BinDCT Design and Implementation on FPGA with Low Power Architecture

Mohamad Hairol Jabbar

MSc in Microelectronic System Design

2008

BinDCT Design and Implementation on FPGA with Low Power Architecture

Mohamad Hairol Jabbar

A dissertation submitted in partial fulfilment of the requirements of Liverpool John Moores University for the award of Master of Science in Microelectronic System Design.

School of Engineering

September 2008

BINDCT DESIGN AND IMPLEMENTATION ON FPGA WITH LOW POWER ARCHITECTURE

MOHAMAD HAIROL JABBAR

Abstract

Image compression is widely used in today's consumer applications such as digital camcorders, digital cameras, videophones and high-definition television (HDTV). As Discrete Cosine Transform (DCT) is dominant in many international standards for image/video and audio compression, the introduction of multiplierless algorithm for fast DCT computation known as BinDCT (Binary DCT) is very well suited for VLSI implementation. Its performances in term of Peak Signal-to-Noise (PSNR), compression ratio and coding gain is proved to be best approximation to the DCT algorithm.

In this work, the design and implementation of 8 x 8 block 2-D forward BinDCT algorithm on a Field Programmable Gate Array (FPGA) is presented. As this algorithm uses simple arithmetic operations (shift and add) rather than floating-point multiplications, low power hardware implementation is very promising. The aim for low power implementation was achieved at architectural level by employing 4 stages pipeline architecture with parallel processing in each stage. However, due to the trade-off between hardware area and speed, this design is focusing on optimising hardware area in each stage such that it can fit the target FPGA device.

The 8 x 8 block two-dimensional (2-D) forward BinDCT implementation can be run at 68.58 MHz with the power consumption of 144.10 mW. This implementation achieved 12.45% less power compare with the implementation of BinDCT presented previously if the design runs at the same speed. Furthermore, results have shown that this implementation achieved good accuracy compare with software implementation as the maximum error of the output from 2-D computation is 1.26 %.

Several works can be done for further power optimisation such as data gating and latency balancing at each stage (which can improves the throughput as well). Besides, the implementation of 8 x 8 block 2-D inverse BinDCT should be carried out such that its accuracy over floating-point DCT in terms of hardware implementation can be analyzed.

Acknowledgements

First of all, I would like to thank to my supervisor, Prof. Dave Harvey for the idea of this work and for the advice to help me making progress throughout the project. I also thanks to him for the valuable discussion about the project and the time he spend for reviewing my writing. I am grateful to work with him for several month of this project.

Next, I would like to thank to my friends who support me while doing this project and keeps me motivating to complete the project. Any help particularly for the writing processes is much appreciated.

Lastly, I would like to thank to my family, who support me emotionally and keep praying for my success in academic. This work is dedicated to them and to show that I am work very hard to achieve my dreams and to make them proud of me.

Table of Contents

List of Figures

List of Tables

Table **Page**

CHAPTER 1

INTRODUCTION

Digital image processing has become widely used in modern electronic applications. As multimedia applications continue to growth rapidly such as videophone, camcorders, internet applications' using mobile phones and high-definition television (HDTV), development of image processing techniques has become more significant to derive further various multimedia applications. Thus, it has been the subject of interest for many researchers. Speed, performance, hardware area, throughput and power consumption are among the main criteria to be concern in the development of image processing techniques.

Image processing requires transformation from one domain into other domain. Transformation is a way of converting time and space domain into spatial or frequency domain such that the image can be transmitted from one point to another. This is known as transform coding. Generally, transform coding has higher compression ratios than predictive coding but requires more computation through quantization processes. This is the reason why many of multimedia applications use transform coding rather than predictive coding or subband coding. Discrete Cosine Transform (DCT) is a type of transform coding and is the popular image compression scheme.

1.1 Image Compression

Compression is used in image and video processing as well as audio or speech processing. It is used to reduce the size of an image in order to transmit or store with acceptable quality degradation. Compression is achieved during quantisation and entropy encoding process. By compressing the image, less bandwidth is required to

transmit the image which means less time is required for the transmission. It is also requires less memory to store the data. For example, for a still image with 1000 x 1000 pixels at 24 bit uses 3 megabytes of storage in uncompressed form. This large amount of storage is not efficient and costly to be used in any applications particularly on mobile devices. Through compression, for example with 50:1 compression ratio, the size of the original image can be reduced to 60 kilobytes. Thus, the image in compressed form can be used efficiently in practical applications with acceptable quality.

There are two types of compression; lossless and lossy. Lossless, as the name suggests, reconstructs the image identical to the original image without losing any bits in the transformation. It has a low compression ratio of 3:1 or lower. While lossy compression, on the other hand, has a high compression ratio for instance 50:1 for images and 200:1 for video, and thus produces reconstructed data that is not identical with the original. The effect of different compression ratio is shown in Figure 1-1. The human eye can see the difference between an original image and compressed image up to a certain limit of compression ratio. As shown in Figure 1-1, the difference between the original image and compressed image with 10:1 ratio cannot be easily identified by the human eye, but can be measured using computer software. Lossy compressions use less storage area and also reduce communication bandwidth. Thus, its applications are many, such as commercial and consumer electronics applications. Whereas, lossless compression is often use in critical applications such as medical imaging systems, surveillance or security systems, and satellite communications.

(a)

(b)

(c)

Figure 1-1: Compression of an image (a) original image, (b) 10:1 compression ratio, (c) 45:1 compression ratio (Smith, 1997)

The growth of image compression techniques as well as its applications is supported by the introduction of international standards such as the Joint Photographic Experts Group (JPEG), MPEG-I, MPEG-II, MPEG-III by Moving Picture Experts Group (MPEG), and H.261, H.263 by International Telecommunication Union (ITU-T). On the basis of its wide applications range from consumer devices to medical equipment, the search for better methods is still an opportunity open for researchers.

The demand of better image compression techniques is growing rapidly due to its wide applications particularly for communications devices such as mobile devices and broadcasting services. It has become a concern for many researches to find better techniques for communication systems developments. The introduction of new standards accelerate these developments as research work focused on complying the standards as well as enhance practical implementations (Alam et al., 2005), (Kyeounsoo and Jong-Seog, 1999), (Cote et al., 1998), (Madisetti and Willson, 1995), (Scopa et al., 1995), (Jutand et al., 1991).

Discrete Cosine Transform (DCT) is a lossy compression technique, first introduced by Ahmed (Ahmed et al., 1974) which has been developed via the Discrete Fourier Transform (DFT). Since the introduction, many researchers proposed better algorithms to compute fast DCTs. DCT has many advantages compare with other compression techniques and therefore it is employed in the international standard such as JPEG, MPEG, H.261, H.263, and DOLBY. DCT techniques use only the cosine element from the DFT and this reduces the number of coefficients needed to be calculated. This is because a DFT comprises complex numbers containing cosine and sine elements which transform the image or video into the much more complex frequency domain. However, using a DCT which has real inputs with the DFT cosine element, the transformation is simpler and thus has many benefits in terms of less arithmetic and faster speed over the DFT. DCT is very close to the optimal discrete time Karhunen-Loéve transform (KLT). It is an optimal transformation from the perspective of energy compaction since it compacts much energy into a few coefficients. However, KLT is signal dependence and requires extensive computation due to the complex basis calculations. DCT, on the other hand, is a much better transform in practice because of signal independence, linear phase, real coefficients, and faster algorithm. DCT has been used in many digital image

and video processing applications due to its advantages over other compression methods (Shabiul Islam et al., 2006, Habibi, 1974, Ahmed et al., 1975, Natarajan and Ahmed, 1977, Wen-Hsiung and Smith, 1977, Madisetti and Willson, 1995). Video and image compression using DCT have become essential technology in today's multimedia applications. However, there is still room for enhancement of the DCT computation for better practical applications.

1.2 Problem Statements

Without image processing techniques, all the mentioned applications are impossible to be practically used in today's society. As many new applications have been developed as well as new devices and equipment introduced, the problems of power consumption become significant particularly for mobile devices. The needs for as low as possible power consumption is important such that many applications can be fully exploited particularly targeting mobile devices.

Multimedia applications often use complex processing which results in considerably high power consumption. Today's multimedia applications as well as electronic devices require low power consumption to support longer functional operations and broad range of applications. For example, video recording in mobile phone uses a lot of power while the source is limited, thus it cannot be used efficiently. The needs for low power consumption image compression techniques are gaining attention of the designers.

This project focuses on low power implementation with regard to the arithmetic operations on architectural level. The dynamic power consumption of add and multiply operations using integers and floating-point numbers is shown in Figure 1-2. This power consumption is based on signed numbers for integers and single precision IEEE 754

standard for floating-point arithmetic and it simulated with the same test vectors. It was estimated using XPower tool from Xilinx ISE 9.2i for a Spartan-3 XC3S200 FPGA device. It shows that floating-point arithmetic uses more power than integer arithmetic for multiplication and addition. The power consumption will be much higher if complex arithmetic operations are involved. On the basis of this problem, this project explores the hardware implementation with the aim to reduce power consumption of FPGA-based DSP application which involves intensive arithmetic operations.

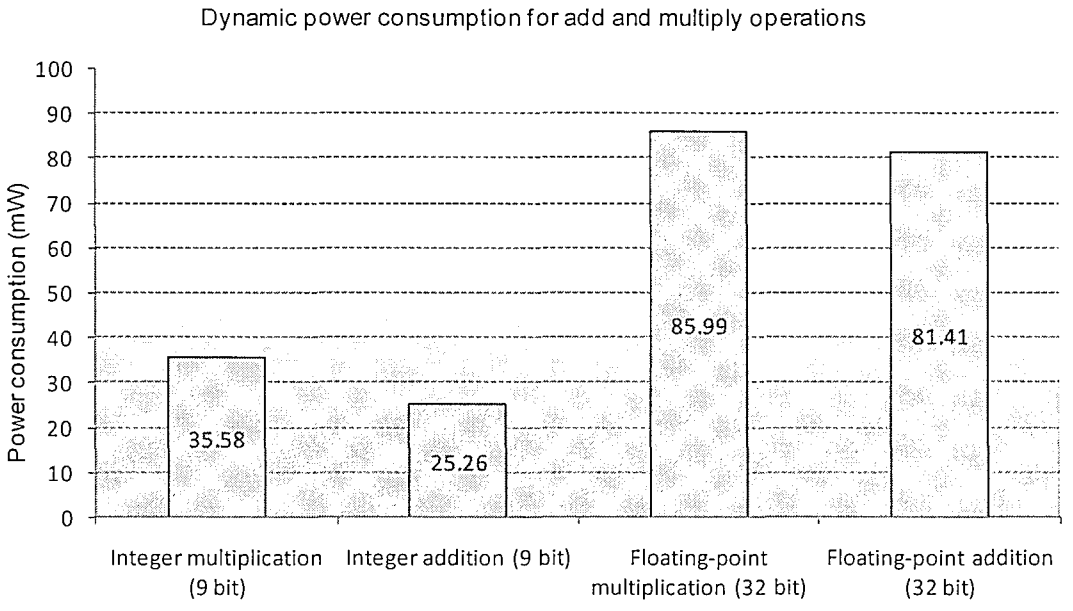

Figure 1-2: Power consumption of addition and multiplication using integers and floating-point numbers

1.3 Project Overview

In this project, the design and implementation of image compression techniques, namely BinDCT, is investigated such that low power consumption can be achieved. The choice of algorithm and hardware devices is discussed further in chapter 2 which includes recent work related to the development proposed.

In this project, Verilog 2001 has been used for the design. Once the design is completed, it is compared with the result from MATLAB software such that the functionality is verified. After the function is successfully simulated, testing was performed with several images and the results analyzed by measuring performance and accuracy parameters. Detailed methodology and results are discussed in chapters 3 and 4.

The project ran from February 2008 until September 2008 as shown in Appendix A and was completed on time. As this project involves hardware implementation, many difficulties were faced and thus the project planning has been revised several times.

1.4 Objectives

The objectives of this project are:

1. To design and implement a 8 x 8 block two-Dimensional (2-D) forward BinDCT on FPGA
2. To analyze the accuracy of the BinDCT implementation over a software implementation
3. To design and implement a low power VLSI architecture based on BinDCT algorithm
4. To analyze and compare the speed-up implementation of BinDCT over floating-point DCT

These objectives were achieved with some limitations as it involved hardware implementation and several other constraints.

1.5 Limitations of the Work

In this project, the design and implementation of 2-D forward BinDCT algorithm is considered to due to time constraints rather than a complete forward and inverse transformation. However, an inverse transform can be realized by reversing the forward transform as it is orthogonal transform. It would still require a significant amount of times to design and debug. In order to analyze the accuracy, the BinDCT implementation is compared with a floating-point DCT implementation. Speed-up implementation is a measure of how much effort is required for each particular implementation. The aim was to design an architecture that is simple such that the implementation requires less effort. Once the design was complete, testing was performed with various benchmark images such that the performance and other parameters of the design can be analyzed.

This design also focused on FPGA hardware types rather than ASIC or dedicated DSP processor due to several reasons discussed in chapter 2. The device, from Xilinx is used for the implementation together with the design tools provided from the same vendor. The Spartan-3 XC3S200 FPGA device has been chosen for the target hardware platform. This first generation of FPGA from 90 nanometres process technology has built in power saving features which save device power automatically with the techniques employed using Xilinx ISE design tools. In addition, routing capacitance can be reduced during power optimisation mode. This device also provides external components power reduction by integrating and saving the power draw of buffers and line drivers.

As specified in the objectives, the main target of this work is to explore low power hardware implementation. While other important parameters such as speed and area also need to be considered in the design, the priority is to find as low as possible power consumption. It is shown there is trade-off between a few parameters for high performance implementation. By achieving low power implementation, other parameters such as chip area, speed, and throughput are also measured.

Low power implementation is explored from the perspective of hardware architecture and not from an algorithm point of views. It can be achieved through pipeline architecture and parallel processing. On the basis of higher computation produces more power consumption, the implementation of low power BinDCT is considering these two techniques. The implementation of 2-D BinDCT is based on row-column method. The 2's complement numbering system is used with fixed-point format of signed numbers. For wordlength, the inputs have 9 bit width signed numbers, while the outputs have 17 bit width signed numbers. Detailed architecture is explained in chapter 3. The image use the grey-scale system where each pixel has the value from 0-255 of 8 bit each, where 0 is the darkest and 255 is the brightest.

1.6 Thesis Organisation

The thesis is organized as follows; chapter 1 explained the overview of the project including problem statements, project objectives and limitations. The reason of low power implementation for image compression method is also discussed.

Chapter 2 discusses the overview of DCT and its fast algorithms. The motivations behind low power hardware implementation and previous related work that has been proposed for fast DCT algorithms and implementations are explored. Several design platforms and target hardware are also covered. Among the proposed fast DCT algorithm, the choice of BinDCT algorithm for hardware realization in this project is explained in detail.

Chapter 3 presents the methodology adopted in this project. It includes the choice of algorithm used for the implementation and its hardware architecture using parallel and pipeline techniques. The design and implementation of 2-D forward BinDCT with low power consumption is described in detail. The 2-D forward BinDCT was constructed based on row-column decomposition where two units of one-dimensional (1-D) BinDCT was used for row-wise and column-wise computation with a transposition matrix between these two units.

Chapter 4 presents the testing that have been performed and evaluation of the design. The result of BinDCT implementation is compared with the MATLAB program such that the correct implementation and its accuracy is verified. Once verified, the performance of the design such as throughput, hardware utilisation and speed was measured and evaluated. The speed-up implementation of the design is compared with true DCT implementation and power consumption is compared with other previous workers.

Lastly, in chapter 5, conclusions are summarized for this project. Future work suggestions to explore further finally explained.

CHAPTER 2

MOTIVATION AND RELATED WORK

Image compression is widely used in digital electronic applications. As communication technology continues to growth, the need for improvement in the image compression techniques has become a major concern. The techniques, particularly dealing with power consumption and performance gain much attention from researchers. Audio bandwidth requires 20 kHz and digital data rate is about 1.4 megabytes per second for high quality stereo sound. For broadcast with high quality video, it requires 10 megabytes per second, while HDTV signals requires 100 megabytes per second. Thus, the role of image compressions is important and become even more significant as the progress of communication technology grows faster. DCT is the dominant algorithms behind many applications that use image compression.

2.1 Discrete Cosine Transform (DCT)

It is first introduced by Ahmed (Ahmed et al., 1974) which was developed via DFT. DCT is a part (cosine) of DFT (cosine and sine) where it has only real value. DCT also transforms a signal more accurately than DFT. On the basis on these characteristics, DCT has gained much attention among the researchers as well as in industry (Blinn, 1993). Since the introduction, a lot of work has been presented with the target of fast DCT calculation for high performance.

DCT has many advantages compared with other compression techniques where it has been used in the international standard such as JPEG, MPEG, and DOLBY. As mentioned in the previous section, DCT compression technique uses only the cosine element from DFT and this reduces the number of coefficients needed to be calculated

during transformation. For example, for an 8-point DFT, 16 coefficients are used (for real and imaginary numbers) rather than 8 coefficients only (for real numbers) in DCT. This is because DFT is a complex transform where the image or video is transformed into the constituent frequency domain which comprises of magnitude and phase information. The DFT has less arithmetic operations compared with the DCT. For instance, an 8 x 8 matrix DFT, the arithmetic operations involve only simple numbers such as 0, 1, -1, and a coefficient which can be realized through fewer multiplications. Whereas in DCT, 64 multiplications are needed to compute all possible products of its coefficients and the inputs for the same matrix size (Blinn, 1993). The advantage of DCT comes from the fact it uses a smaller number of coefficients to get a good approximation while DFT uses more coefficients for a typical signal.

DCT also manipulates the use of human eye characteristics which can only differentiate the colour different known as chrominance. However, it is difficult to distinguish the difference in brightness of an image or picture, which is luminance. On the basis of this, DCT exploits these human visual characteristic in its transformation properties, which focus more on luminance properties of an image. The human eye is more sensitive to low frequency components and overall brightness of an image. Therefore, in DCT, low frequency elements contain more bits than high frequency elements introducing possibilities for high compression, which are one of the reasons for its wide usage in today's multimedia applications.

In theory, KLT is the optimal transformation and DCT is a very close approximation to the KLT compare with other methods. DCT has higher energy compaction and it is very close to the KLT energy characteristic. This is because most of the information of a signal is concentrated in a few components of DCT in the low frequency coefficients. A