FPGA Realization of Lifting Based Forward Discrete Wavelet Transform for JPEG 2000

M.S.Bhuyan, Nowshad Amin, Md.Azrul Hasni Madesa , and Md.Shabiul Islam

***Abstract*—**This paper describes the hardware design flow of lifting based 2-D Forward Discrete Wavelet Transform (FDWT) processor for JPEG 2000. In order to build high quality image of JPEG 2000 codec, an effective 2-D FDWT algorithm has been performed on input image file to get the decomposed image coefficients. The Lifting Scheme reduces the number of operations execution steps to almost one-half of those needed with a conventional convolution approach. In addition, the Lifting Scheme is amenable to "in-place" computation, so that the FDWT can be implemented in low memory systems. Initially, the lifting based 2-D FDWT algorithm has been developed using Matlab. The developed codes are then translated into behavioral level of FDWT algorithm in VHDL. The FDWT modules were simulated, synthesized, and optimized using Altera design tools. The final design was verified with VHDL test benches and Matlab image processing tools. Comparison of simulation results between Matlab and VHDL was done to verify the proper functionality of the developed module. The motivation in designing the hardware modules of the FDWT was to reduce its complexity, enhance its performance and to make it suitable development on a reconfigurable FPGA based platform for VLSI implementation. Results of the decomposition for test image validate the design. The entire system runs at 215 MHz clock frequency and reaches a speed performance suitable for several real-time applications.

***Keywords*—**VLSI, DWT, Lifting, JPEG 2000, Synthesis**.**

I. Introduction

The importance of visual communications has increased tremendously in the last few decades. The development of new technologies and communication networks creates new needs and stimulates the introduction of new functionalities. The current standards in the field of still image coding are inadequate for producing the best quality of performance. To address this concern, ISO committee came up with a new coding system, JPEG 2000 [1]. JPEG 2000 is intended to provide subjective image quality performance superior to other existing standard image file formats viz. JPEG, BMP, and GIF etc. In addition, JPEG 2000 includes many modern features such as lossless to lossy coding with the same algorithm, scalability, etc. This new ISO/ITU-T standard has shown to provide superior coding efficiency to the previous standards. The techniques enabling all new features of JPEG 2000 are a Discrete Wavelet Transform (DWT) followed by an arithmetic coding [1]. The full-frame nature of DWT de-correlates the image over a larger scale and eliminates blocking artifacts at high compression ratios. The advantage of DWT over Fourier Transform is DWT performs multi-resolution analysis of signal with localization in both time and frequency. The DWT decomposes a digital image into different sub bands so that the lower frequency sub bands have finer frequency resolution and coarser time resolution compared to the higher frequency sub bands. The DWT is being increasingly used for image compression due to the fact that the DWT supports features like progressive image transformation (by quality and resolution) ease of compressed image manipulation, region of interest coding etc. The interest in DWT is growing tremendously in recent years due to its adoptions in JPEG 2000 and MPEG-4 [1]. The multi-resolution representation derived from DWT time-frequency decomposition demonstrates extraordinary advantages in signal analysis and compression widely used in image processing, communication and robotics. As a result, efficient VLSI implementations of DWT processor become more and more important recent years.

Since the discovery of DWT by Mallat [2], several architectures for DWT processor have been proposed. The most common one is the filter-bank implementation [2]. Recently an alternative implementation has been proposed, known as the Lifting Scheme (LS) [2]. In addition to providing, a significant reduction in memory uses and computational complexity, lifting provides "in-place" computation of the wavelet coefficients by overwriting the memory locations that contain the input sample values. The wavelet coefficients calculated by lifting are identical to those computed by direct filter-bank convolution. Consequently, the number of multiplications and additions compared to the filter-bank approach are reduced, resulting in a more efficient use of power and chip area. Its modular structure is well suitable for hardware implementation. Because of these advantages, the specification of the DWT kernels in JPEG 2000 is only provided in terms of the lifting coefficients [3].

Manuscript received March 13, 2007; Revised July 2, 2007.

M.S.Bhuyan is with the Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Malaysia (corresponding author : +603-8312 5380; e-mail: shakir_dhaka@yahoo.com).

Nowshad Amin., is with the Department of Electrical, Electronic and System Engineering, Faculty of Engineering, National University of Malaysia, 43600 UKM, Bangi, Selangor, Malaysia. (e-mail: nowshad@vlsi.eng.ukm.my).

Md.Azrul Hasni Madesa , is with the Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Malaysia .

Md.Shabiul Islam, is with the Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Malaysia, (e-mail: shabiul@mmu.edu.my).

Since the emergence of the JPEG 2000 considerable attention has been paid to the development of efficient system architectures of the DWT. FPGA implementations can accelerate DWT by pipelining these operations. Entropy encoding, however, is more difficult to optimize due to its bit-serial structure. Several VLSI architectures based on the DWT [4, 5] have been designed and implemented in order to achieve real-time signal processing [5]. Knowles [6] first proposed a systolic array-based architecture for the 1-D DWT without multiplier, for computing the four-tap Daubechies 2-D DWT was further modified by Lewis and Knowles [7]. However, these architectures need lots of control signals to complete the transform computations. Although these architectures are good for the Daubechies 2-D DWT, they are not suitable for other wavelets. Then, Parhi and Nishitani [8] proposed the folded and digital-serial architectures for DWT. The folded architecture can reduce the latency, but the hardware utilization is poor. The digit-serial architecture requires many registers, and longer latency. Based on the recursive pyramid algorithm, Vishwanath et al. [9] proposed systolic-parallel architecture for the 2-D DWT. However, their architecture cannot provide intact octave band components for perfect reconstruction (PR). Chen [10] proposed a multilevel lifting-based wavelet transform architecture. Chen's architecture takes only half the time for the computation of DWT when compared with other designs, but it still works under a large clock period. An efficient design takes into account aspects such as area, power consumption, throughput, etc. Techniques such as pipelining, distributed arithmetic, etc. help in achieving these requirements. The present paper proposes the FDWT architecture for large throughput and optimized area. Moreover, it is parallel in nature. The design is scalable in terms of number of decomposition levels and image size. To improve the computation we removed area-consuming multipliers from the present realization. Upon the recommendations made by the JPEG-2000 standard committee, the 5/3 Le Gall wavelet filter was used as the wavelet family of choice. In this paper, the focus is on the development of the hardware architecture and the corresponding VHDL model of a JPEG 2000 based IDWT module.

Fig.1 shows an input image is forward transformed before entropy coded in JPEG 2000. To decode the original image, the inverse transform is applied in exactly the reverse order. For simplicity, we have only described the design details of the 2-D FDWT module in this paper. The focus is on the development of the hardware architecture and the corresponding VHDL models. The JPEG 2000 standard committee has recommended using these wavelet filters for integer mode operation [1].

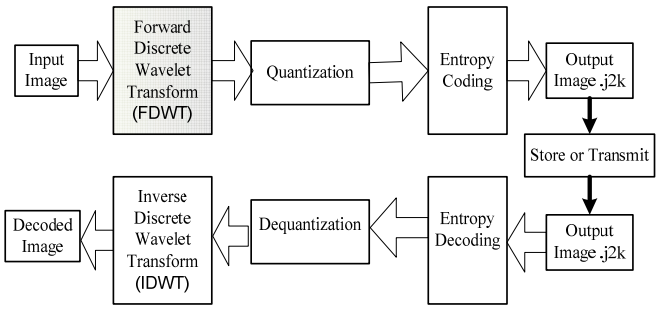


Fig.1 building block diagram of JPEG-2000

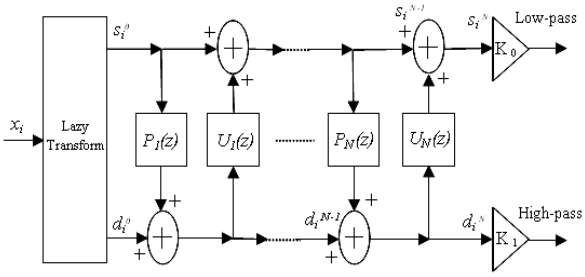


Fig. 2 general diagram of the lifting process

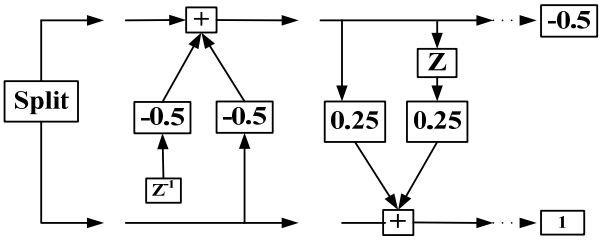


Fig. 3 implementation of Le Gall wavelet filters

II. Discrete Wavelet Transform and Lifting Scheme

The DWT has been traditionally implemented by means of the Mallat filter bank scheme [5] that includes two main steps: signal decimating and filtering with a pair of Quadrate Mirror Filters (QMFs). The filter-bank can be realized using FIR filters. The process consists of performing a series of dot products between the two filter masks and the signal.

A. Lifting Scheme realization

Sweldens [2] proposed the LS where all the operations can be performed in parallel, hence the possibility of a fast implementation. The LS is composed of mainly three steps, namely Split, Predict (P), and Update (U). Fig. 2 describes the LS process. The first step is splitting the input signal x_i into even and odd indexed samples. Then we try to predict the odd samples based on the evens. If the original signal has local correlation, then the prediction should be of high accuracy. For designing the FDWT module, we have chosen the LS approach. Fig. 3 represents the basic building block of the 1-D FDWT using the Le Gall wavelet filters. The LS based

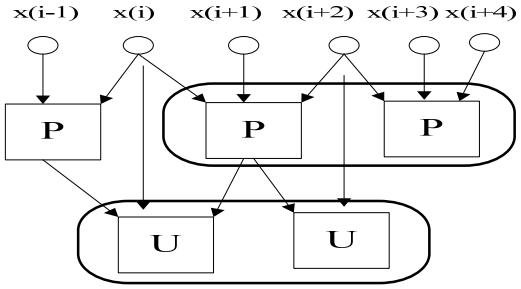


Fig. 4 parallel processing of input data samples

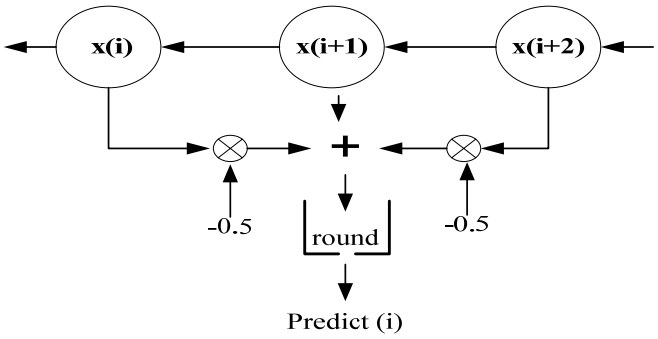


Fig.5 (a) predict filter modules in FDWT

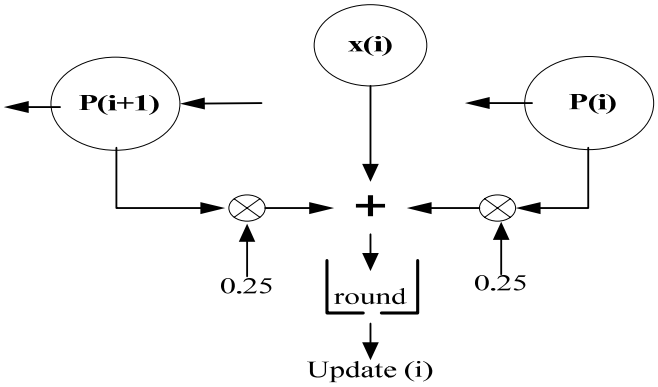


Fig.5 (b) update Filter modules in FDWT

realization allows Integer Wavelet Transform (IWT). The transform coefficients of the IWT are exactly represented by finite precision numbers, thus allowing for truly lossless encoding. This helped us in reducing number of bits for the sample storage and to use simpler filtering units. IWT is achieved by rounding off the output of filters, before addition or subtraction. This led to a very beneficial class of transforms, in terms of computational complexity and memory requirements. Yet, they are highly dependant on the choice of the factorization of the polyphase matrix [5]. Equation (1) shows the LS for the FDWT design.

$$y_{2i+1} = \lfloor -0.5(x_{2i} + x_{2i+1}) \rfloor + x_{2i+1}$$
$$y_{2i} = \lfloor 0.25(y_{2i+1} + y_{2i+3}) \rfloor + x_{2i} \ , where..0 \le i < N/2 \quad (1)$$

III. Design Architecture of the FDWT

In our design a scalable LS based FDWT module has been implemented in FPGA. Design acceleration has been achieved through parallel processing of independent sub-modules, re-usability of image pixel data. Input pixels are accessed through a four sample register (temporary storage), to activate two concurrent predict and update as shown in Fig.4. Registers were used for temporary storage, and reusability of temporary data. Fig 5 (a) and (b) show the block diagram for the predict and update modules respectively. A fixed precision of 8-bits/ per pixels were selected. The error introduced by this precision has been proved, through comparing software and hardware implementations, to be within an accepted range. Since all coefficients are multiplies of 2, all multiplications and divisions were replaced by shifting operations.

Starting with the specifications outlined in [1], the DWT module is supposed to have following parameters: picture width, picture height, levels of decomposition and mode of operation. Since we limited our implementation to only one wavelet family, thus we will have only one mode of operation. Our module has parameterized with picture width 512, height 512, and levels of decomposition (maximum levels 7).

TABLE 1
TRANSFORMED COEFFICIENTS OF FIRST LEVEL

Approximation Coefficients 1st Level (Column)	Detail Coefficients 1st Level (Column)	Approximation Coefficients 1st Level (Row)	Detail Coefficients 1st Level (Row)
0	0	0	0
156	6	117	-156
158	-3	120	-160
156	6	120	-159
154	0	119	-158
155	3	120	-160
154	0	117	-155
159	3	119	-158
156	1	119	-158
154	-3	121	-161

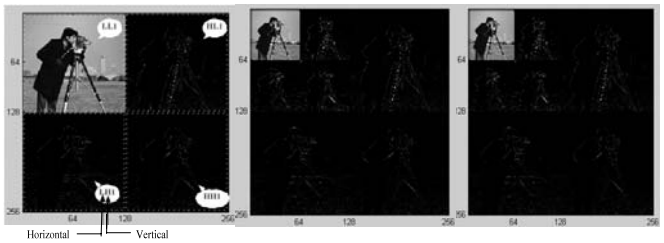
IV. Matlab Simulation

The approximation and the detailed coefficients of test image were computed first in Matlab platform. We calculated approximation and detailed Coefficients for all the rows of the 256 × 256 input image referred as the 1-D FDWT transform

Next, the same Matlab routine was implemented on this coefficients column wise to obtain the 2-D FDWT coefficients. By this consecutive row and column wise operation on the input image data, we get the level-1 decomposition coefficients. Table 1 show few example coefficients extracted from Matlab simulation.

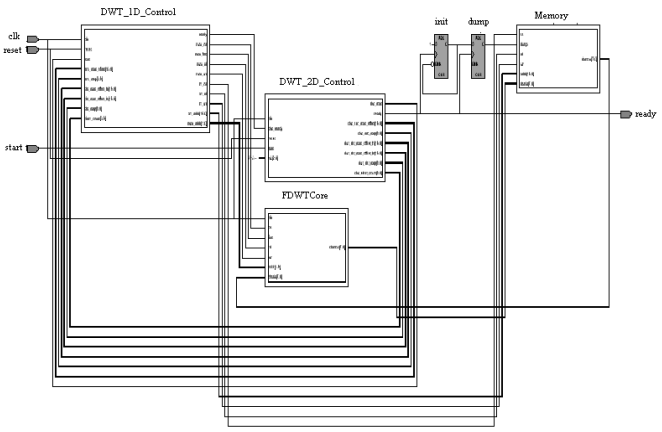
After level-1 decomposition, approximation and the detailed coefficients were assembled together for the level-2 decomposition. Fig. 6 (a) shows the image of level-1 transform coefficients after the assembling process. The same Matlab routine was again applied on the lowest-frequency sub band (LL_1) to get the level-2 approximation and detailed

coefficients shown in Fig. 6 (b). Matlab routine continued to get the level-3 transformation shown in Fig. 6 (c)



(a) (b) (c)

Fig. 6 Matlab simulation (a) level-1, (b) level-2, and (c) level-3 coefficients extracted images



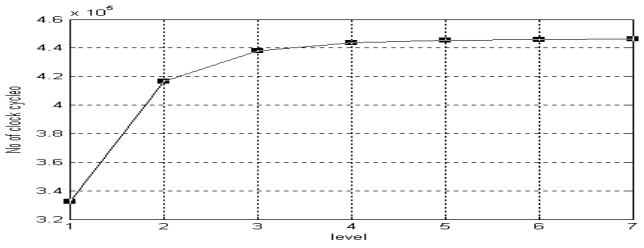
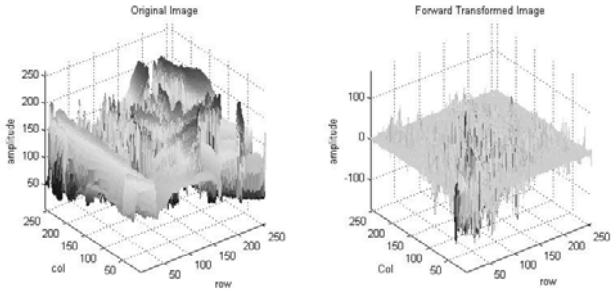


Fig. 8 clock cycles required for various levels



(a) (b)

Fig. 9 plot showing (a) pixel values for the original input image and (b) transformed output resulting from a level-3 FDWT decomposition of the input image

V. Modeling 2-D DWT in VHDL

The developed 2-D FDWT module is composed of four major building blocks: DWT_1D_Control, DWT_2D_control, FDWTCore, and Memory. Detailed architecture is shown in Fig. 7. DWT_1D_Control module performs the 1-D transform row/column wise. After all the rows of the image are transformed, two dimensional controller initiates the transformation process in column wise thus complete the level-1 transformation. Level of computation is controlled by DWT_2D_control module through parameter (NL) signal. The memory consists of only one process. First, an array for the input data and the transformed coefficients are created. Input image data are initialized by the memory read (rd) signal for simulation purposes. The memory is able to process a request only if the wr (*write*) or the rd (read) signal is active (high). The incoming addr signal (address*)* is converted into an integer value for accessing the memory array. The FDWTCore block does the actual computation on the image data. Four input and two output registers are used to hold four input data and two (approximation and detail) output data simultaneously. The data are read sequentially from the memory. Computations are based on equation (1). The higher the level of computation (*nNL*), the higher the number of pixels to be processed thereby requiring increased number of computations. For example, when the level of computation is two, computations are first done for all 65536 pixels 256 × 256 in the original block corresponding to level-1 computation and the results are written to the memory. Then computations are done for all the 1024 pixels 128 × 128 in the LL_1 block of the transformed image. Therefore, the total number of computations in this case is 81920. For each computation level, pixel values are first read in row-by-row fashion. This continues until all pixel values from all rows are read and the transformed coefficients are stored in the memory.

VI. Functional and Timing Simulation of the Design

Simulation is critical in verifying developed design behavior. Functional and timing simulation of the FDWT design was done with Mentor Graphics ModelSim-Altera using developed test bench and appropriate stimuli to validate the design.

The number of clock cycles required for various levels of computation is shown in Fig. 8. These are the number of cycles starting from the time the start signal is asserted until the ready signal is issued. Clearly increasing the number of computation levels increases the number of clock cycles required for performing the tasks Fig 9(a) shows a three-dimensional plot of the pixel intensity values for the input image. It is important to note that the image energy is

distributed across the pixel array. Fig 9(b) shows input image after level-3 FDWT transformation. Note that, the *x* and *y-axis* of the plot correspond to the rows and columns of the 256 × 256 output array. The graph shows that the signal energy from

TABLE 2
COEFFICIENTS OF MATLAB VS. VHDL 1-D FDWT

Coeff. No	Matlab Test1	VHDL Test1	Matlab Test2	VHDL Test2	Matlab Test3	VHDL Test3
1	30	31	30	30	25	23
2	29	29	31	30	28	29
3	29	31	32	32	30	28
4	0	28	0	32	0	23
5	1	0	-1	1	-3	0
6	-0	-1	-0	0	2	0
7	2	0	0	0	1	1
8	0	0	-1	-2	-29	-30

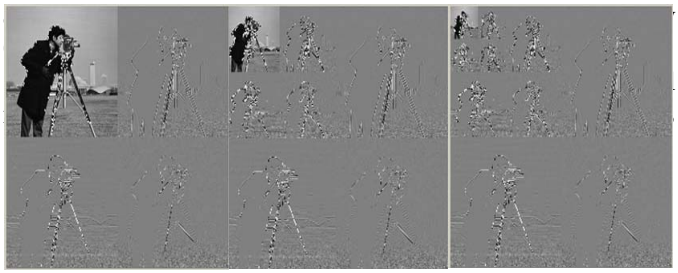
the original image has been decorrelated and then concentrated in a much smaller region corresponding to LL_3 sub band as expected. This concentration of signal energy shows how large compression ratios can be achieved by removing the extraneous information contained in the less-important sub bands. Thus, only small portions of coefficients are needed to encode for compression.

VII. COMPARISON OF SIMULATION RESULTS

The coefficients extracted from different levels of decomposition in Matlab ware compared with the VHDL simulation. First, we did the comparison for 1-D case. Eight arbitrary pixel data were fed into the 1-D Matlab module. The same data were fed into the 1-D VHDL module. The resulting coefficients (decimal) for three different pixel data sets (Test 1, Test 2, Test 3) shown in Table 2. We can observe the Matlab and the VHDL coefficients for the approximation and the details coefficients are almost the same. Then, we perform the comparison between the coefficients for the complete 2-D FDWT module. The input image consisting of 256 × 256 pixels were fed into the VHDL module. The approximation and the detailed coefficients for three different levels of decomposition were compared with the corresponding level of decomposition of the Matlab platform. Fig.10 (a), (b), and (c) were constructed from the 1st, 2nd, and 3rd level of decomposition respectively. Fig 11 shows the difference between software (SW) and hardware (HW) execution time where SW is ten times slower than the HW execution time.

VIII. IMPLEMENTATION RESULTS WITH PERFORMANCE COMPARISON

Quartus II Integrated Synthesis (QIS) tool was used to synthesize the FDWT design codes into gate- level schematic. QIS includes advanced synthesis options and compiler directives (attributes) to guide the synthesis process to achieve optimal results. Fig 12 shows the schematic view of the FDWTCore module of the 2-D FDWT. Fig 13 shows slice of



(a) (b) (c)

Fig.10 Forward Transform (a) Level-1, (b) Level-2, and (c) Level-3 in VHDL simulation

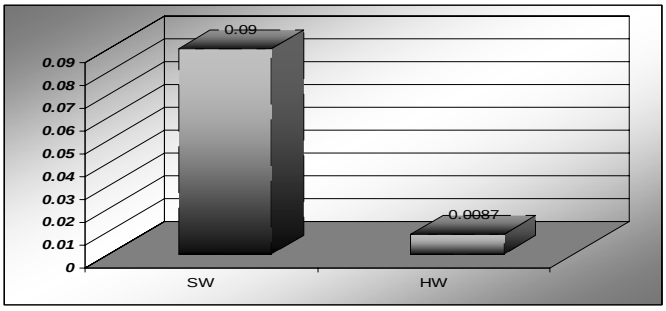


Fig.11 Matlab vs. VHDL on execution time

a performance comparison in terms of clock speed with other existing works reported in the corresponding references. Clearly, our implementation scheme outperforms other implementations.

IX. CONCLUSION

The hardware architecture of a 2-D FDWT processor using 256 × 256 pixels block of 8-bit image information as input has been presented. This architecture employs no multiplier and therefore is an attractive alternative to processors employing computationally expensive multipliers. The processor has been modeled in VHDL and validated using simulation and in-system debugging. The present model is a foundation for development of a comprehensive modeling framework for DWT processors starting from abstract high-level system models to synthesizable models [17].

REFERENCES

[1] ISO/IEC.ISO/IEC.JPEG 2000 coding system.
[2] Sweldens, The lifting scheme: A new philosophy in biorthogonal wavelet constructions', Proceedings of SPIE, pp.68-79, 1995.
[3] ISO/IEC.JTC/SC29/WG11Generic Coding of Audio Visual Objects: Final Draft, 1998
[4] C. Lambert-Nebout, G. Moury and J.-E. Blamont, Status of onboard image compression for CNES space missions, Proc. SPIE'99 3808 (October 1999), pp. 242–256.

[5] MALLAT, S.G.: 'A theory for multiresolution signal decomposition: The wavelet representation', IEEE Trans. Pat. Anal. Mach. Intell.1989, 11, (7), pp. 674-693.
[6] KNOWLES, G.: 'VLSI architecture for the discrete wavelet transform', Electvon. Lett., 1990, 24, (15), pp. 1184-1185
[7] LEWIS, AS., and KNOWLES, G.: 'VLSI architecture for 2D Daubechies wavelet transforms without multipliers', Electron. Letter. 1991, 27,
[8] PARHI, K , and NISHITANI, T.: 'VLSI architecture for discrete wavelet transforms', IEEE Trans. VLSI Syst., 1993, 1, (2), pp. 191-202
[9] VISHWANATH, M., MICHAEL, R.M , and IRWIN, M J.: 'VLSI architecture for the discrete wavelet transform', IEEE Trans. Circuits Syst. -11
[10] P. Y Chen, "VLSI implementation for one-dimensional multilevel lifting-based wavelet transform," IEEE Trans. on Computers, vol. 53, no. 4, pp. 386-398, April 2004.
[11] http://www.barco.com/designservices/Downloads/IPProducts/BA114IDWTFactSheet.pdf.
[12] http://ieeexplore.ieee.org/xpls/abs_all.jsp?isnumber=33117&arnu mber=1557049&count=534&index=185.
[13] http://www.ece.uvic.ca/~nagarwal/resume.html#Related%20Experience %20And%20Courses.
[14] http://portal.acm.org/citation.cfm?id=1057900.
[15] http://portal.acm.org/citation.cfm?id=1062117.
[16] http://portal.acm.org/citation.cfm?id=1169770&jmp=abstract&coll=G UIDE&dl=GUIDE&CFID=15151515&CFTOKEN=6184618.
[17] M.S. Bhuyan, Md. Shabiul Islam, Md.Azrul Hasni Madesa, and Masuri Othman. Design of Lifting Based 2-D Discrete Wavelet Transform in VHDL, ISBN: 978-983-42747-7-7, page 224, 2007.

M.S. Bhuyan obtained his B.Sc. Engg (Hons) from Ahsanullah University of Science and Technology, Dhaka, Bangladesh in Computer Science and Engineering, 2002. He is currently working toward the MEngSc degree at the Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia. His research interests are in the area of VLSI implementation of image processing and video.

Nowshad Amin was born in Chittagong, Bangladesh, on January 5, 1972. He graduated from the Comilla Cadet College, Bangladesh, and later studied at the Universities of Gunma National College of Technology for diploma in Electrical Engineering, Toyohashi University of Technology for BS, Tokyo Institute of Technology for MS and PhD in Electronics, in Japan.

His employment experience includes the Motorola Japan Ltd in Tokyo, the University of South Florida in USA, Freescale Semiconductor Japan Ltd in Japan and the National University of Malaysia in Malaysia. His special fields of interest include Microelectronics, Semiconductor Physics and Devices, Optoelectronics.

Azrul Hasni Madesa Graduated with a diplôme d'ingenieur (MSc. equivalent) from one of France's top Engineering School and backed by 7 years of experience working with companies both local (TM, Sun Microsystems, MDEC) and international(PriceWaterhouseCoopers, France), Azrul currently holds a position as a Chief Technology Officer of Billadam, a UK company specializing in mobile software and content. He is also currently doing his PhD. in Business Process Management Systems.

Md. Shabiul Islam (M'03) received the B.Sc. (Honors) and M.Sc. degrees at the Department of Applied Physics & Electronics, Rajshahi University, Bangladesh, 1985 and 1986 respectively. He has also obtained his M.Sc. in the area of Micro controller Based System Design at the Department of Electrical, Electronic & Systems Engineering, Universiti Kebangsaan Malaysia, Malaysia in 1996. He became a Member (M) of IEEE in 2003. He was born in 1st October 1964 in Gaibandha, Bangladesh. He is a member of Bangladesh Electronics society and Bangladesh Computer Society His research interests are mainly in the field of VLSI Design, Microprocessor Based system Design. He is currently a lecturer at the Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia and submitted his Ph.D. thesis.