

Simulation of floating point generic DWT and IDWT for Denoising in DNA sequence for Exon Region Identification in genes

Yamini Rathore, Vikas Pathak, Harshit Mathur

Department of Electronics and Communication Engineering, Swami Keshvanand Institute of Technology, Management and Gramothan, Jaipur-302017(INDIA)

Email- yaminirtr93@gmail.com

Received 05.07.2018 received in revised form 16.08.2018, accepted 18.08.2018

Abstract:The discrete Wavelet transform is capable of providing the time and frequency information simultaneously, hence giving a time-frequency representation of the signal for which wavelet are discretely sampled. Some of the application of DWT are image processing, data compression, biomedical signal processing. In biomedical signal processing, DWT becomes a powerful technique. There are two architectures for DWT, one is convolution based and the other is lifting based. The lifting scheme provides many advantages, such as in-place implementation, fewer arithmetic operations and easy management of boundary extension. In lifting plan for Daubechies 9/7 channel, each lifting step comprises of one predict and one update step. It offers a higher quality of image restoration, higher coding efficiency. The VHDL code of proposed DWT/IDWT architecture is synthesized using Xilinx ISE 14.4 for FPGA Artix-7 family and simulated using Xilinx Isim simulator. The proposed VLSI architecture of Generic DWT and IDWT is used for denoising in filtered output DNA sequence for Exon region identification in Eukaryotic genes. The advantage of generic DWT/IDWT is that the design is input independent. In this design, the DWT module doesn't depend on number of input and reduces the logic unit in DWT which in turn reducing the chip area.

Keywords:DWT, IDWT, Radix-2, Lifting-Based, IEEE-754, Genomics, Protein Coding Region.

1. INTRODUCTION

Signals in the real world do not exist without noise, which may have negligible i.e. high signal to noise ratio (SNR) under certain condition. But in many cases, noise corrupts the signals, so it must be removed from the signal in order to process the data. Noise reduction is the process of removal of noise from the signal. Denoising is the process from which signal is reconstructed from noisy one. Noise can be random or white noise introduced by signal processing technique. The wide ranges of application of denoising are present such as data mining, radio astronomy, medical image/signal analysis [1]. In medical signal, noise removal needs special care, since denoising involves smoothing of a noisy signal (using low pass filter) may cause loss of fine detail.

Most of the practical signals are time-domain in their raw format i.e. is a function of time. But in many applications, the

most distinguished information is hidden in the frequency content of the signal. The frequency spectrum of a signal is basically the frequency components (spectral components) of that signal that shows what frequencies exist in the signal [2, 3]. The frequency is measured in cycles/second, or in "Hertz". Fourier Transform is the most popular transform being used. There are many other transforms that are Hilbert transform, short-time Fourier transform, Wigner distributions, the Radon Transform and the wavelet transform. Wavelet Transform provides time-frequency representation. There are two types of wavelet transform i.e. continuous wavelet transform (CWT) and discrete wavelet transform (DWT). CWT gives critical redundancy of reproduction whereas DWT gives the adequate data to both synthesis and analysis of the signal. The wavelet series is a sampled version of the CWT and the information it provides is highly redundant.

The input applied for DWT is of real data. So, to represent real value in the digital binary form, there are two possibilities, one is a fixed point and other is floating point representation. Fixed point can lead to a loss of precision when arithmetic operations are performed on two large numbers [4]. Floating point can give a precision to the scale of the number so it can represent a number from very small value to the very large value. It can support a wide range of values. The floating point adder and subtractor are designed and implemented based on single precision floating point (IEEE 754 standards). The floating point adder unit performs the addition and subtraction using substantially the same hardware used for floating point operation. The floating point adder algorithm mainly reduces the overall latency and improves the performance.

Section-2 gives the basic introduction of background of Floating point arithmetic, DWT algorithm and exon region of eukaryotic gene. Section-3 describes the hardware implementation of floating point adders and DWT algorithm. In section-4 various simulations and synthesis results of hardware and MATLAB simulation of lifting based DWT block and their use in identifying the exon region is discussed. Finally section-5 concludes the paper.

2. BACKGROUND

2.1 Floating point arithmetic

Floating-point computation is often found in systems which include very small and very large real numbers, which require fast processing times. Floating point arithmetic is used in embedded arithmetic processors, DSP processors, math coprocessor, and data processing units [5]. They are used where high numerical stability and accuracy is required. Various algorithms and design approaches have been developed by the Very Large Scale Integrated (VLSI) circuit community for implementation of floating point arithmetic.

The IEEE-754 single precision floating point standard has 8-bit exponent (with a bias of 127), a 23-bit mantissa and significant have a precision of 24 bits (7 decimal digits) [6]. The IEEE-754 double precision standard has 11-bit exponent (with a bias of 1023), a 52-bit mantissa and significant have a precision of 53 bits (16 decimal digits) as shown in Figure3.1.

2.2 The DWT Algorithm

Wavelet Transform provides time frequency representation. The wavelet transform has gained widespread acceptance in signal processing and image compression. Because of their inherent multi-resolution nature, wavelet-coding schemes are used in the applications where scalability and tolerable degradation are important [7]. There are two types of wavelet transform i.e. continuous wavelet transform (CWT) and discrete wavelet transform (DWT).

The wavelet series is a sampled version of the CWT and the information it provides is highly redundant. This redundancy requires a significant amount of computation time and resources. The discrete wavelet transform (DWT) provides sufficient information both for synthesis and analysis of the original signal, with a significant reduction in the computation time. It is easier to implement DWT as compared to CWT from hardware implementation point of view.

2.3 Exon Region Identification in Eukaryotic Gene

Any DNA sequence is divided into two parts Genes and intergenic spaces. Genes have two types of sub-regions called as the Exons and Introns. The Prokaryotic gene like bacteria do not have Introns but in Eukaryotic (which have a nucleus) gene both exons and introns are present. These exon regions are responsible for the generation of any protein from gene DNA sequence. Therefore, exons and introns are known as protein coding regions and non-coding regions respectively [8, 9]. Proteins are accountable for the biological function of any living organism. So to generate the proteins from DNA sequence these exon regions have to be identified. There are so many DSP techniques (like DFT, Filtering, CWT) to critically analyze this DNA sequence and to identify protein coding regions. The output of the filtering technique used for

exon region identification problem consists of noise. So this noise is to be removed from the filtered output. Hence DWT technique in this thesis is applied and verified for denoising of this filtered output.

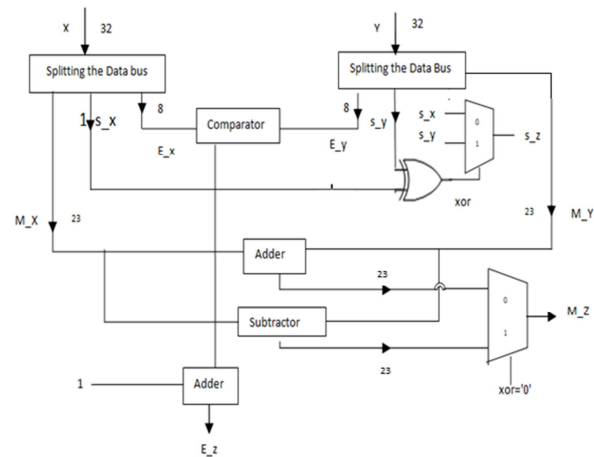
The character sequence of a DNA is converted into a numerical sequence using the EIIP method for applying the DSP techniques in order to analyze the DNA sequence as given in Table I.

Table 1 : EIIP Values for DNA sequence

Nucleotide	EIIP Values
A	0.1260
G	0.0806
T	0.1335
C	0.1340

3. HARDWARE IMPLEMENTATION OF FLOATING POINT LIFTING BASED DWT ALGORITHM

3.1 Floating point adder/subtractor



1 : Block diagram of proposed floating point adder

Figure

Two inputs X and Y of 32 bit is split into the sign (S_X, S_Y) of 1 bit, exponent (E_X, E_Y) of 8 bit and mantissa (M_X, M_Y) of 23 bit as shown in Figure1. In 23 bit comparator E_X and E_Y is compared and difference of E_Y and E_X is taken as 'd'. Xoring of S_X and S_Y is done and output obtained is taken as xor as shown in Figure 1. Xor is taken as a select line for the multiplexer, input to the multiplexer is S_X and S_Y and output is S_Z. When xor = '0', then S_Z = S_X else S_Z = S_Y. M_X and M_Y of 23 bit are given to 23 bit binary adder. Similarly, M_X and M_Y are applied to 23 bit binary subtractor. The output of floating point adder and subtractor are the input of the multiplexer, xor, as a select line and output obtained, is stored in M_Z. By concatenating S_Z, E_Z, M_Z the result of floating point adder/subtractor is obtained.

3.2 Implementation of Generic Discrete Wavelet Transform (DWT)

The 2-level DWT is implemented in this thesis which consists of two predict module and two update module. The VHDL code was synthesized using Xilinx ISE design suite 14.4 tools and simulated using Isim simulator. The generic DWT is designed by using FSM (Finite State Machine) which is input independent. In this design, the DWT module doesn't depend on a number of input and reduces the logic unit in DWT which in turn reducing the chip area.

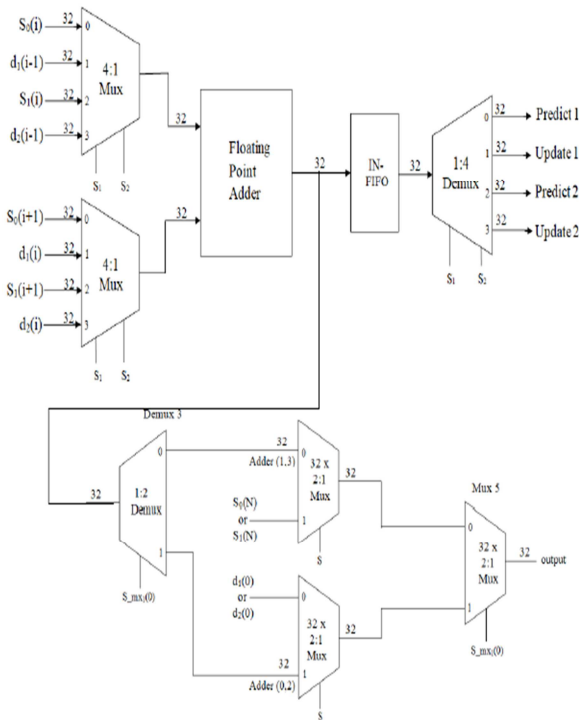


Figure 2: Implementation Block of Generic DWT using FSM

Serially adding Detail coefficient and Approximate coefficient floating point adder is used. When select line of multiplexer is $s1s2 = "00"$ then addition of $S0(i)$ and $S0(i+1)$ is done and output is stored in in-fifo else when select line is "01" then $d1(i-1)$ and $d1(i)$ is added and stored in-fifo and before storing resetting of in-fifo is done. Similarly, $S1(i)$ and $S1(i+1)$, $d2(i-1)$ and $d2(i)$ are added in the floating point adder depending on the select line of multiplexer and output is stored in in-fifo which is of 32 bit as shown in Figure 2.

After addition 4:1 demultiplexer is used, when the select line of demux is "00" then result of addition act as the input to predict1 fifo. When select line $s1s2 = "01"$ the result of in-fifo is used by update1 fifo as an input. In a similar way, the input is supplied to the predict2 and update2 fifo based on the select line of the demultiplexer. In some cases, adder result and some cases directly approximate or detail coefficient i.e. coefficient of predict1 and update1 are applied to the fifo as an input. So, multiplexer and demultiplexer based on the select are used to select between them. Serially, detail and approximate coefficient i.e. coefficient of predict2 and update2 are

calculated for 2-level DWT. After 2-level decomposition, scaling is done for predict2 and update2 coefficient and then, the final approximate and detail coefficient is obtained. This Architecture has 2-level decomposition, detail and approximate coefficient are decomposed two times in DWT.

3.3 Implementation of Generic Inverse Discrete Wavelet Transform (IDWT)

The 2-level IDWT is implemented in this thesis which consists of two predict module and two update module. The VHDL code was synthesized using Xilinx ISE design suite 14.4 tools and simulated using Isim simulator. The generic IDWT is designed by using FSM (Finite State Machine) i.e. it is an abstract machine that can be in exactly one of the finite number of states at any given time and designed architecture is input independent which means it doesn't depend on a number of input.

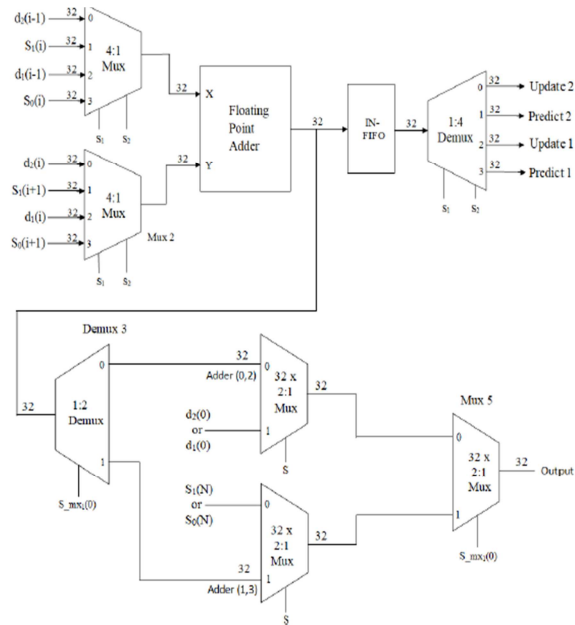


Figure 3: Implementation Block of generic IDWT using FSM

The outputs of DWT act as an input to IDWT Block. The Detail coefficient and Approximate coefficient which is the output of DWT is added serially using a floating point adder. When select line of multiplexer is $s1s2 = "00"$ then addition of $d2(i-1)$ and $d2(i)$ is done and output is stored in in-fifo else when select line is "01" then $S1(i)$ and $S1(i+1)$ is added and stored in-fifo and before storing resetting of in-fifo is done. Similarly, $d1(i-1)$ and $d1(i)$, $S0(i)$ and $S0(i+1)$ are added in the floating point adder depending on the select line of multiplexer and output is stored in in-fifo which is of 32 bit as shown in Figure 3.

After addition 4:1 demultiplexer is used, when the select line of demux is "00" then result of addition act as the input to update2 fifo. When select line $s1s2 = "01"$ the result of in-fifo is used by predict2 fifo as an input. In a similar way, the input is supplied to the predict1 and update1 fifo based on the select

line of the demultiplexer. In some cases, adder result and some cases directly approximate or detail coefficient i.e. coefficient of predict2 and update2 is applied to the fifo as an input. So, multiplexer and demultiplexer based on the select are used to select between them. Serially, detail and approximate coefficient i.e. coefficient of predict1 and update1 are calculated for 2-level IDWT. This Architecture has 2-level decomposition, detail and approximate coefficient are decomposed two times in IDWT.

4. RESULTS AND DISCUSSION

4.1 Methodology

The methodology adopted shown in Figure4 is as follows:

- One dimensional DNA sequence input is applied to MATLAB which is taken from the database. From last 5 years, database of DNA sequence has grown exponentially. Entrez is the search tool for the National Center for Biotechnology Information (NCBI) database. The search starts with a relevant group of database such as nucleotide, protein, etc. Entrez includes GenBank, RefSeq, and PDB. All publically available DNA sequence and the amino acid sequence are present in GenBank. The DNA sequence is in character sequence that is converted into numerical values (real values). The conversion is done using Electron ion interaction potential (EIIP) method. EIIP is the unique number used to represent each amino acid or nucleotide. Analysis of numerical series obtained from the EIIP method is done by digital signal analysis (DWT) methods in order to extract information relevant to the biological function. The real value obtained from the EIIP method is then filtered by an anti-notch filter (small bandpass filter) but after filtering noise is present in the signal.

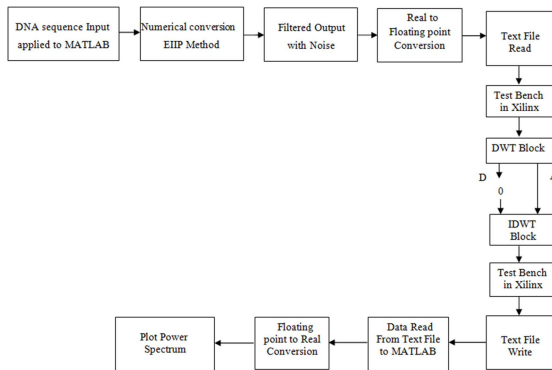


Figure 4: Flow chart of Methodology Adopted for identification of exon region

- In MATLAB, filtered output is converted into floating point numbers. The text file is created by MATLAB contains a floating point number that is given as an input to the DWT block. To perform the floating point addition operation in Xilinx, the implementation of floating point adder is done. Single precision Arithmetic adder/subtractor based on IEEE-754 standards is designed. Floating point

adder/subtractor includes implementation of the priority encoder, left shifter and right shifter.

- Using VHDL Test Bench, DWT block read the Text file created by MATLAB. A test bench is a non-synthesizable VHDL file which compares the output with the expected output. If there is any mismatch, an error is displayed in the VHDL simulator’s log.
- DWT block is implemented by using Lifting based scheme. The output of DWT is write in a text file which contains approximate coefficient (A) and detail coefficient (D). Generic DWT is designed using predict and update which is input independent. The VHDL code was synthesized using Xilinx ISE design suite 14.4 tools and simulated using Isim simulator.
- The text file is again created by Xilinx for the output of DWT Block. The result of DWT contains detailed coefficient and approximate coefficient. By equating the detail coefficient to zero (D = '0') which filter out high pass filter coefficient of Lifting based DWT algorithm and passing approximate coefficient (A) to the input of IDWT block. Text file created for the output of IDWT block is read by MATLAB. Interfacing between Xilinx and MatLab is done through test bench.
- Now MATLAB converts the floating point data which is obtained from the IDWT block is converted into real value. The results of the implementation of DWT and IDWT algorithm is verified with MATLAB for Exon Region Identification in a Eukaryotic gene by plotting the power spectrum.

4.2 Floating point adder

Name	Value	1,999,995 ps	1,999,996 ps	1,999,997 ps	1,999,998 ps	1,999,999 ps
x[31:0]	435a2c44			435a2c44		
y[31:0]	43256322			43256322		
z[31:0]	43bfc7b3			43bfc7b3		
a[31:0]	43bfc7b3			43bfc7b3		
ed[7:0]	00			00		
ex[7:0]	86			86		
ey[7:0]	86			86		
ez[7:0]	87			87		
ey[17:0]	UU			UU		
qx[7:0]	86			86		
qx[7:0]	86			86		
mx[22:0]	5a2c44			5a2c44		
my[22:0]	256322			256322		
mz[22:0]	3fc7b3			3fc7b3		
sg[23:0]	7f8f66			7f8f66		
sg_x[23:0]	da2c44			da2c44		

Figure 5 : Simulation result of floating point adder

Floating point adder performs both addition and subtraction which uses the same hardware for floating point operation. The simulation result of floating point adder has two input of 32 bit represented by x and y and the result of the addition is stored in z which is also of 32 bit as shown in Figure 5. Others are intermediate signal.

4.3 DWT Block

After implementation of 8-input DWT which depends on input, the generic DWT is designed which is input independent. The input of DWT block is denoted by 'inp' of 32 bit. The output of DWT Block is y_s that is approximate coefficient and y_d is detailed coefficient as shown in Figure 6. Other are intermediate signal.

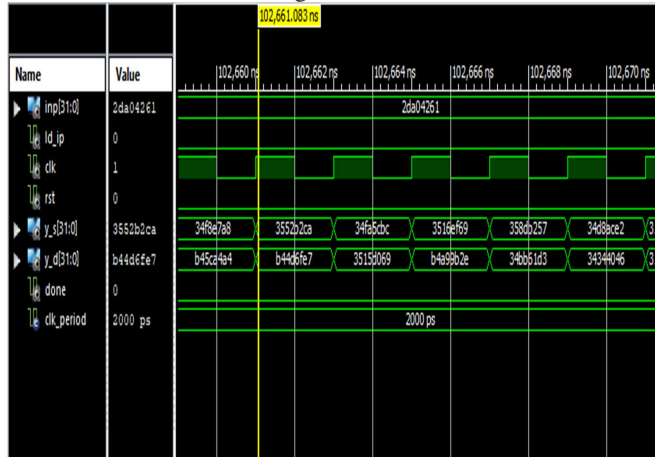


Figure 6 : Simulation Result of generic DWT Block

4.4 IDWT Block

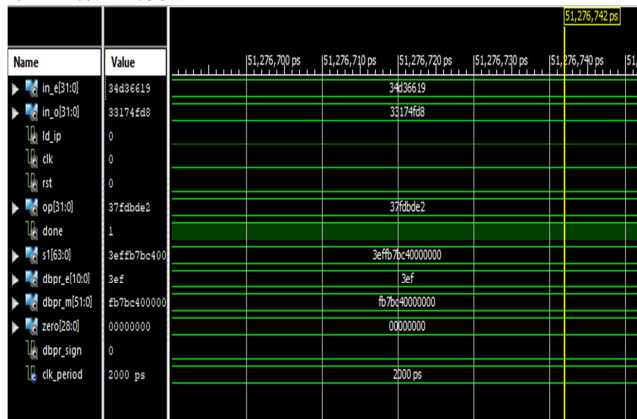


Figure 7 : Simulation Result of generic IDWT Block

Two input is present in the IDWT module in_o and in_e of 32 bit. The output of IDWT is represented by 'op' of 32 bit. It is generic in which IDWT is input independent. Other signals are Intermediate Signal As Shown in Fig 7.

4.5 Verification Of Proposed Design for Denoising of Filtered Output of Exon Region Identification of gene

Figure8 (a) represent Filtered output of gene AJ223321. The plot is between power spectral density and relative position of sequence containing noise. Figure8 (b) represents denoising of filtered output after applying DWT in MATLAB. Figure8 (c) represents denoising of filtered output after applying proposed DWT architecture in Xilinx. Hence, result

is verified for proposed design for denoising of filtered output for exon region identification of gene AJ223321.

5 CONCLUSION

The lifting based 1-D DWT and IDWT Algorithm for 32-bit single precision floating point data is proposed in this dissertation. The VHDL code of proposed DWT/IDWT architecture is synthesized using Xilinx ISE 14.4 for FPGA Artix-7 family and simulated using Xilinx Isim simulator. The proposed VLSI architecture of Generic DWT and IDWT is used for denoising in filtered output DNA sequence for Exon region identification in Eukaryotic genes.

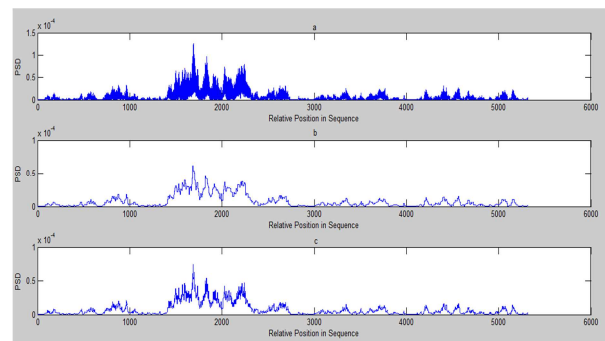


Figure 8 : Power spectrum output of gene AJ223321 for exon region identification (a) Filtered O/P (b) Denoising using MATLAB DWT (c) Denoising using proposed DWT architecture

REFERENCES

- [1] Abhinav V. Deshpande, "VLSI Implementation of Discrete Wavelet Transform (DWT) for Image Compression", IOSR Journal of Electronics and Communication Engineering, Volume 11, Issue 3, Ver. III, pp. 42-45, May-Jun .2016
- [2] Marghny H. Mohamed, Saad Z. Rida and Basma Ahmed A., "Security based Watermarking Algorithm Based on DNA Sequence Using DWT-SVD", An International Journal of Information Sciences Letters, Vol.2, No.1, pp. 1-6, 2005
- [3] Ronald W. Lindsay, Donald B. Percival, and D. Andrew Rothrock, "The Discrete Wavelet Transform and the Scale Analysis of the Surface Properties of Sea Ice", IEEE transactions on Geoscience and Remote Sensing, vol. 34, no. 3, pp. 771-787, May 1996
- [4] Preethi Sudha Gollamudi and M. Kamaraju, "Design Of High Performance IEEE- 754 Single Precision (32 bit) Floating Point Adder Using VHDL", International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 7, pp. 2264-2275, July – 2013
- [5] Najib Ghatte , Shilpa Patil and Deepak Bhoir, "Floating Point Engine using VHDL", International Journal of Engineering Trends and Technology (IJETT), Volume 8, Number 4, pp. 198-203, Feb 2014
- [6] Khushbu Naik and Tarun Lad, "Implementation of IEEE 32 Bit Single Precision Floating Point Addition and Subtraction", International Journal of Computer Application, Volume 5, No. 3, pp. 107-111, April 2015
- [7] Sachin D Ruikar and Dharmal D Doye, "Wavelet Based Image Denoising Technique", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 2, No.3,pp. 49-53, March 2011
- [8] Taegeun Park, Juyoung Kim and Junye Rho, "low-power, low-complexity bit-serial VLSI architecture for 1D Discrete Wavelet Transform, circuits systems signal processing, vol. 26, no. 5, pp. 619–634, 2007
- [9] A. Prochazka, J. Ptacek and I. Sindelarova, "wavelet transform in signal and image restoration", An International Journal of Information Sciences Letters, vol. 5, pp.1-5, 2004