NONLINEAR LEAST SQUARES REGRESSION ANALYSIS AND DIAGNOSIS FOR OPTICAL AND LASER MEASUREMENT SYSTEMS

[1]ZHENGMAO YE, [2]HABIB MOHAMADIAN

[1]Senior Member, IEEE College of Engineering, Southern University, Baton Rouge, LA70813, USA
[2]Fellow, ASME College of Engineering, Southern University, Baton Rouge, LA70813, USA
E-mail: [1]zhengmao_ye@subr.edu, [2]habib_mohamadian@subr.edu

Abstract— Laser Doppler and Raman spectroscopy are widely adopted for biomedical diagnosis. Both Doppler ultrasound and Raman spectra are used to display normal and abnormal signature waveforms that are unique in collected samples, in order to recognize normal and abnormal regions in the spectral display respectively. Visualization of quantitative Doppler ultrasound data and Raman spectra provides a straightforward means to distinguish between normal and abnormal samples, however, mixing artifacts due to motion, instrumentation and background are in fact inevitable. The fluorescence artifact is the dominating factor whose formation mechanism varies. In general, fluorescence waveforms could be formulated as typical Gaussian, Sigmoidal or Lorentzian distribution, while single spectrum is seldom enough to represent the fluorescence waveforms completely to further extract intrinsic spectra with unique signatures. In this case, combination of Gaussian, Sigmoidal and Lorentzian waveforms could be the solution to model the fluorescence artifact. To identify a set of modeling parameters, a simple but powerful nonlinear least squares regression technique is introduced in this article. Case studies on both normal and abnormal sample data are conducted.

Index Terms—Nonlinear Least Squares Regression, Raman Spectroscopy, Doppler Ultrasound

I. INTRODUCTION

For laser excited Raman spectroscopic and Doppler ultrasound techniques, artifacts always occur in various forms of diverse frequency, duration, spike, shape and size. Fluorescence could be the major artifact that has strong impact on the observed spectra. The fluorescence spectroscopic function is usually described as a Gaussian peak, sigmoidal function, Lorentzian peak, or some other nonlinear waveform functions, even as a high order linear polynomial. It is challenging to eliminate effects of artifacts. A linear approach is applied to maximize signal collection efficiency. A near infrared Raman spectroscopy system is employed to detect the near infrared bands hidden among large intensity background spectra so as to examine biological components of chemical concentrations and human tissues. Excited by intense laser light, the stray light influence on spectroscopic systems could be minimized [1]. Noises are unavoidable in any engineering data acquisition system. Noise reduction is necessary under many practical conditions when measurement and observation are severely contaminated by multiple sources of noise. Solid noise reduction schemes should be applied to eliminate the chance of data corruption in the engineering problem. A comparative study is made on a basis of several denoising methods, such as the adaptive recursive least square filter, multi-layer neural network training using backpropagation algorithms and discrete wavelet denoising. The application of laser Doppler noise filtering in the frequency domain is conducted successfully [2]. Signal-to-Noise ratio (S/N) is the criterion that compares the level of useful signal to that of noise. Noise reduction is the process of extracting noise from intrinsic signals so as to enhance signal-to-noise ratio. The objective is to limit noise to an acceptable level. The actual mechanism of noises varies significantly (e.g., thermal noise, shot noise, impulse noise, cross-talk noise and cosmic noise). To make it simple, noise can be grouped into slowly varying noise and rapidly varying noise. For slowly varying noise, simple linear filtering approach can be applied for noise cancellation, such as the linear least squares filter. For rapidly varying noise, the attribute can be highly nonlinear, uncertain and unpredictable, so nonlinear filtering approaches should be implemented, such as fuzzy filtering, neural network training, wavelet denoising, and so on. In a systematic characterization scheme, a linear least squares filter is used to reduce slowly varying noise and a nonlinear fuzzy filter is also applied to reduce rapidly varying noise. The nonlinear filtering based on fuzzy logic is applied to sufficiently improve the signal to noise ratio [3]. For biomedical sample characterization across diversified testing conditions, it is also reasonable for intrinsic spectra extracted to go through the sequence of normalization. Normalization could be processed by either auto-scaling or standard normal variate approach [4]. Artifact sources can be classified into external and internal types. Internal artifacts arise from the body activities owing to movements or sudden potential changes. External artifacts result from coupling with unwanted external interferences. Majority of artifacts can be characterized into 4 types based on the frequency characteristics, sharpness of edge, duration, shape and size [5]. PCA (Principal component Analysis) is effective on linear data classification. It is successfully applied for near-infrared spectral analysis on biomedical samples. PCA represents a quantitative approach that produces new variable sets

from the high dimensional correlated datasets to reduce redundancy. The components of less contribution are subject to cancellation. It serves as a feasible sample characterization method [6]. The artificial intelligence approach is carried out in order to differentiate numerous biomedical samples using Raman spectroscopy. The aim is to secure exact medical diagnosis and decision making. It covers procedures of noise filtering in linear and nonlinear senses, fluorescence identification, optimization, normalization, multivariate statistical analysis, clustering plus decision making. Various schemes of modeling, control and system identification are involved, such as the fuzzy control, genetic algorithms, PCA and system identification. This systematic intelligent control approach is able to classify different sets of biomedical samples [7]. For multivariate signal processing, nonlinear discrete wavelet transform denoising is proposed to extract intrinsic spectra from the noise contaminated original spectra. Fractional wavelet transform is effective in extracting major analytical information from spectral bands. It could also be combined with principal component regression and the partial least squares calibration is conducted for quantitative analysis. The processing of absorption spectra is by means of fractional wavelet transform [8-9]. Noise is mostly random but the artifacts are more deterministic. Seeing the computational complexity involved in some advanced linear and nonlinear approaches, such as fuzzy logic, neural network, and wavelet transform, simple least squares regression analysis [10] is introduced to extract unique signatures from both normal and abnormal biomedical samples for accurate decision making.

II. NONLINEAR LEAST SQUARES REGRESSION

For weak signals like Raman spectra, the fluorescence intensity has a strong impact on the observed waveforms. The artifacts arisen from the fluorescence takes a leading role. Even though the formation mechanism of artifacts varies a lot, such as low frequency versus high frequency, relatively narrow versus relatively broad band, the fluorescence artifact is estimated as combination of typical spectral waveforms with the constant horizontal baseline.

A. Gaussian Waveform Function

The Gaussian waveform function is simplified as (1).

$$f_1(x) = A_1 + B_1 * \exp\left[\frac{-(x-\mu_1)^2}{w_1^2}\right] \quad (1)$$

where A_1 is the amplitude constant offset, μ_1 is the center of Gaussian function, w_1 is corresponding to the width and B_1 is corresponding to the area of the waveform.

B. Sigmoidal Waveform Function

The logistic sigmoid function is defined as (2).

$$f_2(x) = A_2 + \frac{B_2}{1+\exp\left[-w_2(x-\mu_2)\right]} \quad (2)$$

where A_2 is the amplitude constant offset, μ_2 is the center of sigmoidal function, w_2 is corresponding to width adjustment ($w_2>0$) and B_2 is corresponding to height adjustment.

C. Lorentzian Waveform Function

The Lorentzian waveform function is simplified as (3).

$$f_3(x) = A_3 + \frac{B_3}{(x-\mu_3)^2 + w_3^2} \quad (3)$$

where A_3 is the amplitude constant offset, μ_3 is the center of Lorentzian function, w_3 is corresponding to the width and B_3 is corresponding to the area of the waveform.

D. Fluorescence Artifact Model

Each of three typical waveform functions above is regarded as a major component of the fluorescence artifact model. The combination model feature is formulated as (4).

$$f(x) = \theta + \alpha * e^{\frac{-(x-\mu_1)^2}{w_1^2}} + \frac{\beta}{1+e^{-w_2(x-\mu_2)}} + \frac{\gamma}{(x-\mu_3)^2 + w_3^2} \quad (4)$$

where θ represents a horizontal constant baseline. To optimize a set of parameters (w_1, w_2, w_3, μ_1, μ_2, μ_3, α, β, γ, θ), the performance index should be defined accordingly.

E. Nonlinear Least Squares Regression

Assume that m sampling points are collected from tests. Then an iterative nonlinear least squares regression can be made in the optimization problem. Performance index is defined as:

$$F = \frac{1}{2}\sum_{i=1}^{m} r^2(x_i) = \frac{1}{2}\sum_{i=1}^{m} [f(x_i) - y_i]^2 \quad (5)$$

where $r(x_i) = f(x_i) - y_i$ ($i = 1, 2, \ldots, m$). Two vectors in x-coordinate and y-coordinate with m (e.g. m=1000) sampling points also give rise to the vector form $r(x) = f(x) - y(x)$.

For a single sampling point pair (x_i, y_i), the residual function is also the function of 10 parameters. If one waveform function is selected alone, the residual function is an exclusive function of 4 parameters. If two waveform functions are selected, the residual function is a function of 7 parameters. Now we have:

$$r(x_i) = f(x_i) - y_i = f(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \alpha, \beta, \gamma, \theta) \quad (6)$$

For matter of simplicity, it is rewritten as:

$$r(x_i) = f(w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}) \quad (7)$$

With all sampling points being collected, it turns out to the parametric optimization problem covering a parameter set of $w=[w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}]$. This set needs to be optimized so as to minimize the performance index in (5). That is:

$$\text{Minimize } F(w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}) = r(w)^T r(w)/2 \to 0 \quad (8)$$

The number of parametric variables is defined (n=10) in (8). The nonlinear least squares parameter optimization problem is formulated. The gradient

term $\nabla F(w)$ of the performance index is derived by the chain rule:

$$\nabla F(w) = \nabla r(w)^T r(w)/2 = J(w)^T r(w) \quad (9)$$

where $J(w) = \nabla r(w)^T$ is the Jacobian of $r(w)$ as in (10).

$$J(w) = \begin{bmatrix} \frac{\partial r_1(w)}{\partial w_1} & \frac{\partial r_1(w)}{\partial w_2} & \cdots & \frac{\partial r_1(w)}{\partial w_n} \\ \frac{\partial r_2(w)}{\partial w_1} & \frac{\partial r_2(w)}{\partial w_2} & \cdots & \frac{\partial r_2(w)}{\partial w_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial r_m(w)}{\partial w_1} & \frac{\partial r_m(w)}{\partial w_2} & \cdots & \frac{\partial r_m(w)}{\partial w_n} \end{bmatrix} \quad (10)$$

The Hessian is computed by the chain rule again.

$$\nabla F^2(w) = \nabla r(w) \nabla r(w)^T + \sum_{i=1}^{m} r_i(w) \nabla^2 r_i(w) \quad (11)$$

The Hessian of the nonlinear least squares performance index includes one term with 2 first order derivatives and the other term with 1 second-order derivative. At each iteration k, the model is linearized by the first-order Taylor series expansion about w_k, while higher-order terms are neglected. The normal equation in the matrix form is shown as (12) whose optimal solution in the least squares sense is shown as (13).

$$[J(w)^T J(w)] \Delta w = J(w)^T \Delta y \quad (12)$$

$$\Delta w = [J(w)^T J(w)]^{-1} J(w)^T \Delta y \quad (13)$$

If matrix $J(w)^T J(w)$ is nonsingular, then parametric vector w can be determined directly.

III. CASE STUDY A: COMPARISONS OF NORMAL AND ABNORMAL RAMAN SPECTRA

Raman instrumentation consists of argon laser, a set of optics, a Raman spectrometer, a detector and a computer. It is designed to monitor and observe low frequency vibrational modes. The laser light scattered from the sample surface includes Raman scattered light and Rayleigh scattered light. Raman scattered light is based on the monochromatic laser source at the near infrared, near ultraviolet or visible range, whose intensity is relatively weak compared with intense Rayleigh scattered light. Thus holographic notch filter is used to reject high degree intense Rayleigh scattered light. Laser collection optics system is used to focus laser spots onto the sample and spectrometer entrance slit. Via Raman energy shift information between incident light and scattered light, natural frequency of samples can be determined. The Raman scattered light is measured by a multi-channel CCD detection system with a cooled detector head and a controller. The computer program is to adjust all instrumental parameters to produce the real-time optimized parameter value such as slit width, scan step size, acquisition time and pixel parameters. Via A/D conversion at the detector interface unit, digital spectra are generated. Then source Raman spectra are captured by the data acquisition system. The slowly varying smooth fluorescence artifact has strong impact on spectrum observation that is a dominating factor which inhibits the intrinsic spectrum to be perfectly extracted. The schematic diagram of Raman spectroscopy is in Fig. 1

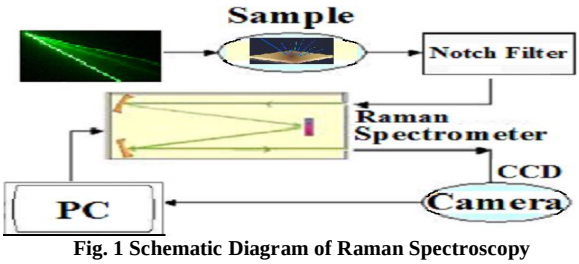

Fig. 1 Schematic Diagram of Raman Spectroscopy

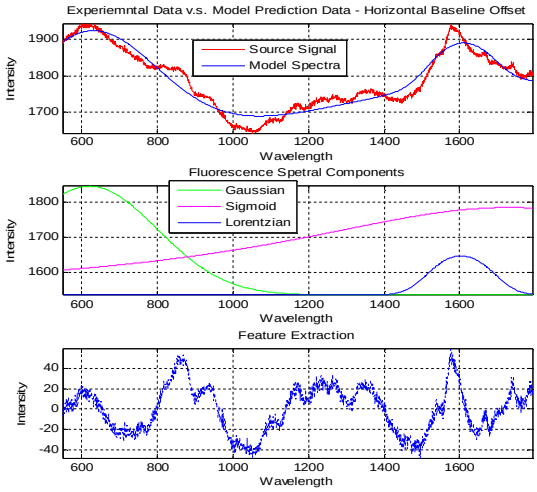

Fig. 2. Raman Spectra of Normal Lung Sample

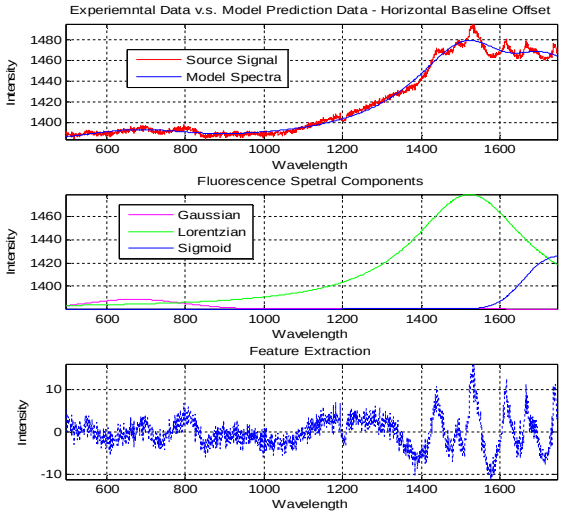

Fig. 3. Raman Spectra of Injected Lung Sample

In Figs. 2-3, Raman spectra of normal and abnormal mice lung samples are shown, whose spectroscopic function is estimated as combination of Gaussian, Sigmoidal and Lorentzian wave forms. Similar results are also shown in Figs.4-5, where combination of exclusive Gaussian and Lorentzian waveforms are applied. The laser light interacts with molecular vibrations or phonons, resulting in the energy shift of

photons. The shift in energy provides information on system vibrational modes. This frequency shift is associated with the vibrational modes that is corresponding to peaks of the waveform. In each of Figs. 2-5, the intrinsic signal is extracted sharply after the removal of spectra arisen from major fluorescence artifacts. From visual appeals, it is observed that the frequency shift information (relevant to peaks) occurs much more frequently in abnormal samples than normal samples. Meanwhile, more precise spectra extraction is made when the estimation model covers all three popular spectra than only two typical Gaussian and Lorentzian spectra. Similar quantitative data will also be disussed in next session. The proposed approach can easily be extended to decision making among normal, benign and abnormal samples.

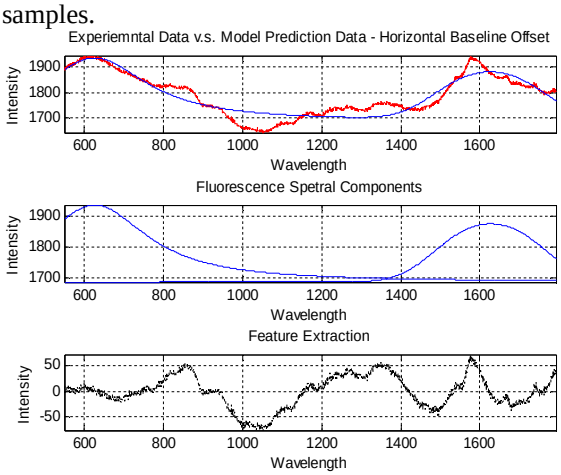

Fig. 4. Raman Spectra of Normal Lung Sample (Gaussian + Lorentzian)

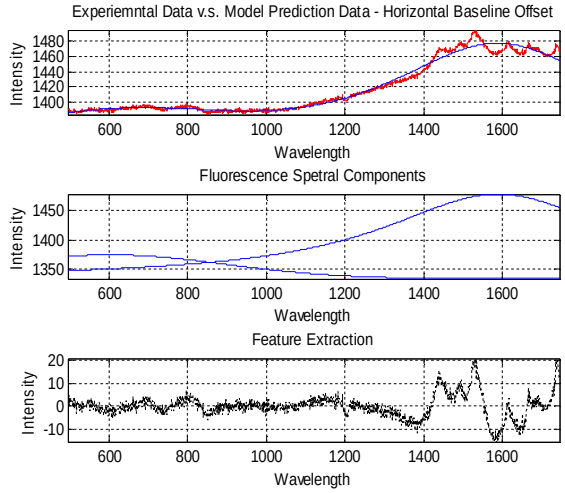

Fig. 5. Raman Spectra of Injected Lung Sample (Gaussian + Lorentzian)

VI . CASE STUDY B: COMPARISONS OF NORMAL AND ABNORMAL DOPPLER ULTRASOUND SPECTRA

In Laser Doppler scanning systems, the near infrared (NIR) laser source and Doppler ultrasound scans are applied together to conduct non-destructive and non-invasive tests. The blood flow is monitored and estimated by bouncing the high frequency ultrasound sound waves off circulating blood cells. The near infrared laser has a deep penetration depth which is less dependent on skin colors due to the long wavelength and low melanin absorption. For a small scale near infrared imaging, laser Doppler scanning is made pixel by pixel according to the row index number or column index number. The non-contact laser beam is projected directly onto skin and the scattered laser light is used for non-destructive biomedical diagnosis. Measurement of scattered light Intensity over the whole region provides pictorial representation that will outperform the spot measurement. Doppler shift can be captured in terms of the echoes from intersection of the ultrasound beam and blood flow. The Doppler shift of the scattered laser light is proportional to the average flow rate of blood cells, which will be detected by near infrared spectrometer. The real time measurement of the shift information quantifies differences in the blood flow patterns over time. Spectral Doppler is then used to display the unique normal and abnormal signatures. Comparision of waveforms between the normal skin region and burnt area can be made to estimate the burn depth of skin. Quantitative visualization data in context is collected from row index based ultrasound scans of the skin surface. The schematic diagram of laser Doppler ultrasound scanning is in Fig. 6.

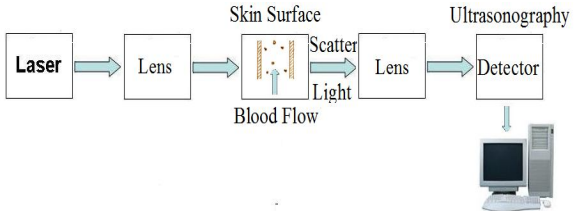

Fig. 6 Schematic Diagram of Laser Doppler Ultrasound Scan

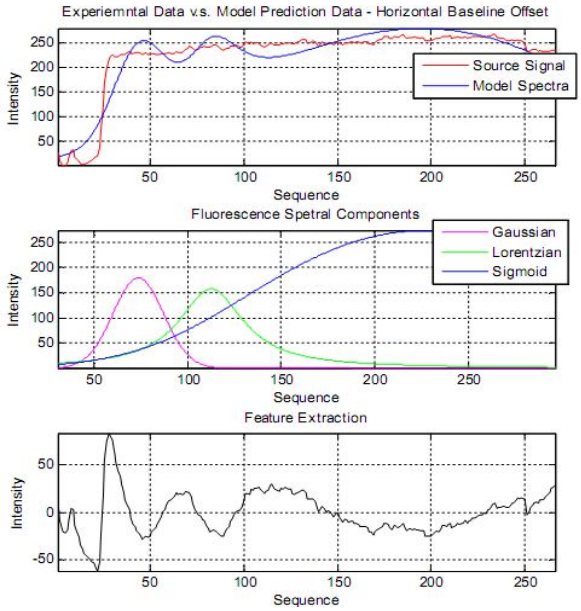

Fig. 7. Doppler Ultrasound Spectra from Surface Scan A

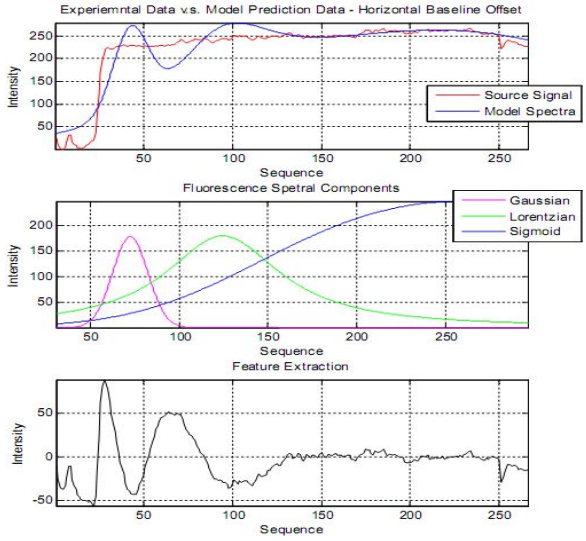

Fig. 8. Doppler Ultrasound Spectra from Surface Scan B

In Figs. 7-8, Doppler ultrasound scan spectra of normal and abnormal skin surface regions are shown, whose spectroscopic function is estimated as combination of Gaussian, Sigmoidal and Lorentzian waveforms. Intrinsic sigmals are extracted successfully, where the range of waveforms with considerable difference (at high sequence number) will tell the difference between normal and abnormal regions of the skin surface. A fraction of laser beam penetrates the skin and interacts with moving blood cells and static tissues. The penetration depth depends on wavelength, morphology and absorption. The Doppler effects lead to changes in the wavelength of scattered light, which is depended on speed and density of moving cells.

Experiments with 2-spectra models are also conducted where 10 parameters are substituted by 7 parameters instead. The combinations of Gaussian and sigmoidal waveforms, Gaussian and Lorentzian waveforms, and sigmoidal and Lorentzian waveforms, are applied respectively. All outcomes from 3-spectra model and 2-spectra models are satisfactory. In order to further compare the 3-spectra model with 2-spectra models, the metric of the relative root-mean-square error (RRMSE) is defined as (14). RRMSE results are shown in Table I.

$$RRMSE = \frac{\sqrt{\sum_{i=1}^{m}[f(x_i)-y_i]^2}}{\sqrt{m}\,(f_{max}+y_{max})/2} \qquad (14)$$

where m is the total number of sampling points; f_{max} and y_{max} stand for the largest intensity values from either the estimation model or the measurement data.

TABLE I. RELATIVE ROOT MEAN SQUARE ERROR OF DOPPLER DATA

RRMSE %	3-Spectra	Gaussian Lorentzian	Gaussian Sigmoidal	Sigmoidal Lorentzian
Scan A	8.31%	12.97%	10.67%	11.77%
Scan B	9.45%	12.96%	10.66%	11.76%

Table I is made based on the pixel by pixel scanning of 256 data points according to the row index number. It is clearly shown that the 3-spectra model provides more accurate estimation than those individual 2-spectra models being selected.

CONCLUSIONS

Biomedical spectra are always corrupted with diverse types of artifacts. The artifact overlaps with the intrinsic spectrum over a broad range of frequency band that also exhibits nonlinear attributes. Nonlinear rather than linear modeling should be applied for parametric identification. Since it is important to distinguish between normal and abnormal conditions from the biomedical spectral analysis, the nonlinear least squares regression has been applied to multiple case studies of biomedical sample differentiation. The complex fluorescence artifact spectrum has benn approximated as the combination of several typical Gaussian, sigmoidal and Lorentzian waveforms. It is subject to elimination from the collected source signals. Spectral extraction is thus made so that the intrinsic signal can be located to capture unique signatures based on generated waveforms. Distinctive intrinsic features are well captured for biomedical diagnosis.

REFERENCES

[1] J. Brennan, Y. Wang, R. Chandra, R. Dasari and M. Feld, “Near-Infrared Raman Spectrometer Systems for Human Tissues Studies”, Journal of Applied Spectroscopy, Vol. 51, No. 2, pp. 201-207, 1997

[2] Z. Ye and H. Mohamadian, "Comparative Study of Noise Reduction Approaches with Engineering Applications in Time Domain and Frequency Domain", Proceedings of the 2012 IEEE International Conference on Computer Science and Engineering (ICCSE 2012), pp. 1190-1194, July 14-17, 2012, Melbourne, Australia

[3] Z. Ye and G. Auner, “Linear Filtering and Nonlinear Fuzzy Logic Filtering for Sample Identification with Raman Spectroscopy”, Proceedings of the 2003 IEEE International Conference on Systems, Man and Cybernetics (SMC 2003), pp. 4619-24, Oct 5-8, 2003, Washington, DC, USA

[4] R.Barnes, M.Dhanoa, S.Lister, "Standard Normal Variate Transforma-tion and De-trending of Near-Infrared Diffuse Reflectance Spectra", Applied Spectroscopy, pp.772-777, Volume 43, Issue 5, July 1989

[5] M. Islam, N. Tuan, Y. Zhou, and Z. Yang, “Analysis and Processing of In-Vivo Neural Signal for Artifact Detection and Removal”, pp.437-442, 2012 5th International Conference on BioMedical Engineering and Informatics, Oct 16-18, 2012, Chongqing, China

[6] Z. Ye and G. Auner, “Principal Component Analysis for Biomedical Sample Identification”, Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics (SMC 2004), pp. 1348-1353, Oct. 10-13, 2004, Hague, Netherlands

[7] Z. Ye, “Artificial Intelligence Approach for Biomedical Sample Characterization Using Raman Spectroscopy”, IEEE Transactions on Automation Science and Engineering, Vol. 2, No. 1, pp. 67-73, Jan 2005

[8] Z. Ye and H. Mohamadian, "Spectral Feature Extraction and Characterization via Simple Linear and Nonlinear Technology Integration", Proceedings of the 2011 IEEE International Symposium on Computational Intelligence and

Informatics, pp. 121-125, November 21-22, 2011, Budapest, Hungary

[9] E. Dinc, F. Demirkaya and D. Baleanu, "New Approach for Simultaneous Spectral Analysis of a Complex Mixture Using the Fractional Wavelet Transform", Communications in Nonlinear Science and Numerical Simulation, vol. 15, no. 4, p 812-818, April 2010

[10] K. Astrom and B. Wittenmark, "Adaptive Control", 2nd Edition, Prentice Hall, 1995

★★★