Chemometrics and Intelligent Laboratory Systems 128 (2013) 1–8

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

A real-time hyper-accuracy integrative approach to peak identification using lifting-based wavelet and Gaussian model for field mobile mass spectrometer

Cuiping Li [a,b,*], Jiuqiang Han [a], Qibin Huang [b], Ning Mu [b], Baoqiang Li [b], Bingqing Cao [b]

[a] Ministry of Education Key Lab for Intelligent Networks and Network Security, Xi'an Jiaotong University, XianYang west road 28# Xi'an PO 710049, China
[b] Beijing Institute of Pharmaceutical Chemistry, PO BOX 1044-403, Beijing PO 102205, China

ARTICLE INFO

Article history:
Received 7 February 2013
Received in revised form 21 June 2013
Accepted 22 June 2013
Available online 4 July 2013

Keywords:
Peak identification
De-noising
Lifting-based wavelet transform
Gaussian model
Mobile mass spectrometer
Environment pollution monitoring

ABSTRACT

Field mobile mass spectrometer is pivotal apparatus for real-time qualitative and quantitative analyses of chemical substances *in situ* environment pollution detection. To solve spectrum peak signal interfered by complicated noise, and to recognize irregular peak shape as well as quick monitoring, a real-time denoising and hyper-accuracy peak identification integrative approach for field mobile mass spectrometer using lifting-based wavelet transform (LWT) and Gaussian model has been developed. First, LWT was applied to eliminate the noise and to search for mass peak parameters in raw spectral peak data. Then, fitting the irregular mass peaks with Gaussian multi-peaks, a regular spectrum signal was obtained for further processing. Both of synthetic and apparatus experiment results show that LWT is a fast and effective denoising and peak identification method and retained the original peak features. The denoising effect (SNR/RMSE) by LWT was superior to Savitzky–Golay method used widely by experimental mass spectrometer, and the processing time was shortened obviously. Moreover integrated with Gaussian fitting algorithm, the peak parameters (the peak area A, centroid c, and half peak's width w) had been optimized. As the result, qualitative and quantitative accuracies of FMMS increased consequently. In addition, the approach achieved data compression.

1. Introduction

There is an increasing interest in the mass spectrometry that has field mobility including vehicle and portability [1]. That is so called field mobile mass spectrometry (FMMS) which combined with their direct inlet sample on-line, rapid analysis and high mobility. FMMS is an ideal equipment for various detection applications *in situ* that includes environmental and industry process monitoring, chemical emergency response applications, chemical or/and biological agents detections [2].

The conversion of the "raw" ion count data acquired by the machine into peak lists or mass spectra for further processing is usually called *peak picking* [3] or peak identification. The stages of peak identification usually include peak recognition, peak position and intensity calculation. Peak recognition routines carry to discriminate between true peaks and noise [4].

Unlike experimental mass spectrometry (MS), spectra peak signal of FMMS working at field complicated environment is often complicated by serious noise, irregular peak shape, and baseline drift which may be due to interfering physical or chemical processes, instrumental instability and temperature fluctuation. As the measured signal is close to the detection limit or low-abundant component peaks, some weak peak information is usually drowned in the complicated background. Irregular peak shape means the overlap, bifurcating and asymmetrical peaks (tailing peak or leading peak). Overlap peaks occur if the beam current did not drop below the threshold, particularly with instruments of lower resolution. These phenomena often cause difficulties with further processing, and either false negative or positive identifications of sample components owing to false signal. In addition, it can reduce mass accuracy due to shifting centroids of peaks. Denoising and regulating the shape of the MS are obviously important to increase the measurement accuracy.

In order to improve the S/N ratio by reducing noise, some peak identification methods were employed, such as threshold and a preliminary digital filtering or smoothing routine [4]. Several filtering methods, such as Fourier filtering method [4], Savitzky–Golay smoothing method [5], and Kalman filtering method [6] had been developed. An improved smoother based on penalized least squares was much better than the SG method in smoothing results, speed and memory usage when sparse linear algebra was used [7]. In addition, the median filter technique was well used for broad features with narrow noise spikes. If the peak shape function was well-known, a nonlinear least-squares multivariate fitting or matched filtration was used for denoising and peak picking [8,9]. Several papers introduced approaches for denoising and baseline subtraction based on wavelet translation in LC–MS and other experimental

* Corresponding author.
E-mail addresses: cuipingli86@126.com, lcplcp86@163.com (C. Li).

0169-7439/$ – see front matter.
http://dx.doi.org/10.1016/j.chemolab.2013.06.010

spectra [3,8,10,11], and proved that wavelet denoising algorithms have been confirmed superior to the Savitzky–Golay smoothing algorithm and the Fourier transform filtering algorithm. In addition, wavelet translation added advantage of compressing data in literature [8].

However, those effective algorithms for experimental MS are complicated, time-consuming, post-processing (off-line) and unsuitable for FMMS which should simultaneously resolve the problems of serious noise and detect irregular peaks shape. Meanwhile, analysis *in situ* is required to preserve chemical information in the limited memory with less time, particularly in the situation of applications, such as chemical agent detection and toxic pollutant measurement [1]. The chemical emergency response has demonstrated the need for real-time detection, high specificity and high sensitivity for low levels of chemical agents in air [12]. Increasing demands for the field analyst are to obtain the results as quickly as possible, so on-line and real-time monitoring is important for FMMS.

In conclusion, *in situ* accurate analysis of low concentration chemical compounds has become technological bottlenecks for FMMS. The algorithm should be powerful for denoising and peak detecting, and run in real time. That means the time spending on processing data should never exceed the time of acquiring it.

In this paper, we proposed an integrated algorithm that addresses the above mentioned goals. We advanced the real-time algorithm of denoising and peak detection for FMMS *in situ* detection. The algorithm sequentially utilized (1) lifting wavelet translation (LWT) to remove noise and search for irregular peak parameters (maximum and its position, left and right endpoints, and half peak's width) quickly and accurately, and (2) Gaussian fitting to modify irregular peak shape and to improve mass accuracy, at the same time, to achieve data compression. Using synthetic and experimental data obtained by self-design vehicle quadrupole mass spectrometry, the integrated algorithm was compared with that of algorithm used in experimental MS.

2. Theory and model

2.1. Fast lifted wavelet transform

Unlike a Fourier transform, the wavelet transform (WT) has dual localization both in scale (frequency) and in position (time) [13]. It is particularly suited for the processing of measured spectrum signals on different scales. Using the WT, the signals are decomposed into different frequency ranges or length scales that can be regarded independently of each other. Apparently, considering the signal at the correct scale, estimate of the typical peak width effectively suppresses both baseline and noise, keeping only the contribution due to the analytical signal. Nevertheless, WT is complicated, time-consuming, we had to find a simple and fast algorithm. The lifting scheme proposed by Sweldens [14] for both designing fast wavelets and performing the discrete wavelet transform is more delicate and facilitative. Lifting scheme can adaptively design wavelets by custom and speed-up the wavelet transform [15].

Take an initial set of biorthogonal scaling functions and wavelets $\left\{\varphi, \tilde{\varphi}^0, \psi^0, \tilde{\psi}^0\right\}$. Then a new set $\left\{\varphi, \tilde{\varphi}, \psi, \tilde{\psi}\right\}$, which is formally biorthogonal, can be found as [15]:

$$\psi(x) = \psi^0(x) - \sum_k s_k \varphi(x-k) \tag{1}$$

$$\tilde{\varphi}(x) = 2\sum_k \tilde{h}_k^0 \tilde{\varphi}(2x-k) + \sum_k s_k \tilde{\psi}(x-k) \tag{2}$$

$$\tilde{\psi}(x) = 2\sum_k \tilde{g}_k \tilde{\varphi}(2x-k). \tag{3}$$

Where the coefficients s_k can be freely chosen by custom. We can start from a simple or trivial set of biorthogonal functions and use formula (1) to choose s so that the wavelet after lifting has some desirable properties.

From formula (1), (2) and (3), evidently, an initial set of quadruplet filters $\left\{h, \tilde{h}^0, g^0, \tilde{g}\right\}$ of wavelet can be lifted into $\left\{h, \tilde{h}, g, \tilde{g}\right\}$. Coming from Mallat's fast wavelet translation, instead of explicitly constructing the filter $\tilde{h}$ and g, it can reduce the number of operations as compared to the standard algorithm that works with h, $\tilde{h}^0$, g^0, $\tilde{g}$ and s. The basic idea of the fast lifting-based wavelet transforms (LWT) is to first perform a classical subband filter with simple filters and later "lifting" the lower subband with the help of the higher subband, which named as *primal lifting*. In the case of *dual lifting* the higher subband would be lifted with the help of the lower one [14]. A block scheme is depicted in Fig. 1.

Forward transform:

—Stage I (splitting: calculate the unlifted wavelet coefficients c and d using the sample filters $\tilde{h}^0$ and $\tilde{g}$):

$$c_{j,l} := \sqrt{2}\sum_k \tilde{h}_{k-2l}^0 c_{j+1,k} \tag{4}$$

$$d_{j,l} := \sqrt{2}\sum_k \tilde{g}_{k-2l} c_{j+1,k}. \tag{5}$$

—Stage II (prediction or update: calculate the lifted wavelet coefficients):

$$c_{j,l} := c_{j,l} + \sum_k s_{l-k} d_{j,k}. \tag{6}$$

Inverse transform:

—Stage I:

$$c_{j,l} := c_{j,l} - \sum_k s_{l-k} d_{j,k}. \tag{7}$$

—Stage II (merging):

$$c_{j+1,k} := \sqrt{2}\sum_l h_{k-2l} c_{j,l} + \sqrt{2}\sum_l \tilde{g}_{k-2l}^0 d_{j,l} \tag{8}$$

where $c = \{c_0, c_1, \ldots, c_l\}$ is the low frequency component (outline), and $d = \{d_0, d_1, \ldots, d_l\}$ is high-frequency component (detail). From formula (4)–(8), the stage I of inverse transform is simply undoing stage II of the transform, only changing the positive sign to negative sign. So lifting speed up the Mallat algorithm.

2.2. Model of signal

The measured spectrum of MS is of an inherently multiscale nature, and is described as following [16]:

$$f(x) = Ns(x) + b(x) + n(x) \tag{9}$$

where $f(x)$ is the observed signal, x is time sequences representing as m/z, $b(x)$ is low-frequency baseline or background term, often described as a slowly varying trend under the spectra, $s(x)$ is the signal produced by the component fragment peaks present in the sample, where $s(x)$ occupies a frequency range in between noise and baseline, N is a normalization factor, and $n(x)$ is a general term for the high-frequency noise term, which encompasses both the chemical background and the stochastic noise.

In this paper, baseline drift is described as a linear function according with our FMMS:

$$b(x) = kx + b_0 \tag{10}$$

where $b(x)$ is the baseline term, x is time sequences representing as m/z, k is the slope of linear function, and b_0 is baseline drift constant.