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, Mayaguez, PR
2 shawn@ece.uprm.edu, Electrical and Computer Engineering, University of Puerto Rico, Mayaguez, 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. 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.](image)

CCD sensors are used to detect the Raman Scattering and the measurements are affected by noise. The most important noise sources are:[3]

- Noise spikes or cosmics 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(t), the Raman scattering signal. The impulsive noise is represented by v[n] and the rest of the noise is summarized in a global random noise v[n].

IMPULSIVE NOISE FILTER

The impulsive noise filter is composed of two parts. Once detected the cosmos, 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.

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[n], preserving the signal s[n]. Four wavelets are selected to make the signal decomposition: sym8, db5, coif5, bior6.8. The analysis of the signal is done using five levels of decomposition. Four thresholding methods have been considered: Stein’s Unbiased Estimate of Risk (rigorous – RSURE - and heuristic - HSURE), MiniMax (MM) and universal threshold (UHT).[5]

RESULTS

To compare the Savitzky-Golay algorithm with the wavelets denoising, we have designed a synthetic sample characterized by Lorentzian peak[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 12.8611 for SNR = 5.

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

<table>
<thead>
<tr>
<th>Thr. Method</th>
<th>RSURE</th>
<th>HSURE</th>
<th>MM</th>
<th>UHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>sym8</td>
<td>1.6123</td>
<td>1.6867</td>
<td>2.2824</td>
<td>3.9485</td>
</tr>
<tr>
<td>db5</td>
<td>1.8950</td>
<td>2.2637</td>
<td>2.8562</td>
<td>5.0310</td>
</tr>
<tr>
<td>coif5</td>
<td>1.8286</td>
<td>1.8033</td>
<td>2.7080</td>
<td>5.0372</td>
</tr>
<tr>
<td>bior6.8</td>
<td>1.7827</td>
<td>1.7724</td>
<td>2.8283</td>
<td>3.7563</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Thr. Method</th>
<th>RSURE</th>
<th>HSURE</th>
<th>MM</th>
<th>UHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>sym8</td>
<td>5.0432</td>
<td>5.2665</td>
<td>7.1468</td>
<td>11.254</td>
</tr>
<tr>
<td>db5</td>
<td>6.2363</td>
<td>8.3731</td>
<td>8.4113</td>
<td>14.145</td>
</tr>
<tr>
<td>coif5</td>
<td>6.9922</td>
<td>7.9480</td>
<td>9.6533</td>
<td>16.394</td>
</tr>
<tr>
<td>bior6.8</td>
<td>5.6854</td>
<td>5.1811</td>
<td>6.1373</td>
<td>10.131</td>
</tr>
</tbody>
</table>

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

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 s[n]. Two alternatives are evaluated, the classic smoothing algorithm of Savitzky-Golay and the Wavelets Denoising algorithm.

Figure 2. Preprocessing system proposed.

![Figure 2. Preprocessing system proposed.](image)

RELEVANCE TO CenSSIS STRATEGIC RESEARCH PLAN

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

REFERENCES


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 ECC-9908212).