Chemometrics and Intelligent Laboratory Systems 163 (2017) 41–48

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

# Signal processing for fast Raman imaging under low SNR condition and its applications in Raman imaging

Xin Wang, Hao He, Xian'guang Fan*

*School of Aerospace Engineering, Xiamen University, Xiamen 361005, China*

ARTICLE INFO

*Keywords:*
Raman imaging
Low SNR
CPSA
Sparse approximation

ABSTRACT

Imaging in Raman spectroscopy is a valuable tool for analytical chemistry, cell research and so on. At present, low SNR (signal-noise ratio) of the Raman spectral data set and the data processing speed of the imaging process are the two constraints of observing some dynamic change reactions. In this paper, a fast univariate Raman imaging algorithm named CPSA (Characteristic Peak Sparse Approximation) which is based on sparse approximation, least square and Butterworth low pass filter was presented to generate the Raman image under low SNR condition, ant it was proved to be timesaving as well as high imaging quality. In CPSA algorithm, sparse approximation is used to eliminate the noises of the target Raman signals; least square is used to enhance the useful Raman signals; Butterworth low pass filter is used to improve the image quality. Three experiments were conducted and the results were presented. The first experiment was used to validate the reliability of the proposed method, and the left two experiments were its applications in medical and human cancerous cell imaging.

## 1. Introduction

Raman spectroscopy, as a vibrational spectroscopy technique which owns many advantages such as noninvasive and fingerprint detection [1,2], plays a significant role in many fields such as pharmaceutical analysis [3,4], cell research [2,5], and food safety [6]. Raman imaging technique is a chemical characterization imaging method based on Raman spectroscopy, which can be performed with little sample preparation due to its highly sensitive to sample's conformation and chemical environment [7]. In terms of data acquisition method, Raman imaging technique can be divided into scanning imaging and single wavelength wide-field imaging, of which scanning imaging can be divided into point scanning imaging and line scanning imaging [7,8]. In this paper, we mainly study the scanning imaging, single wavelength wide-field imaging is not within the scope of this paper. For point scanning imaging, a point shaped specified monochrome laser is used to excited the sample point by point on the two-dimensional plane, and each scan produces a Raman spectrum corresponds to the point laser position on the sample. Line scanning imaging is an extension of point scanning imaging method which using a line shaped monochrome laser as the excitation light, and each scan produces Raman spectra of a line correspond to the line laser position on the sample. Therefore, compared with the point scanning method, the line scanning Raman imaging method significantly reduces the data acquisition time, but also leads to the resulting SNR of Raman spectral data set decreased. A Raman image can be obtained by mapping characteristic value of the Raman spectral data set onto the two-dimensional plane of the sample, in which these characteristic values include peak intensity, peak area and peak Raman shift and so on. The Raman images generated by different characteristic values have different physical meanings. For example, Raman peak shift images reflect the stress distribution in the material, peak area and peak intensity Raman images reflect the chemical components' distributions in the sample. In this paper, we mainly use the Raman image to study the distribution of chemical composition in the sample and the method we proposed below is not suitable for peak Raman shift imaging.

The developed instruments, for example, multi-focus Raman microscopy and silt-scanning Raman microscopy [9], as well as data processing methods may benefit the Raman imaging. These data processing methods can be summarized as multivariate analysis methods and univariate analysis methods, of which the multivariate analysis methods aim at researching the multi-components distribution in the sample which use the large part of the spectrum or entire spectrum to generate Raman images. Compared with multivariate method, the univariate method is used to research the specific chemical component's distribution in the sample through using a single characteristic value(single peak intensity, single peak area or peak ratio, etc) of Raman spectrum to generate image.

* Corresponding authors.
*E-mail address:* fanxg@xmu.edu.cn (X. Fan).

http://dx.doi.org/10.1016/j.chemolab.2017.01.022
Received 2 September 2016; Received in revised form 6 January 2017; Accepted 7 January 2017
Available online 20 February 2017

Many kinds of multivariate algorithms have been proposed to analysis the syntheitcal Raman spectral data set since from the Raman imaging technique was been invented. Typical multivariate Raman imaing methods such as KMCA(K-Means Clustering Analysis) [10], PF-MCR(Principle Factor-Multivariate Curve Resolution), OP-MCR(Othogonal Projection-Multivariate Curve Resolution) [11], MALS(Modified Alternative Least Square) [12,13] and MCR-LLM(Multivariate Curve Resolution by Log-Likelihood Maximization) [14], decompose the spectral data set into several clusters of pure spectra correspond to the different chemical components in the sample. Although multivariate methods have multi-components analysising ability, but they are not robust enough to deal the low SNR spectral data sets. Beside, since the Spectral data sets are usually consist of tens of thousands spectra, multivariate methods need much computational time. At last, some of the multivariate methods need prior knowledges of the chemical components which limit their universality [10,15]. However, in some circumstances, in order to observe the dynamic changes of one specific chemical component in the sample, fast line scanning should be adopted to finish the acquisition of spectral data set which means that the resulting SNR of the data set is bad. In this case, it is appropriate to use univariate methods to process the spectral data set rather than multivariate methods. Unfortunately, few scholars have studied the univariate Raman imaging analysis methods so far. The only feasible method at present seems that is to denoise every spectrum one by one in the huge data set and then substract the baselines of them finally imaging them. But this is too inefficient and computational large method which is hard to achieve. Therefore, fast univariate Raman imaging method should be developed and the challenge lies in cutting the processing time of the data set, meanwhile, maintaining high Raman image quality.

In this paper, we proposed an univariate analysis method based on sparse approximation aims at reconstructing Raman image under low SNR condition, which was proved to be timesaving as well as high imaging quality acquired. We call this algorithm as CPSA (Characteristic Peak Sparse Approximation), since its kernel is to reconstruct the interesting band of the spectrum according to sparse approximation. Sparse approximation is a powerful tool in signal processing field. To our knowledge, sparse approximation is widely used in 2-D image processing [16,17] and audio analysis for representation [18], however, few people use it to deal Raman spectral data.

Three algorithms were used in the CPSA: Sparsity-constrained Batch-OMP, least square and Butterworth low-pass filtering, of which the Batch-OMP is very particularly suitable for processing the spectral data set since it can speed up the sparse approximations processes of large amounts of data using the same dictionary. A brief outline of sparse approximation was first been described, and then we elaborately introduced the Sparsity-constrained Batch-OMP algorithm especially the adaptive dictionary construction method. In the next, we described the least square and Butterworth low pass filter. Finally, experiments and results were presented.

## 2. Method

In this section, we divided the CPSA into three steps: Sparse approximation, least square, and post-processing by Butterworth low-pass filter. Sparse approximation is the first and foremost step which plays a role of eliminating the noise of Raman signals. Least square is the second step which aims at enhancing the useful Raman signals obtained by the last step. The final step is post-processing, which is performed by Butterworth low pass filter and used to improve the image quality.

### 2.1. Sparse approximation

Consider a linear system of equations:

$$x = D\cdot\gamma \tag{1}$$

where $x \in R^m$, $y \in R^n$, $D \in R^{m\times n}$, $D$ is a predesigned matrix, called as dictionary. $D$ itself is over-complete, and the vectors ($\in R^m$) in $D$ called as atoms. The sparse approximation of $x$ is a linear combination of a few atoms of the given dictionary. The problem, in a nutshell, is to estimate the signal $\gamma$ subject to the constraint that it is sparse. Sparse implies that only a few coefficients in the vector $\gamma$ are non-zero and most of it are zero.

A tested Raman spectrum, in general, is the combination of pure Raman signal, background and random noise, which can be described as follow [1]:

$$x = p + b + r \tag{2}$$

where $x \in R^m$ is the tested Raman signal, $p \in R^m$ is the pure Raman signal, $b \in R^m$ is the background, $r \in R^m$ is the random noise. In fact, we eliminated the background simply by one-order polynomial approach before the sparse approximation, so that the signal $x$ can be described as:

$$x = p + r \tag{3}$$

Usually we consider that the random noise is independent of the pure Raman signal. It is expected that the pure Raman signal could be expressed and the random noise could be restricted through sparse algorithms if the dictionary is designed properly.

#### 2.1.1. Sparsity-constrained Batch-OMP

There are several algorithms [19–21] have been developed for solving sparse approximation problems, such as OMP(Orthogonal Matching Pursuit), K-SVD etc. In this paper, a Sparsity-constrained Batch-OMP algorithm was utilized to solve the sparse approximation problem. Batch-OMP is an extension of OMP, which can speed up the sparse approximation process of large amounts of data using the same dictionary through pre-computed some parameters [21].

The Sparsity-constrained Batch-OMP aims to estimate the signal $\gamma$ given by [19–21]:

$$\hat{\gamma} = \underset{\gamma}{Arg\ \min}\|x - D\gamma\|_2^2 \quad subject\ \ to \quad \|\gamma\|_0 \le K \tag{4}$$

Table 1 depicts the detail step of the Sparsity-constrained Batch-OMP algorithm.

In the Table 1, where $I$ is the collection of the index of the selected atoms in $D$, $G_I$ is the sub matrix of $G$ contains the columns indexed by $I$, $G_{I,I}$ refers to the sub matrix of $G$ contains the rows and columns indexed by $I$. $L$ is the Cholesky factorization of $G_{I,I}$. The output $\gamma$ is the sparse approximation of $x$ based on the dictionary $D$. In that case of $x$ is the Raman signal given by Eq. (3), that the pure Raman signal can be approximately described as:

$$p \approx D\cdot\gamma \tag{5}$$

**Table 1**
Sparsity-constrained Batch-OMP algorithm.

| **Sparsity constrained Batch Orthogonal Matching Pursuit** |
|---|
| **1. Input:** $\alpha^0 = D^T x$, $G = D^T D$, $K$ **(target sparsity)** |
| **2. Output:** $\gamma$, **approximant** $\hat{x} \approx D\gamma$ |
| **3. Init: set** $I := ( )$, $L := [1]$, $\gamma := 0$, $\alpha := \alpha^0$, $n := 1$ |
| **4. While** $(n \le K)$ **do** |
| **5.** $\hat{k} := \underset{k}{Arg\ \max}\{|\alpha_k|\}$ |
| **6. if** $(n > 1)$ |
| **7. solve for** $L$, **given by** $LL^T = G_{I,I}$ |
| **8. end if** |
| **9.** $I := (I, \hat{k})$ |
| **10. solve for** $\gamma_I$, **given by** $LL^T\gamma_I = \alpha_I{}^0$ |
| **11.** $\alpha := \alpha^0 - G_I\gamma_I$ |
| **12.** $n := n + 1$ |
| **13: End while** |