We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,800
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Discrete Wavelet Transform in Compression and Filtering of Biomedical Signals

Dora M. Ballesteros¹, Andrés E. Gaona² and Luis F. Pedraza³

¹Faculty of Engineering, University Military Nueva Granada
²Faculty of Engineering, University Francisco José de Caldas
³Faculty of Technology, University Francisco José de Caldas

Colombia

1. Introduction

Biomedical signals are a kind of signals that are measured from a specific part of the body, for example from the heart (electrocardiography: ECG), muscles (electromyography: EMG) and brain (electroencephalography: EEG). This kind of signals have a non-stationary behavior, it means the behavior through the time is changing every time window. For this reason, the pre-processing, processing, and analysis should be different of the deterministic and stationary signals. One of the methods used in the last years to examine biomedical signals is the Discrete Wavelet Transform (DWT), it represents both time and frequency the signal’s characteristics in a multi-resolution mode.

In this chapter, we are going to present two applications of the DWT in biomedical signals, it is known as filtering and compression. When you have a device that measures the body’s signals, it is desired that the information stored or transmitted have high quality and low redundancy; this corresponds to apply a filter and compress the signal. These two blocks (filtering and compression) are added once the signal is acquired and processed by digital signal processing methods. The goal of using the DWT in an algorithm of filtering and compression biomedical signals is the possibility of choosing the signal’s coefficients with a significant energy and discards the others that have a very low percentage of all energy. This is possible because in every level of decomposition, the energy of different frequencies and time position is related to a specific coefficient.

In the first part, we present one model of filtering of biomedical signals based on Discrete Wavelet Transform. We analyze the different parameters in the model and its relation to the quality of the new signal. Every parameter affects in low or high manner the quality of the filtered signal and we present the most common test to probe the signal’s distortion when the coefficients with low energy have been removed. Additionally, we present some results with one real EMG signal with different configuration of the parameters.

In the second part, we extend the model of filtering to include the stage of compression; we explain the encoding block, which is added to the compression model. Two lossless encoding methods are explained and compared. The compression of some records of ECG is presented.
Finally, in the third part, the architecture of the Discrete Wavelet Transform on a FPGA is shown. The convolution and sub-sampling processes are modeled using VHDL and its performance is simulated using a CAE tool.

2. Filtering technique for background noise

The basic idea of filtering technique is to improve the signal to noise ratio, in fact to reduce the background noise in the biomedical signal. Because noise can affect the reading and interpretation of the signal, a pre-processing step is desirable before the computer analysis. Because the external noise does not have a specific band and its frequency is commonly superposed to the biomedical frequency, it is necessary to design an intelligent model which can be adaptable to different kind of signals. It is possible with the Discrete Wavelet Transform.

The classic technique (Donoho & Johnstone, 1994) includes three important stages: the decomposition of the signal; the identification of low energy coefficients and its rejection (thresholding); and finally, the reconstruction of the new coefficients. It is shown below;

![Diagram of the filtering technique](Fig. 1. The filtering technique.)

The selection of the DWT is due to the simultaneous representation of the signal and noise in time and frequency. The technique is applied for the model with additive noise, according to:

\[ ns = s + an \]  

\[ NS = DWT{ns}; \quad S = DWT{s}; \quad AN = DWT{an} \]

Then

\[ NS = S + AN \]  

If the external signal corresponds to white noise, its energy is sparse with low amplitude. Then, the wavelet coefficients of \( ns \) with low amplitude correspond to the noise of the signal. The noise can be eliminated if the coefficients below a threshold are turned to zero (thresholding).

Every stage of the Figure 1 has parameters related to the performance of filtering. Specifically, the decomposition and reconstruction involve the base wavelet and the number of levels; and the thresholding involves the threshold and the rule.

2.1 Parameter selection

The time and frequency characteristics can vary from signal to signal, then it is necessary to establish a method to identify the best conditions for each specific type of signal (normal, clean signal, noisy signal).
with pathology, of woman, of man, ...). In this section, we present the methodology of selection of the main parameters in the filtering model.

To evaluate in objective form the performance of every combination, there are measurements of the quality of the filtered signal. One of the most used is the PRD (Percentage of RMS), which is calculated according to:

$$PRD = \sqrt{\frac{\sum (ns - f)^2}{\sum (ns)^2}}$$

(3)

f represents the filtered signal. If PRD is high, the filter could have eliminated important components of the signal; while, if the PRD is very low, the filter could have not eliminate the noise.

2.1.1 Wavelet family

In Table 2, the most common families are presented. This list is supported by Matlab ©. The index is related to the length of the filter, for example, for sym4 the length is eight. We suggest selecting the base according to the similarity with the biomedical signal. At the most it looks like, better is the representation.

In relation to the length of the filter, it is not recommended to use long filters for short time signals, for example, if the time is 10ms, sym10 is better than sym45.

<table>
<thead>
<tr>
<th>Wavelet Family</th>
<th>Wname</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daubichies</td>
<td>db1 or haar, db2, ... db10, ... db45</td>
</tr>
<tr>
<td>Coiffltes</td>
<td>coif1,..., coif45</td>
</tr>
<tr>
<td>Symlets</td>
<td>sym2,...sym4,...sym45</td>
</tr>
<tr>
<td>Discrete Meyer</td>
<td>Dmey</td>
</tr>
<tr>
<td>Biorthogonal</td>
<td>bior1.1, bior1.3, bior1.5, bior2.2, bior2.4, bior2.6, bior2.8, bior3.1, bior3.3, bior3.5, bior3.7, bior3.9, bior4.4, bior5.5, bior6.8,</td>
</tr>
</tbody>
</table>

Table 1. Wavelet Family

2.1.2 Levels of decomposition

The number of levels of decomposition (N) depends on the relation between the sampling frequency and the bandwidth of the signal. A big N is required if the relation is high; an initial rank can be 3 to 10 levels.

2.1.3 Thresholding rule

There are two important thresholding rules applied in most of papers of biomedical signal denoising: soft and hard threshold. However, other rules have been proposed (Quian, 2000).

Soft threshold is defined by (Donoho, 1995):

$$g(x) = \text{sgn}(x)(|x| - |\text{th}|) \quad \text{if } |x| \geq |\text{th}|;$$

$$g(x) = 0 \quad \text{else}$$

(4)
Where \( th \) is the threshold, \( x \) is the initial amplitude (input) and \( g(x) \) is the result (output). In Figure 2 the function is represented.

\[
\begin{align*}
g(x) &= x & \text{if } |x| \geq |th|; \\
&= 0 & \text{else}
\end{align*}
\]

The difference between the soft and hard rules is the output when the input exceeds the threshold. In both cases, the output is zero when the input is less than the threshold. The function is presented in Figure 3.

**Fig. 2. Soft threshold.**

Hard threshold is defined by (Mallat, 1998):

\[
\begin{align*}
g(x) &= x & \text{if } |x| \geq |th|; \\
&= 0 & \text{else}
\end{align*}
\]

The difference between the soft and hard rules is the output when the input exceeds the threshold. In both cases, the output is zero when the input is less than the threshold. The function is presented in Figure 3.

**Fig. 3. Hard threshold.**

### 2.2 Results: A case of study of filtering SEMG

The electromyography signal is one of the biomedical signals, which correspond to record of the muscle activity. Because in a typical record, the activity and non-activity regions are acquired and transmitted, it is desirable to clean the signal for improvement the contrast between the two regions.
We analyze the relation of the triangle: threshold, energy retained and percentage of zeros. When the threshold increases the percentage of zero increases too, but, the energy decreases. The appropriate point of the triangle corresponds to the maximum percentage of zeros with the maximum energy retained. Experimentally, we found that the 95% of energy retained is adequate for a right interpretation of the signal (Ballesteros & Gaona, 2007, 2008). In table 2, the results of our study are presented.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PRD</th>
<th>Zeros %</th>
<th>Energy Retained %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sym6</td>
<td>sym8</td>
<td>sym6</td>
</tr>
<tr>
<td>Th1 &amp; hard</td>
<td>0.14</td>
<td>0.16</td>
<td>71.7</td>
</tr>
<tr>
<td>Th1 &amp; soft</td>
<td>0.25</td>
<td>0.27</td>
<td>71.7</td>
</tr>
<tr>
<td>Th2 &amp; hard</td>
<td>0.22</td>
<td>0.25</td>
<td>77.5</td>
</tr>
<tr>
<td>Th2 &amp; soft</td>
<td>0.36</td>
<td>0.38</td>
<td>77.5</td>
</tr>
</tbody>
</table>

Table 2. Results of filtering: N=3.

Figure 4 presents the results with different parameters. In the left side, the sym6 base was used in the decomposition, threshold was equal to 0.26 and the hard thresholding was applied. It obtained the 95% of the energy retained for the 77.5% of the wavelet coefficients set to zero. In the right, the rule was soft. In this case, the 83% of the energy was conserved with the same number of wavelet coefficients set to zero.

According to table 2 and figure 4, the best combination is th2&hard with the sym6 base. This satisfies the balance criteria.

3. Compression model

The basic idea of the compression model is to reduce the amount of information. Although in some applications the quality in the compressed signal is not important, in the case of biomedical signals the difference must to be the minimum. The purpose is to find the redundancy in the information and eliminate it.

In addition to the three stages of filtering model, an encoding block should be used to improve the Compression Relation (CR). The compression model is presented in Figure 5.
The encoding can be by lossless or lossy methods (Hankerson et al., 2005). First, the data in the receiver are the same that in the transmitter; while in the second, a part of the information is lost in the process. Because in biomedical signal compression model is appropriate to retain the non-redundant information, we suggest methods without loss of information.

### 3.1 Lossless methods

We present two of the lossless methods: Run-Length and Huffman. Both take advantage of the thresholding stage to increase the compression ratio (CR).

#### 3.1.1 Run-Length encoding (RL)

This method is used when a number (commonly zero) is repeated many times in a sequence, and then the data in the original stream is replaced by the number and its repetition. (Smith, 2003). The length of the new data decreases when the quantity of zeros increases. Suppose you have the following data stream:

```
1 5 0 0 0 0 0 0 0 10
```

The output stream with run-length encoding is:

```
1 5 0 7 10
```

And the CR is $10/5=2$.

#### 3.1.1 Huffman encoding

Huffman encoding defines the codebook according to the repetition of every data. It uses more bits in the no-frequent data and fewer bits for the data with higher occurrence (Huffman, 1952). An important feature of Huffman code is that no code can be the header of another; the decoding of data is unique.

The steps for creating the code are:

1. Sort the data from high to low level of repetition.
3. Repeat second step until all data have been combined.
4. Draw the Huffman tree with branches of two nodes, where data sets with higher levels of repetition are located to the left of the tree and the lowest level on the right. Assign
‘1’ to the data of the left and a ‘0’ to the right. Huffman code is read from top to bottom of the tree.

Suppose we have the following list of repetitions into a stream:

<table>
<thead>
<tr>
<th>Date</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetition</td>
<td>60</td>
<td>15</td>
<td>22</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 3. Date example Huffman code.

The tree according to the four previous steps is:

![Huffman tree](image)

Fig. 6. Huffman tree.

And the codebook is presented in Table 4:

<table>
<thead>
<tr>
<th>Date</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huffman Code</td>
<td>1</td>
<td>001</td>
<td>01</td>
<td>000</td>
</tr>
</tbody>
</table>

Table 4. Codebook.

### 3.2 Results: A case of study in compression of ECG

The first step to compress a biomedical signal is to analyze its characteristics and determine the parameters for decomposition and the thresholding stage (section 2.1) No matter what encoding method, the output coefficients of the threshold block must retain much of its energy.

#### 3.2.1 Compression with DWT and Run-length/Huffman encoding

Wavelet Transform followed by run length encoding has been used in the last decade in biomedical signal compression. Chen et al (2006) consider that the Huffman code is less
robust against the different statistical characteristics, because it is necessary a prior knowledge of its symbol statistics; but we develop a solution with a static unique code which can be used in different type of ECG.

Ballesteros & Gaona (2009, 2010) developed a compression model for ECG. This work had the following parameters:

- **Family:** Daubichies, Symlets and BiorSplines (db6, sym6, bior5.5)
- **Levels of decomposition:** 3 and 4 levels.
- **Amplitude of the threshold:** it is calculated by
  \[
  \text{th}_p = 0.0075 \times \max(w_p) \quad (6)
  \]
  where \( \text{th}_p \) is the threshold of level \( P \), \( w_p \) is the wavelet coefficients of the detail/coarse of level \( P \) and \( \max(.) \) is the maximum function.
- **Rule of thresholding:** hard.
- **Encoding method:** run-length and Huffman.
- **Time:** 2, 4, 5 and 10 seconds of the ECG.
- **ECG:** records 100, 101, 104, 107, 108 and 200 of the MIT Database.

**Run length encoding:** it used the consecutive zeros to form the output sequence. The wavelet coefficients of four levels of decomposition are the input of the encoding algorithm: the input (D) is composed of detail coefficients (d4, d3, d2, d1) and coarse coefficients (c4). The value of consecutives zero is assigned to the output.

**Huffman encoding:** two codebooks for the first and second level of decomposition were computed. Additionally, in the third level, one codebook was calculated to the coarse coefficients and other to the detail coefficients. Every codebook was composed by thirty-two codes. Because most of detail coefficients are zero after the output of the threshold block and the coarse coefficients are non-zero, two different kind of tree were estimated. For detail

---

Fig. 7. Huffman tree: detail (left) and coarse (right).
codebook, the average of all detail coefficients of the six signals was computed. The average signal is divided in thirty-two ranges and the Huffman code was estimated. The coarse codebook is calculated according to the distribution. The thirty-two codes are assigned according to the coarse amplitude: bigger amplitude has shorter code and smaller amplitude has longer code.

Fig. 8. Histograms: 101 (top); 104 (middle) and 107 (bottom).

The previous picture presents the histograms of the 104 and 107 records, which have different characteristics in time and frequency, but, significant similarities in its histograms. According to Figure 8, the detail coefficients (d1, d2, d3) have a histogram with a significant concentration in one bar; while, the coarse coefficients (c3) have the energy distributed in many amplitudes.

The length of the output stream is theoretically calculated as:

\[
\text{Length of output stream} = \sum_{i} \text{frequency}_i \times \text{length}_i
\]
Length = \sum (Pi*HCLi) \quad \text{for } i=1,\ldots,32 \quad (7)

\(Pi\) is the probability in the range I; HCLi is the Huffman Code Length in the range i.

The results of the 101 record with the base sym6 are presented in Figure 9. It obtained PRD=1.35 and CR=9.24 for run-length encoding and PRD=0.98 and CR=9.32 for Huffman encoding. The figure 10 presents the result for the base db6. Run-length encoding obtained PRD=1.1 and CR=9.11; while Huffman encoding obtained PRD=0.91 and CR=9.4.

Fig. 9. Record 101: case 1.
Although the record 104 has a pathological behavior opposed to the record 101, it is possible to compress the signal with the same codebook. According to the figure 11, the quality in the signal with RL encoding is better than Huffman encoding. But, the CR is better in the second method.

Fig. 10. Record 101: case 2.

Fig. 11. Record 104. Run length: PRD=0.901 and CR=8.78; Huffman: PRD=1.08 and CR=9.08.
The review results for the records 101, 104 and 107 are presented in table 5.

<table>
<thead>
<tr>
<th>Record</th>
<th>Method</th>
<th>CR</th>
<th>PRD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>db6</td>
<td>sym6</td>
</tr>
<tr>
<td>101</td>
<td>Huffman</td>
<td>9.32</td>
<td>9.28</td>
</tr>
<tr>
<td></td>
<td>RL</td>
<td>8.98</td>
<td>6.91</td>
</tr>
<tr>
<td>104</td>
<td>Huffman</td>
<td>8.98</td>
<td>9.06</td>
</tr>
<tr>
<td></td>
<td>RL</td>
<td>8.53</td>
<td>8.84</td>
</tr>
<tr>
<td>107</td>
<td>Huffman</td>
<td>7.62</td>
<td>7.78</td>
</tr>
<tr>
<td></td>
<td>RL</td>
<td>8.22</td>
<td>8.42</td>
</tr>
</tbody>
</table>

Table 5. Compression results

4. Hardware implementation

In previous works, the authors have developed strategies of hardware processing in the areas of filtering and compression. The first approach presented the Wavelet Transform architecture in decomposing process for denoising of electroencephalographic (EEG) signals (Gaona & Ballesteros, 2005). The second, it presented the comparison in hardware implementation between FIR and IIR (Corredor & Pedraza, 2009). The last project organized the above results in the architecture for biomedical compression based on Discrete Wavelet Transform (DWT) and Run Length encoding (Ballesteros et al., 2010).

4.1 Model of Discrete Wavelet Transform

The Discrete Wavelet Transform is composed by two stages: the convolution of the input signal by the wavelet base and the subsampling process. The convolution is performed by the equations:

\[
y_1[n] = \sum_{i=1}^{k=1,2,\ldots,m-1} h_i x[n-k]
\]

and

\[
y_2[n] = \sum_{i=1}^{k=1,2,\ldots,m-1} g_i x[n-k]
\]

In the above equations, \(y_1[n]\) and \(y_2[n]\) are the outputs of the FIR filters, \(h_i\) is the impulse response of the lowpass filter, \(g_i\) is the impulse response of the highpass filter and \(x[n]\) is the input (signal). The expression \([n-k]\) corresponds to the delay in the input. The value of \(m\) depends on the length of the wavelet base. For example, if the base is sym6, then \(m\) is equal to twelve.
The subsampling is calculated, according to:

\[ y_3[n] = y_1[2n] \]  
(11)

\[ y_4[n] = y_2[2n] \]  
(12)

To implement the equations (9) to (12), we propose the following blocks:

- **Bank of register**: This block stores in a \( m \)-position register the samples \( x[n] \) in an orderly way, depending on its arrival time. At the output is showed the current input value and the last \( m-1 \) inputs. If the reset signal is active, all the outputs are set to zero.

- **Coefficients Block**: Two \( m \)-positions memories keep the value of the wavelet filters: \( h_i \) and \( g_i \).

- **Control Unit**: It generates the control signal to select \( h_i \) or \( g_i \). Because the subsampling process eliminates the half of operations, it is most efficient to calculate only the half of operations, it means every two cycles of clock.

- **Multiplier / Adder**: It computes the mathematics operations of the equations (9)-(10) and (11)-(12).

In figure 12, \( m \) corresponds to the length of the FIR filter; \( sel \) is the bit of selection between the low-pass or high-pass filter and \( res \) is the number of bits of every input data.

**4.2 Model of thresholding and encoding**

To complete the compression model, two stages are added to the previous one: the thresholding and the encoding.

- **Thresholding block**: A comparison between a threshold value and the data from the low-pass or high-pass filter is doing. It computes the hard rule and the data with a lower value than the threshold are modified to zero.

- **Encoding block**: It performs the run-length method. At the output of this block there are two signals that go to a memory. The first signal indicates the position in the memory where the second signal ought to be stored.

**4.3 Results: Compression and filtering on FPGA**

In Figure 13 we present the RTL of the Multiplier/Adder for eight coefficients (\( m=8 \)). The eight multipliers and seven adders were modelled.
In Figure 14, the RTL of the thresholding stage is presented. The comparator is the main structure of the thresholding block. Its input is the output of the Multiplier/Adder block.

The architecture was modeled in a FPGA Spartan 3XC3S200. It obtained the following performance:
In this chapter we present the basic idea about the filtering and compression of biomedical signals with the Discrete Wavelet Transform. Although many works have been develop in this theme, we suggest a simple way to select the best parameters and we propose a novel static Huffman encoding for the compression of ECG signals.

For the younger researchers, the methodology in the selection of decomposition-thresholding parameters has been presented with a real case of SEMG signal. For the researchers with more experience in this area, we describe a novel method to generate static codebooks that it can be used in signals with different characteristics; this theme could be explored further.

Finally, the real time processing with the Discrete Wavelet Transform in filtering and compression of biomedical signals is conceived on FPGAs. The advantage of hardware solutions over software is the time of response which is lower in the first. The easy math of the DWT allows very rapid prototyping.

### 6. Acknowledgment

This work was supported in part by the University Military Nueva Granada under Grant ING290 and ING641.

### 7. References


The discrete wavelet transform (DWT) algorithms have a firm position in processing of signals in several areas of research and industry. As DWT provides both octave-scale frequency and spatial timing of the analyzed signal, it is constantly used to solve and treat more and more advanced problems. The present book: Discrete Wavelet Transforms - Biomedical Applications reviews the recent progress in discrete wavelet transform algorithms and applications. The book reviews the recent progress in DWT algorithms for biomedical applications. The book covers a wide range of architectures (e.g. lifting, shift invariance, multi-scale analysis) for constructing DWTs. The book chapters are organized into four major parts. Part I describes the progress in implementations of the DWT algorithms in biomedical signal analysis. Applications include compression and filtering of biomedical signals, DWT based selection of salient EEG frequency band, shift invariant DWTs for multiscale analysis and DWT assisted heart sound analysis. Part II addresses speech analysis, modeling and understanding of speech and speaker recognition. Part III focuses biosensor applications such as calibration of enzymatic sensors, multiscale analysis of wireless capsule endoscopy recordings, DWT assisted electronic nose analysis and optical fibre sensor analyses. Finally, Part IV describes DWT algorithms for tools in identification and diagnostics: identification based on hand geometry, identification of species groupings, object detection and tracking, DWT signatures and diagnostics for assessment of ICU agitation-sedation controllers and DWT based diagnostics of power transformers. The chapters of the present book consist of both tutorial and highly advanced material. Therefore, the book is intended to be a reference text for graduate students and researchers to obtain state-of-the-art knowledge on specific applications.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License, which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.