 1, 2, 3, … \end{array}\right) \qquad (7)$$

Modeling and Evaluation

Several wavelet moment invariants were selected by stepwise regression (a systematic method for adding and removing independent variants from a multilinear model based on their statistical significance in a regression), and used to establish the final linear models, which reflect the relationship between the selected wavelet moment invariants and the concentrations of compounds. The performances of models were evaluated by means of their statistical parameters such as the correlation coefficient R^2, adjusted correlation coefficient R_{adj}^2, mean square error (*MSE*), *F*-test for the model and *t*-test for the regressive coefficients in the model. Furthermore, leave-one-out (LOO) and Jackknife cross-validations were employed to estimate the reliability of models.

All calculation programs were written in M-file based on MATLAB 7.0, and carried out with PC (CPU P4 1.6 GHz/ RAM 1 GB).

Results and Discussion

Moment Invariants

As a sort of statistical indexes to describe grayscale image, moment invariants possess invariability in many operations such as shifting, scaling and rotation. Wavelet transforms have become a powerful tool to reveal the properties of a signal in localized regions of the time and frequency space simultaneously. Wavelet moment invariants are constructed by the combination of wavelet

Table 2 The wavelet moment invariants in the models

Model No.	F_{mnq}	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Model 1 (for Trp)	F_{022}	2.4735E+03	8.4291E+02	2.9735E+02	1.6179E+03	1.2312E+03
	F_{331}	2.8491E+01	7.2162E+01	6.4143E+02	2.5119E+02	1.9843E+02
Model 2 (for Tyr)	F_{111}	8.0886E+03	4.7084E+03	9.4391E+03	8.3953E+03	8.1531E+03
	F_{112}	9.0140E+03	1.0556E+04	1.5314E+04	1.8601E+04	1.3912E+04
Model 3 (for Phe)	F_{223}	4.1496E+02	4.2937E+02	4.0109E+03	2.0073E+03	1.9729E+03
	F_{300}	1.5854E+03	1.4805E+03	1.4144E+03	9.7498E+02	5.7701E+02

transforms and image moment invariants. There are several mother wavelet functions, in which cubic B-spline wavelet function is not only near optimal in terms of its time-frequency space localization, but also takes advantage of the wavelet inherent property of multi-resolution analysis. Furthermore, the cubic B-spline wavelet function has a compact support. Therefore, the wavelet moment invariants based on cubic B-spline wavelet function are capable to extract multi-resolution features from grayscale image effectively. Taking the advantage of multi-resolution property, we can select several wavelet moment invariants used in quantitative analysis for the interested compounds in mixture.

The invariability of wavelet moment invariants with higher orders is decreased due to the increasing error of discrete numerical method. In this work, the calculation of wavelet moment invariants was carried out within the three of maxim order (m) and the four of maxim q in Eq. (7). Then a total of 136 wavelet moment invariants were obtained, and the wavelet moment invariants selected in the models are listed in Table 2.

Linear Models

By means of the selection with stepwise regression method, several wavelet moment invariants (listed in Table 2) were

used to establish the final linear models for the quantitative analysis of three compounds, respectively. The linear models obtained are listed as follows:

> Model 1 for Trp: $C_{Trp} = -1.5188E\text{-}06 + 1.6749E\text{-}09F_{022} + 1.5926E\text{-}09F_{331}$
> Model 2 for Tyr: $C_{Tyr} = 2.2611E\text{-}05 - 3.0509E\text{-}09F_{111} + 6.8336E\text{-}10F_{112}$
> Model 3 for Phe: $C_{Phe} = -2.572E\text{-}04 + 2.542E\text{-}07F_{223} + 9.822E\text{-}08F_{300}$

The performances of these models are displayed in Table 3, the concentrations calculated with these model are listed in Table 1 (as Calc. columns), and the comparison of concentrations between experimental values and calculated values is shown as Fig. 3.

As can been seen from Table 3 and Fig. 3, the correlation coefficient R^2 and adjusted correlation coefficient R_{adj}^2 are very satisfactory. The results of t-test indicate that the regressive coefficients have statistical significance. F-test for the models proves that there have best linear relationships between the concentrations of three compounds in mixture samples and the selected wavelet moment invariants that derived from the image of the three-dimensional fluorescence spectra. These statistical parameters indicate that the linear models are reliable.

Table 3 The performances of the linear models

Model No.	R^2	R_{adj}^2	MSE	F-test		R_{cv}^2	R_{jv}^2
				F-value	P-value		
Model 1 (for Trp)	0.9999	0.9998	2.848E–16	9106	1.1E–4	0.9994	0.9993
Model 2 (for Tyr)	0.9905	0.9811	5.612E–13	105	9.5E–3	0.9573	0.9359
Model 3 (for Phe)	0.9999	0.9997	3.940E–11	6884	1.4E–4	0.9990	0.9986

R^2 correlation coefficient

R_{adj}^2 adjusted correlation coefficient

MSE mean square error

R_{cv}^2 correlation coefficient with LOO cross-validation

R_{jv}^2 correlation coefficient with Jackknife cross-validation
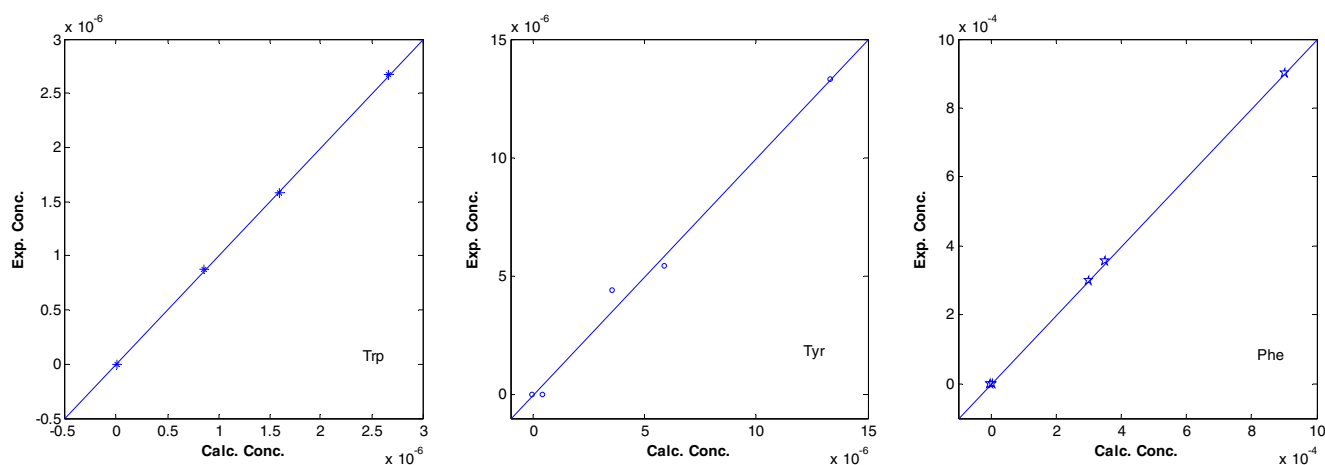
Fig. 3 Experimental concentrations vs. calculated concentrations

The Validation of Models

Cross-validation is a cunning technique for assessing how the results of a model will generalize to an independent data set. Here both LOO and Jackknife cross-validations were employed to validate the obtained linear models.

(1) LOO validation

LOO cross-validation procedure consists of removing one sample from the training set, reconstructing the model on the basis only of the remaining training data, and then predicting on the removed sample. This is repeated such that each sample in training subset is used once as the validation data. Then the correlation coefficient R_{cv}^2 is computed based on the predicted values and experimental values, which is used to evaluate the generalization ability of the model. In our study, the values of R_{cv}^2 are more than 0.95 for three models, and demonstrate the creditability of obtained models.

(2) Jackknife validation

As another method, the Jackknife cross-validation uses resampling to estimate the bias of a sample statistic, which is thought the most objective and rigorous way in comparison with sub-sampling test or independent dataset test, and usually employed for assessing the uncertainty in parameters estimates [27]. During the process of Jackknife analysis, a sample will be in turn moved from each to the other, and all the rule parameters are calculated based on the remaining. The information loss resulting from jackknifing will have a greater impact on the small subsets than the large ones. Nevertheless, as shown in Table 3, the Jackknife correlation coefficients (R_{jv}^2) are satisfactory.

The results of LOO and Jackknife validation support that our approach is efficient, and the obtained models could be used in the quantitative analysis.

The Comparison Between PARAFAC and Our Approach

A three-way data array can be constructed with the matrices of different samples, which presents trilinearity in the fluorescence intensity, excitation emission and wavelengths. PARAFAC is one of important analytical methods for the three-dimensional fluorescence spectra of mixture. However, there are still several hard nuts to crack. First, it is difficult to decide the best rank of a three-way array in PARAFAC model due to many factors such as noise signals and too many components. Second, Rayleigh scatter is not trilinear in its nature, and should be avoided or eliminated in the decomposition. In the analysis of this dataset by PARAFAC [17], compared with the four-component model, the three-component model was readily accepted because that the three loadings resemble the pure spectra of three compounds. The fourth component seemed to be scatter effects that caused some additional systematic variation. In order to eliminate the influence of Rayleigh scatter, the region of excitation wavelengths (range: 240–249 nm) was removed. Hence, the data array decomposed is only 5 samples×51 excitations×201emissions.

Owing to the characteristic of multi-resolution in the time-frequency space simultaneously, wavelet moment invariants with different orders can represent the various signals (such as noises, compounds and Rayleigh scatter) in the grayscale image of three-dimensional fluorescence

Table 4 The comparison of results between PARAFAC and our approach

Compound	Exp. conc.	PARAFAC	Our approach
Trp	8.79e–7	7.8E–7	8.60E–7
Tyr	4.40e–6	3.5E–6	3.54E–6
Phe	2.97e–4	2.3E–4	3.01E–4

spectrum. Therefore, in our approach, the quantitative analysis for the three compounds in mixture can be done directly in whole data array without any cutting. It is easy to find out the special wavelet moment invariants that only correspond to the signals of interested compounds. Thus the noise and Rayleigh scatter signals cannot affect the analytical results. As shown in Table 4, the satisfactory results were obtained by our approach.

Conclusion

The quantitative analysis of multi-target compounds in mixture is always important issue in analytical chemistry. Three-dimensional fluorescence spectrum provides the enough information for the quantitative analysis of interested compounds in mixture. Utilizing the techniques of digital image processing, especially the characteristic of multi-resolution in wavelet moment invariants of grayscale image, a novel and simple approach to the quantitative analysis was developed. Our study indicates that this approach is more efficient, and the results obtained are reliable and accurate, which could decrease the difficulty of separation and the analytical costs for mixtures. This idea can also be further expanded to the analyses of other three-dimensional spectra.

References

1. Christensen J, Nørgaard L, Bro R, Engelsen SB (2006) Multivariate autofluorescence of intact food systems. Chem Rev 106:1979–1994
2. Strasburg GM, Ludescher RD (1995) Theory and applications of fluorescence spectroscopy in food research. Trends Food Sci Technol 6:69–75
3. Harshman RA (1970) Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multi-modal factor analysis. UCLA Working Papers in Phonetics 16:1–84
4. Carroll JD, Chang JJ (1970) Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika 35:283–319
5. Tomasi G, Bro R (2006) A comparison of algorithms for fitting the PARAFAC model. Comput Stat Data Anal 50:1700–1734
6. Wu HL, Shibukawa M, Oguma K (1998) An alternating trilinear decomposition algorithm with application to calibration of HPLC-DAD for simultaneous determination of overlapped chlorinated aromatic hydrocarbons. J Chemom 12:1–26
7. Xia AL, Wu HL, Fang DM, Ding YJ, Hu LQ, Yu RQ (2005) Alternating penalty trilinear decomposition algorithm for second-order calibration with application to interference-free analysis of excitation-emission matrix fluorescence data. J Chemom 19:65–76
8. Tauler R (1995) Multivariate curve resolution applied to second order data. Chemometr Intell Lab Syst 30:133–146
9. Pravdova V, Boucon C, de Jong S, Walczak B, Massart DL (2002) Three-way principal component analysis applied to food analysis: an example. Anal Chim Acta 462:133–148
10. Bro R (1996) Multiway calibration. Multilinear PLS. J Chemom 10:47–61
11. Olivieri AC (2008) Analytical advantages of multivariate data processing. One, two, three, infinity? Anal Chem 80:5713–5720
12. Escandar GM, Olivieri AC, Faber NM, Goicoechea HC, Muñoz de la Peña A, Poppi RJ (2007) Second- and third-order multivariate calibration: data, algorithms and applications. Trends Anal Chem 26:752–765
13. Pereira A, Reis M, Saraiva P (2008) Quality control of food products using image analysis and multivariate statistical tools. Ind Eng Chem Res 48:988–998
14. Borin A, Ferrão MF, Mello C, Cordi L, Pataca LCM, Durán N, Poppi RJ (2007) Quantification of Lactobacillus in fermented milk by multivariate image analysis with least-squares support-vector machines. Anal Bioanal Chem 387:1105–1112
15. Reis MS, Bauer A (2010) Image-based classification of paper surface quality using wavelet texture analysis. Comput Chem Eng 34:2014–2021
16. Zhai HL, Hu FD, Huang XY, Chen JH (2010) The application of digital image recognition to the analysis of two-dimensional fingerprints. Anal Chim Acta 657:131–135
17. Bro R, PARAFAC (1997) Tutorial and applications. Chemometr Intell Lab Syst 38:149–171
18. Hu MK (1962) Visual pattern recognition by moment invariants. IRE Transactions on Information Theory 8:179–187
19. Liyun W, Hefei L, Fuhao Z, Zhengding L, Zhendi W (2009) Spermatogonium image recognition using Zernike moments. Comput Meth Prog Bio 95:10–22
20. Xiao X, Wang P, Chou KC (2008) Predicting protein structural classes with pseudo amino acid composition: an approach using geometric moments of cellular automaton image. J Theor Biol 254:691–696
21. Marengo E, Bobba M, Liparota MC, Robotti E, Righetti PG (2005) Use of Legendre moments for the fast comparison of two-dimensional polyacrylamide gel electrophoresis maps images. J Chromatogr A 1096:86–91
22. Chen YM, Chiang JH (2010) Fusing multiple features for fourier Mellin-based face recognition with single example image per person. Neurocomputing 73:3089–3096
23. Shen D, Ip HHS (1999) Discriminative wavelet shape descriptors for recognition of 2-D patterns. Pattern Recogn 32:151–166
24. Murtagh F, Starck JL (2008) Wavelet and curvelet moments for image classification: application to aggregate mixture grading. Pattern Recogn Lett 29:1557–1564
25. Reginska T, Wakulicz A (2009) Wavelet moment method for the Cauchy problem for the Helmholtz equation. J Comput Appl Math 223:218–229
26. Zhang F, Liu SQ, Wang DB, Guan W (2009) Aircraft recognition in infrared image using wavelet moment invariants. Image Vision Comput 27:313–318
27. Miller RG (1974) The jackknife-a review. Biometrika 61:1–15