IJCTA, 9(22), 2016, pp. 325-334

Raman Spectra De–Noising Using Wavelet Packet Thresholding

Nitendra Kumar* A. H. Siddiqi and Khursheed Alam*

***Abstract* :** Wavelet Packet Analysis is an important generalization of wavelet analysis, pioneered by Ronald Coifman, Yves Meyer and Victor Wickerhauser. Nowadays, Wavelet Packet Transform (WPT) is now becoming an efficient tool in signal processing, comparing with the ordinary wavelet analysis, it has special abilities to accomplish higher discrimination by analyzing the higher frequency domains of a signal. The frequency domains divided by the wavelet packet can be easily selected and classified according to the characteristics of the analyzed signal. So the wavelet packet is more suitable than wavelet in signal analysis. In this paper the principles of wavelet packet thresholding (Daubechies 3 with level 3) for de-noising were analyzed for Raman Spectral data of Sr^{2+} modified PMN-PZT.

Keywords : De-noising, Wavelet Packet Transform, Raman Spectroscopy.

1. INTRODUCTION

Nowadays, Raman spectroscopy plays an important role in various fields as a non-destructive, rapid analysis, high-reliability, and detection technology. In the process of Raman spectra acquisition, especially in small samples of low concentration, a lot of background spectrum, noise and other kinds of random factors are existent; therefore the data collection is inevitable effected on. Which makes it difficult to extract the useful information; the noise can even cause a wrong result. The noise greatly controls the attainment of accuracy analytical result.

There are several methods for signal de-noising of Raman spectra. One of the popular recent methods is based on the wavelet transform because of its excellent properties. The Wavelet transform was inspired by the idea that we could vary the scale of the basis functions instead of their frequency. The idea behind wavelets is to analyze according to scale, instead of representing a function as a sum of weighted delta functions (as in time domain), or as a sum of weighted sinusoids (as in the frequency domain), it represents the function as a sum of time - shifted (translated) and scaled (dilated) representations of some arbitrary function, which is called wavelet.

An advantage of wavelet transforms is that, wavelet analysis allows the use of long time intervals for low – frequency information, and shorter regions for high – frequency information. Wavelets have been shown to be effective as low and high frequency filters, revealing the middle frequency Raman features. Wavelet method is an accurate and reliable tool for studying signals with sudden changes of phase and frequency. It is very useful in diverse fields such as signal processing, machine learning, signal smoothing and signal de-noising, speech recognition, data mining and pattern recognition, face recognition biomedical imaging, data compression.

Wavelet Packet Analysis is an important generalization of wavelet analysis, pioneered by Ronald Coifman, Yves Meyer and Victor Wickerhauser. Nowadays, Wavelet Packet Transform (WPT) is becoming an efficient tool in signal processing as comparing to ordinary wavelet analysis; it has special abilities to accomplish higher discrimination by analyzing the higher frequency domains of a signal. The frequency domains divided by the wavelet packet can be easily selected and classified according to the characteristics of the analyzed signal. So the wavelet packet is more suitable than wavelet in signal analysis.

* School of Engineering & Technology, Department of Applied Science, P.N. -32-34, Knowledge Park-III, Sharda University Greater Noida - 201306 INDIA Email: nkshukla.kumar4@gmail.com, siddiq.abulhasan@gmail.com, khursheed.alam@sharda.ac.in

2. WAVELET PACKET ANALYSIS AND THRESHOLDING

In the wavelet packet frame, the de-noising idea is same as in the wavelet frame, except decomposing the low and high frequency part of spectrum at the same time. Apart from that the method of thresholding process in the wavelet packet is the same as in the wavelet analysis.

The continuous wavelet transform for signal is $f(t) \in L^2(R)$ is $C_f(a,b) = < f, \psi_{a,b} > = |a|^{-\frac{1}{2}} \int_R f(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt$

and its reconstruction formula (inverse transform) is $f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{a^2} C_f(a,b)\, \psi\left(\frac{t-b}{a}\right) da.db$ where

$C_\psi = \int_0^\infty \frac{|\psi(\bar{\omega})|}{|\bar{\omega}|} d\bar{\omega} < \infty$, $\psi(t)$ stands for wavelet function, $\overline{\psi(t)}$ for complex conjugate function, C_ψ and $f(t)$ is independent

$\psi(t) =$ will become $\psi_{a,b} = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right)$ by dilating and translating, $a, b \in R$, $a \neq 0$. if we set ,

$a = j$, $b = 2^j kb^0$, k ranges over z, and b_0 is normalized, then there exist $\Psi_{a,b} = \frac{1}{\sqrt{a}}\left(\frac{t-b}{a}\right) = a^{-\frac{j}{2}} \psi(2^{-j} t - k) = \psi_{j,k}$,

and we get the discrete wavelet transform coefficients $C_{j,k} = \int_{-\infty}^{+\infty} f(t) \overline{\psi_{j,k}(t)} dt = < f, \psi_{j,k}(t) >$, its reconstruction

formula is $C_{j,k} = C \sum_j \sum_k C_{j,k} \psi_{j,k}(t)$, C is a constant independent of signals.

During the wavelet packet decomposition, the high-frequency and low-frequency band in the frequency domain occupied respectively half of the wide frequency band. When decomposed next time, low-frequency signal is divided into two equally frequency wide band, so did high- frequency. The decomposition next time is the same as before. The process of WPT by filters is similar to the WT. the difference between WPT and WT is that only low-frequency band is divided into two bands continuously in WT, while both a high low frequency band is divided in WPT. Finally, the entire band in WPT is divided into uniform frequency bands.

The coefficients of $f(t)$ in the wavelet transform is $C_k^{n,j}$, the Nth wavelet coefficients is

$C_k^{n,j} = 2^{-\frac{j}{2}} \int_{-\infty}^{\infty} f(t) \overline{\psi_n(2^{-j} t - k)} dt$. Wavelet packet decomposition algorithm is: with $C_k^{n,j}$, the n^{th} and j^{th} layer

coefficients $C_k^{n,j-1}$ is decompose into wavelet coefficients $\{C_k^{2n,j}, C_k^{2n+1,j} : k \in Z\}$, that is $\begin{Bmatrix} C_k^{2n,j} = \sum_k h_{l-2k} C_k^{n,j-1} \\ C_k^{2n+1,j} = \sum_k g_{l-2k} C_k^{n,j-1} \end{Bmatrix}$

As shown in figure 1, where $\downarrow 2$ means down-sampling and $\uparrow 2$ means up-sampling.

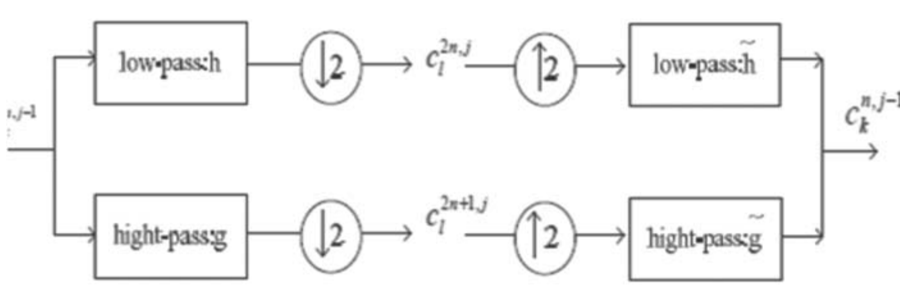

Fig. 1. Wavelet packet decomposition and reconstruction based on filters

Wavelet packet reconstruction algorithm is : with $c_k^{2n,j}, c_k^{2n+1,j}, c_l^{2n,j-1}$, that is $c_l^{2n,j-1} = \sum_k h_{2k-1} c_l^{2n,j} + \sum_k g_{2k-1} c_l^{n,j}$,

where $g_{2k} = (-1)^k \bar{h}_{-k+1}$ And up-sampling must be processed during reconstruction.

H can be taken as low-pass filter coefficients; g as high pass filter coefficients. Response function of the low-pass filter is $H(\omega) = 2^{-0.5} \sum_{k \in z} h_k e^{-ik\omega}$, high pass filter response function is $G(\omega) = 2^{-0.5} \sum_{k \in z} g_k e^{-ik\omega} = 2^{-0.5} \sum_{k \in z} (-1)^k \bar{h}_{-k+1} e^{-ik\omega}$, wavelet function is $\psi(x) = 2^{-0.5} \sum_{k \in z} g_k \phi(2x - k)$, the scaling function $\phi(x) = 2^{-0.5} \sum_{k \in z} h_k \phi(2x - k)$, $\{c_k^{n,j}\}$ is called the wavelet packet on h_k. Following figure 2 is shown for three-layer space spatial analysis by wavelet packet decomposition.

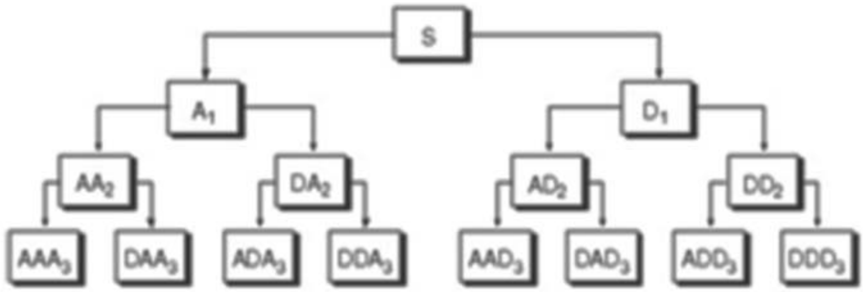

Fig. 2. Wavelet Packet decomposition tree

In figure 2, S is the signal, A is the low frequency (approximation part), D is high-frequency (detail part). The number is the order number of decomposition (the number of scales). The signal S can be expressed as a variety of decomposition methods, for S = A1 + AD2 + ADD3 + DDD3 and so on.

In figure 2 the first line represents the frequency the frequency band belongs to original signal, the following lines number represents the scale of wavelet decomposition, the number of columns is the frequency and location parameter. From the figure 2 we can see that after decomposition the signal frequency band is divided into two parts, get the second layer which has a high and low frequency sub-band. Only low frequency in WT is decomposed continuously, high frequency sub-band unchanged. In WPT each sub-band is divided into another two lower levels sub-band. The whole frequency band is covered by layers of sub-band signals.WPT breaking limitations that the high frequency space can't be decomposed in WT although it can divide the low frequency limitations of space. So WPT can get more extensive time-frequency localization information, and is more suitable for transient signals, non- stationary signal analysis and detection.

2.1. Thresholding Principle

The theory of thresholding is based on; orthogonal wavelet transform in particular has a strong décor- relation to the data. It enables the signal to focus the energy on some large wavelet coefficients in the wavelet domain, while the noise is distributed in the whole wavelet domain. Thus, by the wavelet decomposition, the signal amplitude of the wavelet coefficients of magnitude is greater than the noise factor. The relative large amplitude wavelet coefficients can be considered generally the main signal, while the smaller amplitude coefficient largely noise. By the threshold method the signal factor can be reserved, most of the noise factor can be decreased to zero. Finally, the effective signal can be restored through reconstruct wavelet coefficients obtained by inverse wavelet transform.

The selection of thresholds has been two ways: hard and soft threshold. For the soft threshold the absolute value of the signal is compared with the threshold value, when the data of absolute value is less than or equal the threshold, it will reduce to zero; when greater than the threshold value, the data becomes the difference value between the data and threshold value. For the hard threshold the data less than or equal to the threshold point becomes zero, otherwise the data unchanged. These formulas are as following:

Hard threshold method : $$W_{j,k} = \begin{cases} W_{j,k}, |W_{j,k}| \geq \lambda \\ 0, |W_{j,k}| < \lambda \end{cases}$$

Soft threshold method : $$W_{j,k} = \begin{cases} \text{sgn}(W_{j,k})(|W_{j,k}| - \lambda), |W_{j,k}| \geq \lambda \\ 0, |W_{j,k}| < \lambda \end{cases}$$

For the hard threshold method, the department $W_{j,k}$ is discontinuous at $W_{j,k} = \lambda$, which brings to the reconstructed signal oscillation; and soft threshold method to calculate $W_{j,k}$ with overall good continuity. Soft threshold is chosen in this article.

2.2. Wavelet Packet De-noising Algorithm

The one dimensional model of signals with additive noises can be shown as follows :

$$y(n) = x(n) + \sigma e(n), n = 0, 1, 2, ..., N$$

where, $y(n)$ denotes noise-containing signals, $x(n)$ denotes real signals, $e(n)$ is white Gaussian noise with a normal distribution, and N(0, 1) denotes the deviation of noise signals.

The de-noise algorithm is given bellow :

- Initially, decompose the input signal using Wavelet Packet: choose a wavelet and determine the decomposition level of a wavelet packet transform N, then implement layers wavelet decomposition of signal S.
- Calculate the optimal tree (*i.e.*, to determine the best wavelet basis). Calculate the best tree with given entropy standard.
- Select the thresholding method and theresholding rule for quantization of wavelet coefficients. Apply the thresholding on each level of wavelet decomposition and this thresholding value removes the wavelet coefficients above the threshold value (soft and hard thresholding).

Finally, reconstructed wavelet packet, according to the N^{th} low-frequency wavelet coefficients and N^{th} high frequency coefficients quantified, reconstructing wavelet packet.Among these four steps, the most critical steps is to select the threshold and how to quantify the threshold. To some extent, it is directly related to the quality of signal de-noising. It is directly reflex the quality of de-noising there are several major entropy, Shannon entropy, P-order standard entropy, log energy entropy, entropy thresholding, SURE entropy. Shannon entropy with fix from thresholding (scaled with white noise) is chosen for this paper.

3. TEST ANALYSIS AND EVALUATION

3.1. Signal – to – Noise Ratio (SNR)

In order to verify the performance of the proposed de-noising approach, computer generated noises with variable amplitudes are added to well-known benchmark signals; moreover, the classical algorithms are performed for comparison. A number of quantitative parameters can be used to evaluate the performance of the de-noising procedure in terms of the reconstructed signal quality. In this case, the following parameters are compared:

$$\text{SNR}(db) = 10 \ln \frac{\sum_{k=1}^{N} x^2(k)}{\sum_{k=1}^{N} [x(k) - x'(k)]^2}$$

Where $x'(k)$ is the de-noise signal, and $x(k)$ is the original signal. The constant, N, is the number of sample composing the signals.

3.2. Root-Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) is a frequently used measure of the difference between values predicted by a model and the values actually observed from the environment that is being modelled. These individual differences are also called residuals, and the RMSE serves to aggregate them into a single measure of predictive power.

4. EXPERIMENT AND RESULT DISCUSSION

Raman Spectral data of Sr^{+2}‘modified PZT-PMN (with Sr = 0.025, 0.5. 0.75, 1.0) have been collected by Raman spectrometer of Remisaw Rm-1000 spectro scope, which use 1.19 nm as a unit the range of 200 – 1000 cm^{-1}. The experiment was carried at a room temperature about 25°C. The data has been obtain in .txt format and has two kinds of 876X2 data that represents the intensity-Raman shift and the intensity-pixel. Export the intensity value (876X1 only and take it as one-dimensional signal).

The major steps of the thresholding -based wavelet packet de-noising methods for Raman Spectral signal are:

1. Decompose the noisy signal using db3 (the Daubechie’s wavelet) in three levels with wavelet packet transform, and then get a full decomposition tree.
2. Calculate the optimal tree (*i.e.*, to find out the best wavelet packet basis). Calculate the best tree with given entropy standard.
3. Threshold the best bases of wavelet packet decomposition by using the fixed from threshold function mentioned in this paper.
4. Reconstruct the wavelet packet decomposition coefficients, and get de-noised signal.

Fig. 1, 3, 5. 7 shows the spectrum of the original signal and Fig. 2, 4, 6 and 8 shows the spectrum of the de-noised signal by using the soft threshold function.

Table 1. Denoising Results

Db3	*SNR*	*RMSE*
Sample 1	40.9029	27.8381
Sample 2	40.5453	27.9932
Sample 3	39.7494,	25.0349
Sample 4	38.2964,	25.1362

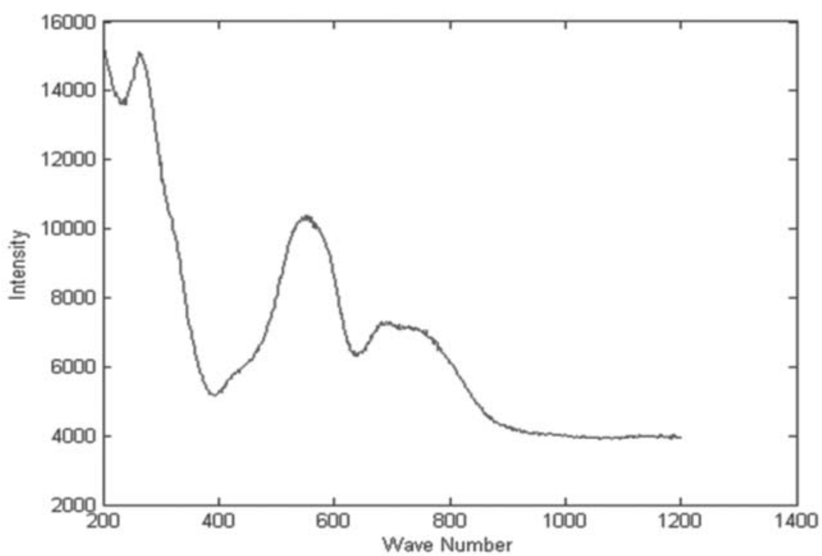

Fig. 2. Original Raman Spectra of Sample 1

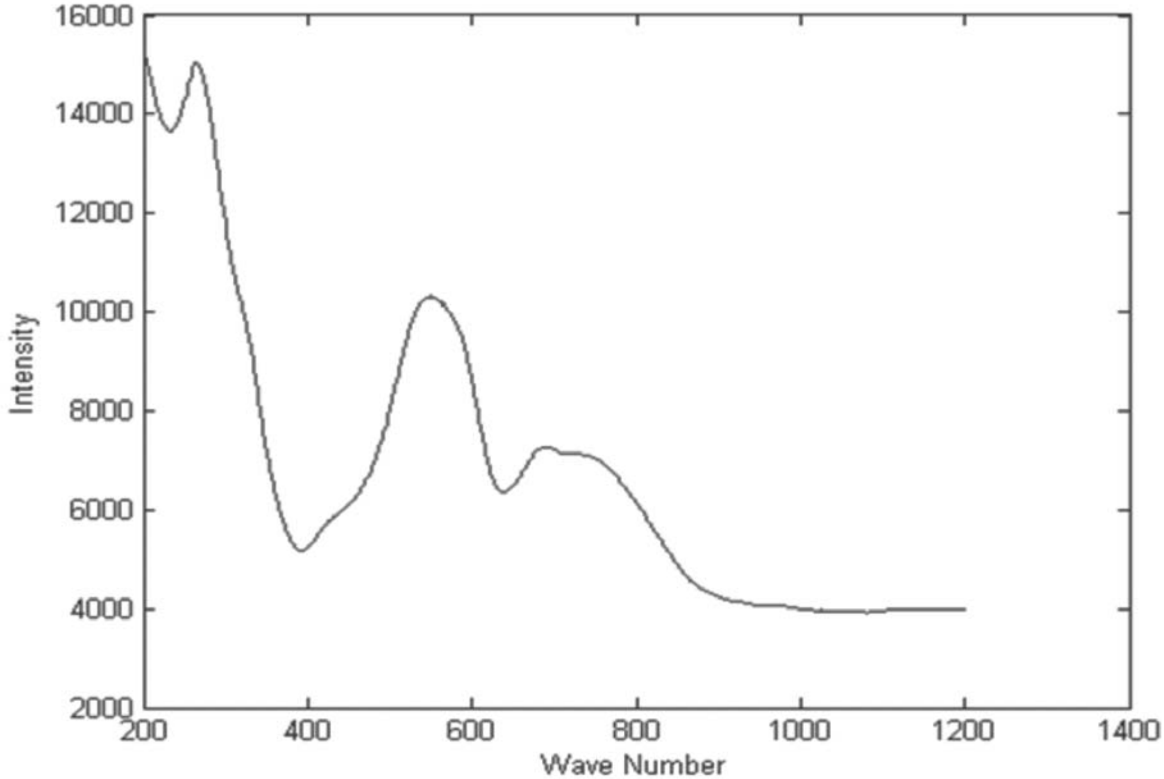

Fig. 2. Denoising Raman Spectra of Sample 1

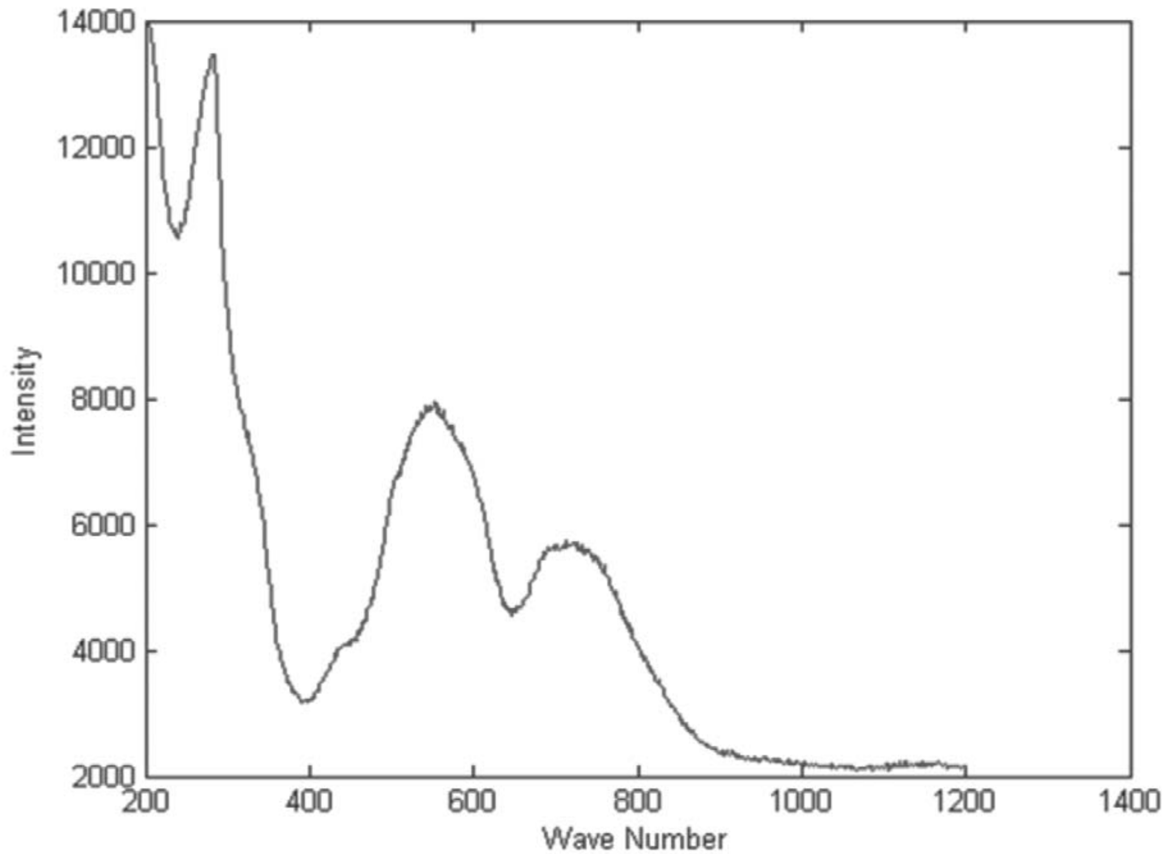

Fig. 3. Original Raman Spectra of Sample 2

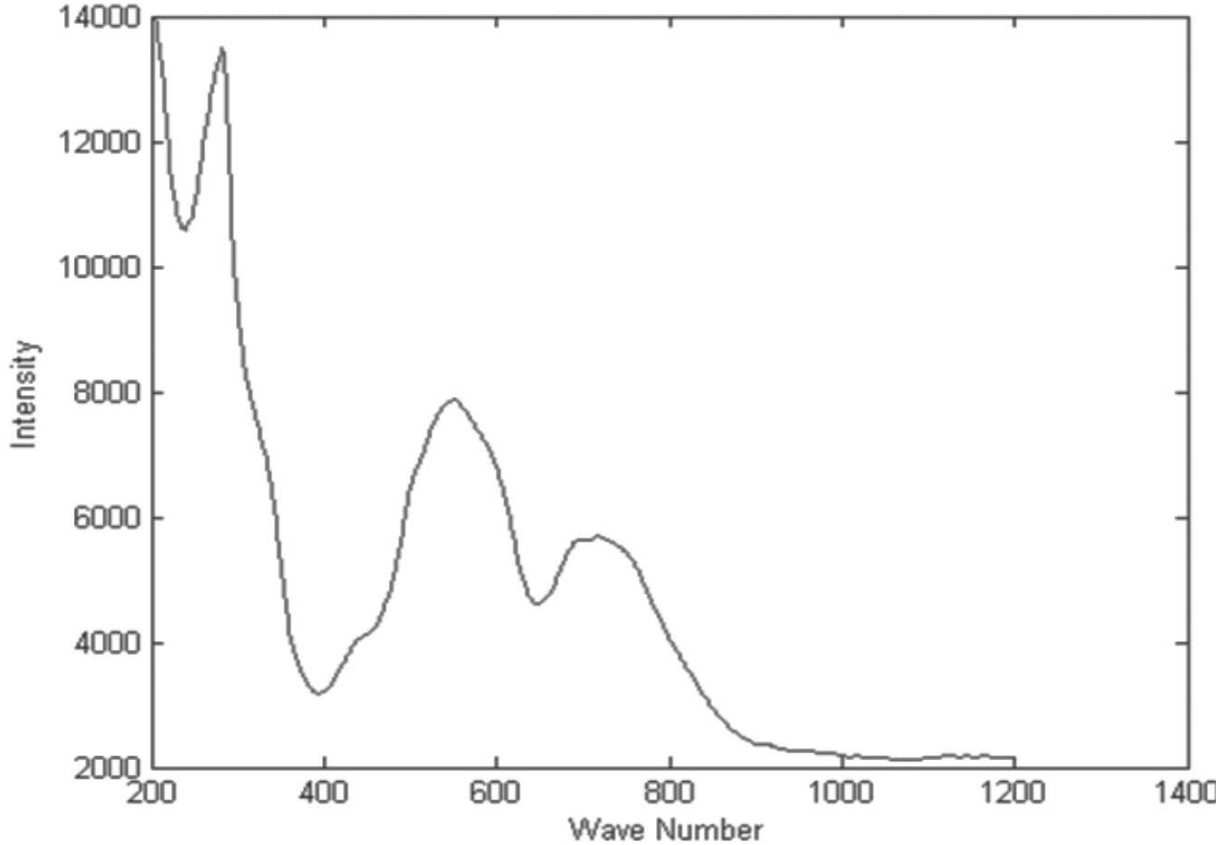

Fig. 4. Denoising Raman Spectra of Sample 2

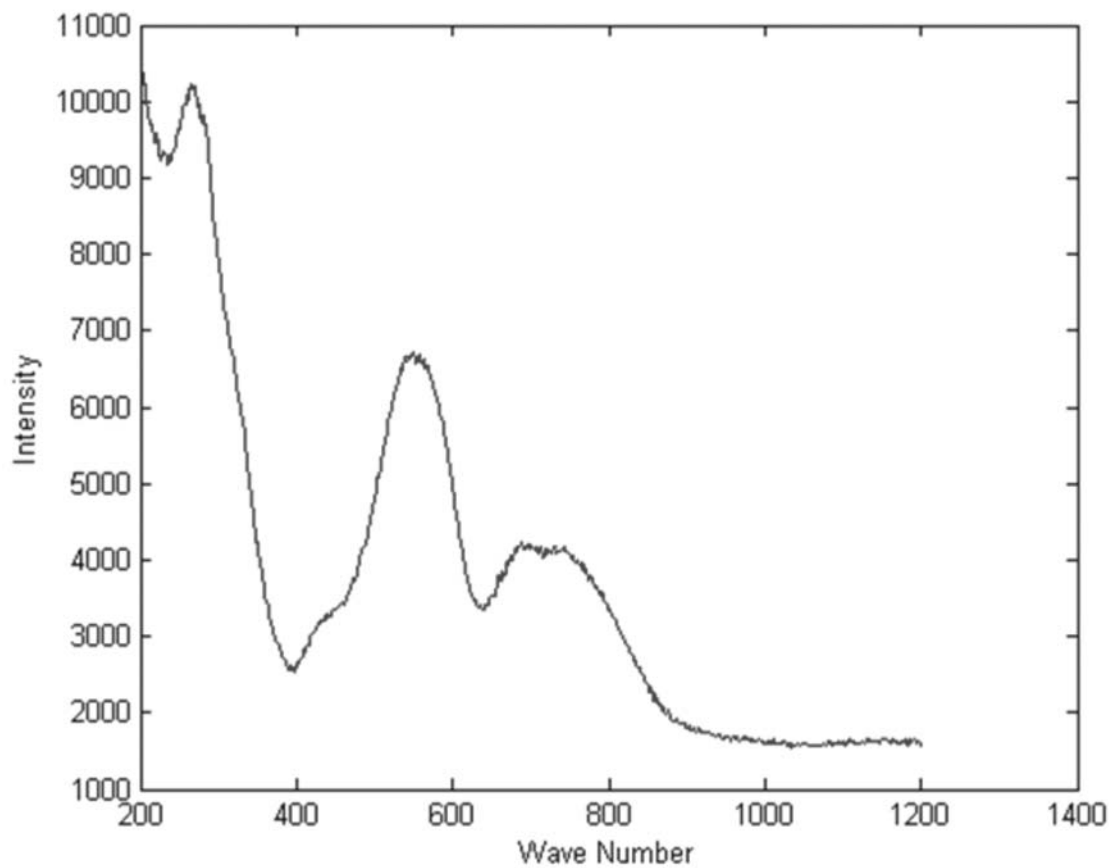

Fig. 5. Original Raman Spectra of Sample 3

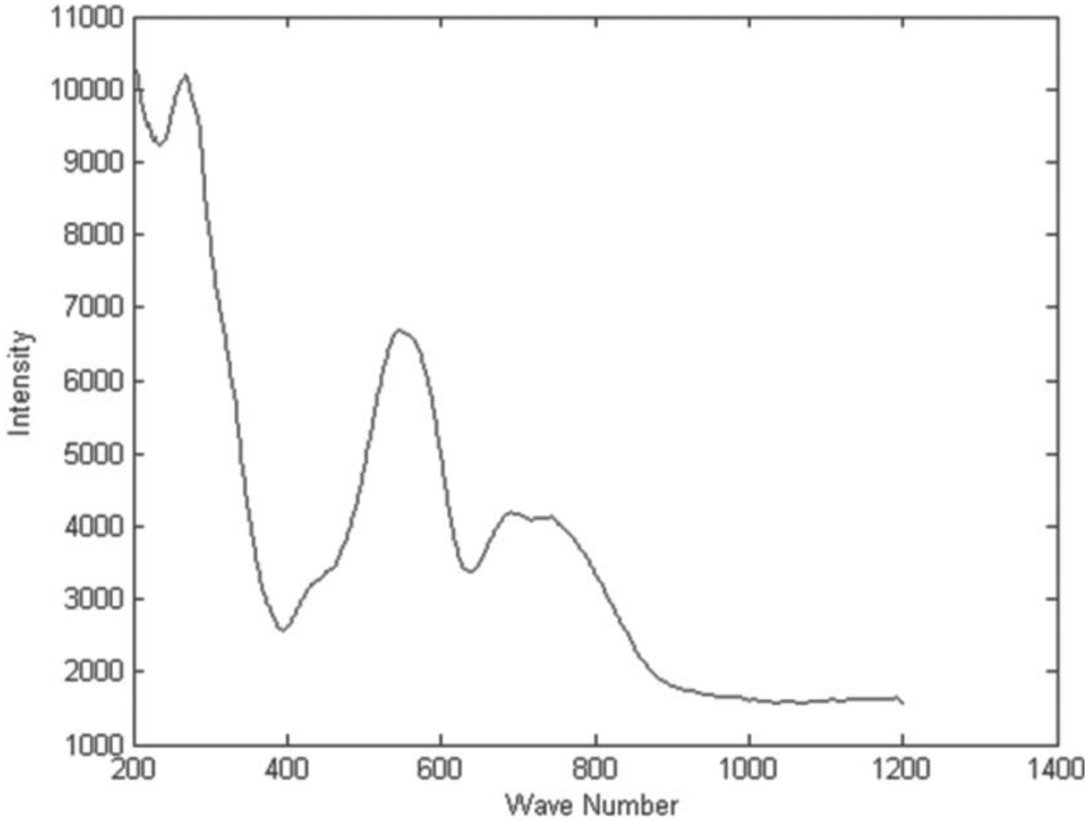

Fig. 6. Denoising Raman Spectra of Sample 3

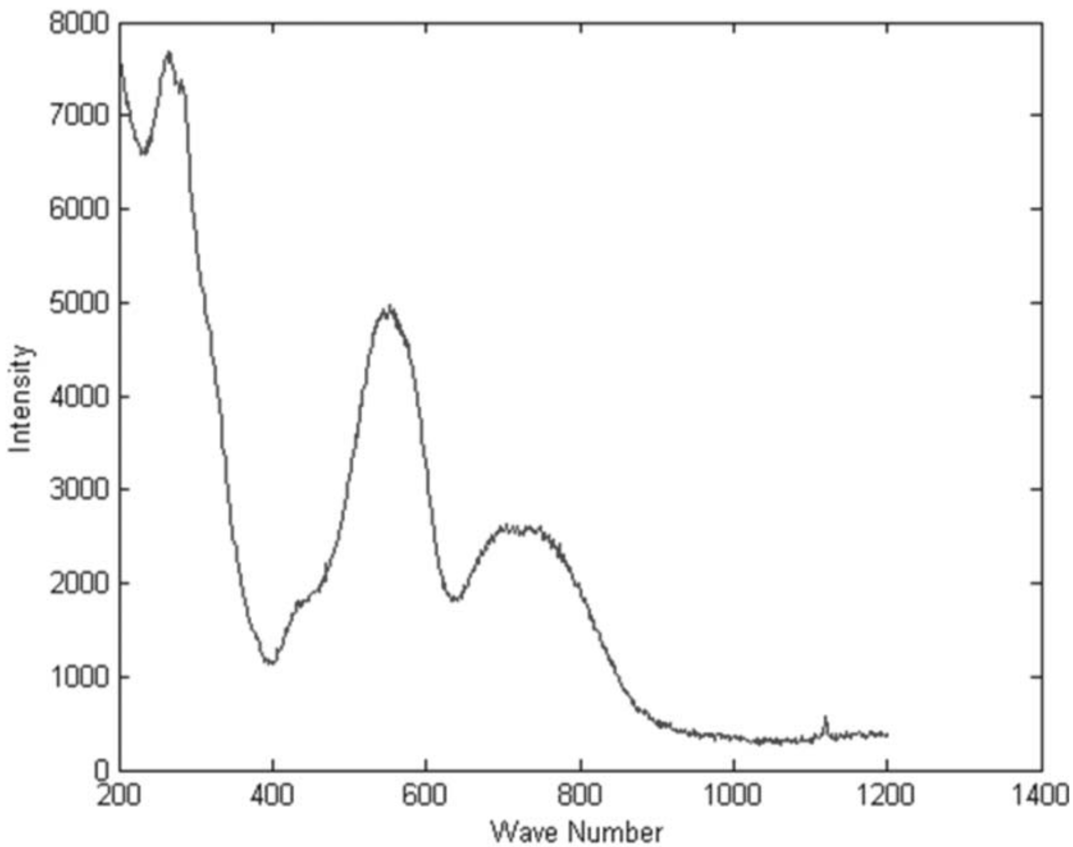

Fig. 7. Original Raman Spectra of Sample 4

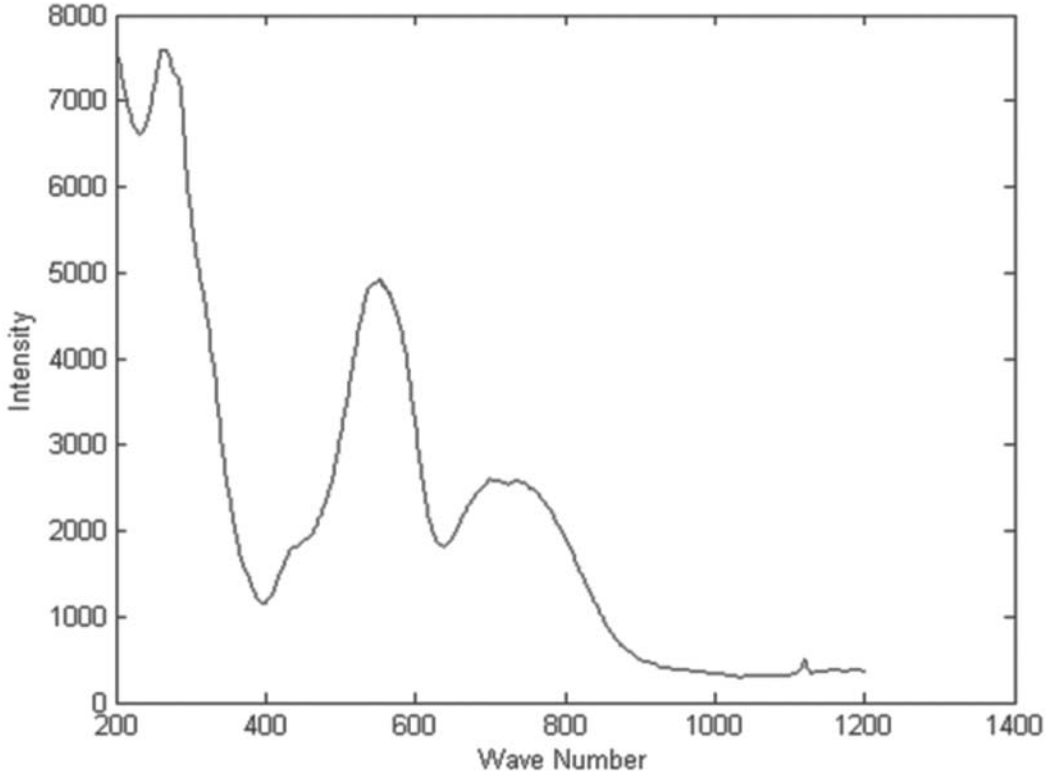

Fig. 8. Denoising Raman Spectra of Sample 4

5. CONCLUSION

This paper describes the principle of wavelet packet transform, thresholding and the general implementation steps, and by Matlab simulation results, related data prove that the signal de-noising based on wavelet packet is valuable.

In the de-noising process by WPT, the key is to select the de-noising threshold and to select the WPT function based on experience (db3 with level 3). Also the signal-to-noise ratio and root-mean square error shows that the de-noised results improve the background of original result shown in the table [1].

6. REFERENCES

1. C. V. Raman,"A new type of secondary radiation", *Nature*, 121, 619, 1928.
2. Y. Yang, L. Guo and F. Li, "Application of wavelet transform in spectral data de-noising", *IEEE*, 978-1-0246-4/11, 2011.
3. L. Gang, "Noise removal of Raman spectra using interval thresholding method", *IEEE,* DOI 0.1109/IITA, 573, 2008.
4. C. Chen, F. Peng, Q. Cheng and D. Xu, "Application of Wavelet Packet Transform to Compressing Raman Spectra Data", *Proc. of SPIE*, 7280, 728009-1, 2009.
5. F. Ehrentreich, L. Summchen,"Spike removal and De-noising of Raman spectra by wavelet transform methods", *Analytical chemistry*, Vol. 73, no. 17, 2001, pp. 4364-4373.
6. W. Cai, L. Wang, Z. Pan, J. Zuo, C. Xu, X. Shao,"Application of the Wavelet of Raman spectra", *Journal of Raman Spectroscopy*, Vol. 32, No. 3, 2001, pp. 207-209.
8. Ramos, P. M, Ruisa'nchez I.,"Noise and background removal in Raman spectra of ancient pigments using wavelet transform", *Journal of Raman Spectroscopy*, 36(9), 2005, pp 848–856.
9. Hu Y, Jiang T, Shen A, Li W, Wang X, and Hu J,"A background elimination method based on wavelet transform for Raman spectra", Chemometer Intell. Lab. Syst, 85(1), 2007, pp 94-101.

10. Kumar A. and Mishra S. K., Effects of Sr^{2+} substitution on structural, dielectric and piezoelectric properties of PZT-PMN ceramic, *Int. J. Miner. Metall. Mater*,Vol. 21, No. 2. 2014.

11. H.–Y. Gao, “Choice of thresholds for wavelet shrinkage estimate of the spectrum,” *J. Time Series Analysis*, Vol. 18, pp. 231–51, 1997.

12. Donoho, D.L.; Johnstone, I.M. “Adapting to unknown smoothness via wavelet shrinkage”, J. *Am. Statist. Assoc*, Vol (90), 1200–1224, 1995.

13. Agnieszka Martyna, Aleksandra Michalska, Grzegorz Zadora, “Interpretation of FTIR spectra of polymers and Raman spectra of car paints by means of likelihood ratio approach supported by wavelet transform for reducing data dimensionality”, *Analytical and Bioanalytical Chemistry*, Springer, Volume 407, Issue 12, pp 3357–3376, 2015.

14. Yu, Xiao; Xu, Liang; Mo, Jia-qing; Lü, Xiao-yi, “Raman spectroscopy de-noising based on EEMD combined with VS-LMS algorithm”, Optoelectronics Letters,Springer, Volume 12, Issue 1, pp.16-19, 2016.

15. Foo-Tim, Yi-zeng Liang, Junbin Gao and Xue – Guang Shao, “Chemometrics From Basics to Wavelet Transform”, *Wiley – Interscience*. Vol - 164. pp 99-141, 2004.

16. Andrzej Kwiatkowski, Marcin Gnyba, Janusz Smulko and Pawe³ Wierzba, “Algorithms of Chemical Detection using Raman Spectra”, *Polish Academy of Science*, Vol - XVII, No.- 4. pp 549-560, 2010.