

January 01, 2007

Denoising of Raman Spectroscopy Signals

Luis A. Quintero

University of Puerto Rico - Mayaguez

Shawn Hunt

University of Puerto Rico - Mayaguez

Max Diem

Northeastern University

Recommended Citation

Quintero, Luis A.; Hunt, Shawn; and Diem, Max, "Denoising of Raman Spectroscopy Signals" (2007). *Research Thrust R2 Publications*. Paper 2. <http://hdl.handle.net/2047/d10008949>

Denoising of Raman Spectroscopy Signals

LUIS A. QUINTERO¹, SHAWN HUNT² and MAX DIEM³

¹ luis.quintero@ece.uprm.edu, Electrical and Computer Engineering, University of Puerto Rico, Mayagüez, PR
² shawn@ece.uprm.edu, Electrical and Computer Engineering, University of Puerto Rico, Mayagüez, PR
³ m.diem@neu.edu, Department of Chemistry & Chemical Biology, Northeastern University, MA



ABSTRACT

The detection system (Charge-Coupled Device, CCD) in Raman spectroscopy measurements are affected by spurious signals and noise, mainly produced by cosmic rays, shot noise and thermal noise. Generally, due to the nature of the noise signals, the spectrum estimation is divided in two sequential stages. The first stage removes the impulsive noise caused by cosmic rays. The second stage attempts to remove the rest of the noise, it is assumed that the statistics of the noise follows a Poisson process. In this work, the algorithm for removing the impulsive noise is based on a system which uses both a median filter and classic pattern recognition techniques. For the second stage is considered the Wavelet transform like alternative to denoise the spectra and it is compared with the classical smoothing method of Savitzky-Golay. The implemented algorithms are tested with synthetic and real spectra; real spectra are from Raman Imaging of biological materials which were provided by the research group led by professor Max Diem at Northeastern University research group led by professor Max Diem at Northeastern University. The algorithms are useful for all software tools that analyze Raman spectroscopy data.

PROBLEM STATEMENT

Raman micro-spectroscopy is a powerful tool used in the estimation of chemical and physical properties of materials. An scheme of a Raman spectroscopy is showed in the Figure 1.

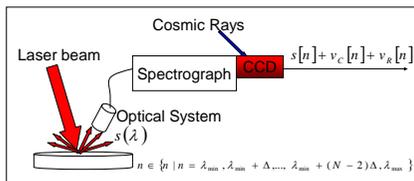


Figure 1. Raman micro-spectroscopy and signal model.

CCD sensors are used to detect the Raman Scattering and the measurements are affected by noise. The most important noise sources are[1]:

Noise spikes or cosmic: are produced by high energy radiation from local or extraterrestrial sources (cosmic rays). In this work, this noise is called *impulsive noise*.

Signal shot noise: It is a inherent noise related with counting based methods like CCD detectors. Shot noise is characterized by variance which is proportional to the mean value of the measurement. Generally, this noise is distributed according to the *Poisson probability density function*[2].

Detector dark noise: also known as thermal noise. Spontaneous generation of electrons in the detector is produced mainly by thermal generation. This noise is reduced by mean a cooling system in the detector.

Readout noise: It is related with the digitalization of the electrons, produced by the photons, in the CCD detector

Figure 1 shows the signal model used, where $s[n]$ is the discrete estimate of $s(\lambda)$, the Raman scattering signal. The *impulsive noise* is represented by $v_c[n]$ and the rest of the noise is summarized in a global *random noise* $v_r[n]$

IMPULSIVE NOISE FILTER

The *impulsive noise filter* is composed of two parts. Once detected the *cosmics*, the *spikes* are removed and replaced by the result of the *Missing Point Filter* with an added noise. The noise is added in order to have the same statistics for the replaced points, as the rest of the signal.

SAVITZKY-GOLAY FILTER

Savitzky-Golay (SG) filter[3] is a smoothing algorithm based in a least-squares polynomial fitting within a window. Other smoothing algorithms has been proposed to remove the random noise, but the SG filter is the most widely used in analytical chemistry. For this work, is used a *third-order* and a *nine-point* window.

SIGNAL PROCESSING SYSTEM

A signal processing system to estimate the signal $s[n]$ based on the assumed signal model is proposed. It consists of two sequential stages, the first one is an *impulsive noise filter* that detects and removes the *spikes (cosmics)* in the spectra. The second stage is a *denoising or smoothing filter* designed to make a good estimate of the signal $s[n]$ from the measured signal $\hat{x}[n]$. Two alternatives are evaluated, the classic smoothing algorithm of Savtzky-Golay and the Wavelets Denoising algorithm.

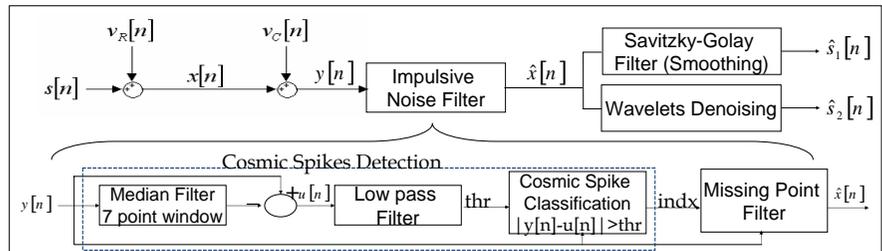


Figure 2. Preprocessing system proposed.

WAVELETS DENOISING

By mean of wavelets, a representation that discriminates the signal from the noise is found. Wavelet Shrinkage Estimators[4] are used to attenuate the noise $v_r[n]$, preserving the signal $s[n]$.

Four wavelets are selected to make the signal decomposition: symlet of eighth order (*sym8*), coiflets of fifth order (*coif5*), daubechies of fifth order (*db5*) and biorthogonal of sixth order for reconstruction and eighth order for decomposition (*bior6.8*). The analysis of the signal is done using *five levels of decomposition*.

Four thresholding methods has been considered: Stein's Unbiased Estimate of Risk (rigorous - *RSURE* - and heuristic - *HSURE*), MiniMax (*MM*) and universal threshold (*LITH*) [5].

RESULTS

To compare the Savitzky-Golay algorithm with the wavelets denoising, we have designed a synthetic sample characterized by *Lorentzian peaks*[6]. The performance of the estimation of $s[n]$ is measured by mean the mean squared error (MSE). The MSE for applying the Savitzky-Golay filter is: 2.8276 for SNR = 22 and 12.8401 for SNR = 5.

Table 1. MSE for Wavelet denoising, soft thresholding; SNR=22

Thr. Method	RSURE	HSURE	MM	UTH
sym8	1.6123	1.6867	2.2824	3.9485
coif5	1.8950	2.2637	2.8562	5.0310
db5	1.8286	1.8633	2.7080	5.0372
bior6.8	1.7627	1.7724	2.2623	3.7563

Table 2. MSE for Wavelet denoising, soft thresholding; SNR=5

Thr. Method	RSURE	HSURE	MM	UTH
sym8	5.4032	5.2695	7.1468	11.254
coif5	6.2363	8.3731	8.4113	14.145
db5	6.9922	7.9480	9.6533	16.394
bior6.8	5.6854	5.1811	6.1373	10.131

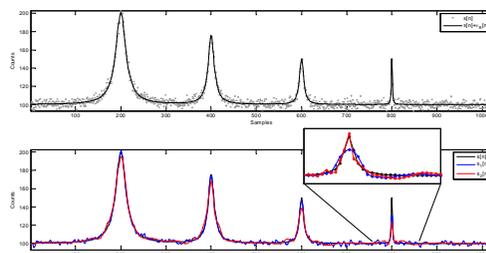


Figure 3. Simulated data and estimates of $s[n]$. Wavelets: sym8, RSURE

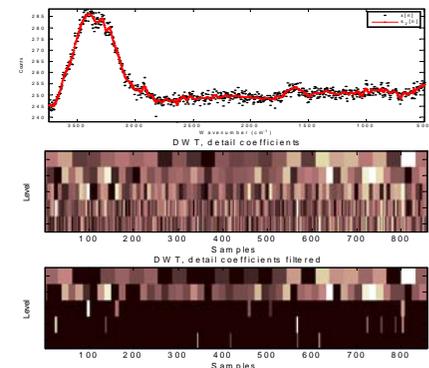
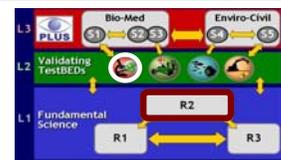


Figure 4. Real noisy data (black) and wavelet estimate in red (sym8, RSURE). Detail coefficients before and after the thresholding, below.

RELEVANCE TO CenSSIS STRATEGIC RESEARCH PLAN



The algorithms will help to improve the processing and the understanding (R2) of Raman spectrums used in the estimation of chemical and physical properties of biological materials (BioBED).

REFERENCES

- [1] R. L. McCREERY, *Chemical Analysis: Raman Spectroscopy for Chemical Analysis*, Vol 157, Wiley, New York, 2000.
- [2] S. E. BIALKOWSKI, *Data Analysis in the Shot Noise Limit. 1. Single Parameter Estimation with Poisson and Normal Probability Density Functions*, Analytical Chemistry, Vol. 61, No. 22, 1989. p. 2479-2483.
- [3] A. SAVITZKY, M. J. E. GOLAY, *Smoothing and Differentiation of Data by Simplified Least Squares Procedures*, Analytical Chemistry, Vol. 36, No. 8, July, 1964. p. 1627-1639.
- [4] S. MALLAT, *A Wavelet Tour of Signal Processing*, Academic Press, 1999. p. 434-486.
- [5] P. BESBEAS, I. FEIS, T. SAPATINAS, *A Comparative Simulation Study of Wavelet Shrinkage Estimators for Poisson Counts*, International Statistical Review, Vol. 72, No. 2, 2004. p. 209-237.
- [6] L. S. GREEK, et al. *SNR Enhancement and Deconvolution of Raman Spectra Using a Two-point Entropy Regularization Method*. Applied Spectroscopy, Vol. 49, No. 4, 1995. p. 425-431



This work was supported in part by Gordon-CenSSIS, the Bernard M. Gordon Center for Subsurface Sensing and Imaging Systems, under the Engineering Research Centers Program of the National Science Foundation (Award Number EEC-9986821).

