FPGA based implementation of baseline JPEG decoder

Conference Paper · January 2009
DOI: 10.1145/1838002.1838035 · Source: DBLP

4 authors, including:

Jahanzeb Ahmad
Altera Europe
8 PUBLICATIONS 41 CITATIONS

Kamran Raza
Iqra University
7 PUBLICATIONS 14 CITATIONS

Mansoor Ebrahim
Sunway Education Group
19 PUBLICATIONS 49 CITATIONS
FPGA based implementation of Baseline JPEG decoder

Jahanzeb Ahmad, Mansoor Ebrahim
Faculty of Engineering, Sciences and Technology, IQRA University, Karachi Pakistan.
jahanzeb.ahmad@iqra.edu.pk, mansoorebrahim99@hotmail.com

Abstract

The JPEG standard (ISO/IEC 10918-1 ITU-T Recommendation T.81) defines compression techniques for image data. As a consequence, it allows to store and transfer image data with considerably reduced demand for storage space and bandwidth. From the four processes provided in the JPEG standard, only one, the baseline process is widely used. In this paper FPGA based High speed, low complexity and low memory implementation of JPEG decoder is presented. The pipeline implementation of the system, allow decompressing multiple image blocks simultaneously. The hardware decoder is designed to operate at 100MHz on Altera Cyclon II or Xilinx Spartan 3E FPGA or equivalent. The decoder is capable of decoding Baseline JPEG color and gray images. Decoder is also capable of downsampling the image by 8. The decoder is designed to meet industrial needs. JFIF, DCF and EXIF standers are implemented in the design.

1. Introduction

Communication and storage cost are reduced by doing data compression. Data compression techniques can be divided into two categories “lossy” and “lossless”. Lossless compression model are based on entropy coding schemes. This model is widely used for text and data compression. In lossless compression model exact data is obtained at the receiver. Lossy compression model produces close approximation of the original data at the receiver. Video, Image and audio compression commonly use lossy compression. Compression ratio up to 100:1 can be achieved depending on the fidelity of the data.

There are several standards/formats for image compression/decompression. Joint Photographic Experts Group (JPEG) [1, 17], Graphics Interchange Format (GIF) [7 8], Portable Network Graphics (PNG) [9], JPEG 2000 [10], Tagged Image File Format (TIFF) [11]. JPEG is a very well know image compression standard. It is widely adopted as compression standard for still images. Joint Photographic Expert Group (JPEG) is a joint workgroup of three international standard organizations, International Organization for Standardization (ISO), International Telegraph and telephone consultative committee (CCITT) and International Electrotechnical commission (IEC).

Enormous amount of data storage is required for digital images/video. An uncompressed color image requires 24 bits for each picture element (pixel). A 6 Mega pixel (3038 X 2012) camera requires 17.5 Mega Bytes, when stored uncompressed, same image when compressed with JPEG take almost 1.7 Mega bytes depending on the compression ratio. En-hui Yang, Longji Wang [18] proposed an algorithm which can further improve this ratio, the algorithm is iterative, which is more complex to implement in Hardware.

Digital devices are now more popular then analog devices especially in the field of multimedia (Audio, Video and Image) because of amazing improvement in digital signal processing algorithms and fast hardware. Digital storage media is more reliable and less effected by noise and distortion.

Real-time implementation of JPEG encoder or decoder requires efficient and fast hardware architecture. So architecture specific implementation is required to achieve real-time results. Variety of architecture designs capable of supporting real time image/video processing already exists such as ASIC, FPGA, Microprocessor and Digital signal processor based design, which implements different algorithms for image and video processing. But only a few efficient architectures are implemented for Image and video compression, decompression, processing [12, 13, 14, 15, 16, 19, 20, 21, 22, 23, 32]. Shizhen Huang and Tianyi Zheng [12] proposed an architecture for PNG image decoding, they used combination hardware and software approach which reduce the throughput of the system. Zulkalnain MohdYousof, et al. [13] proposed a Digital Signal processor based JPEG Decoder but it
can only support small resolution images. R. P. Jacobi et al. [14] proposed an FPGA based JPEG decoder design but its maximum operating frequency is 38.7 MHz on Vertex 6 which is very slow for commercial design. Mario Kovac and N. Ranganathan [15] presented encoder architecture which is capable of operating at 100 MHz and can support 1024x1024 spatial color image resolution. Mohammed Elbadri et al [16] also proposed a FPGA based design for JPEG decoder this design also has low operating frequency, 67 MHz. Kyeong-Yuk Min and Jong-Wha Chong [19] proposed an architecture for JPEG Encoder. Zulkalnain Mohd Yusof et al [20], proposed a Digital Signal Processor (DSP) based architecture, DSP based systems have low development time and cost but low throughput as compare to FPGA.

FPGA is relatively young technology. FPGA can provide speed, performance and flexibility because parallel and pipelined implementation of Algorithm is possible. FPGA provide a better solution because hardware is designed for specific algorithm.

In this paper we proposed a FPGA based JPEG decoder architecture, which gives fast and efficient results. The paper is organized as follow: In section 2 we discuss JPEG in general. In section 3 JPEG stream is discussed. Hardware implementation is discussed in section 4. Synthesis reports are discussed in subsequent section. Finally, results and conclusions are discussed.

2. JPEG compression overview

Principles of JPEG can be explained better to take a look at the steps of encoding rather than decoding. Therefore, despite the fact that a decoder has been developed, due to better understanding the steps of encoding. The steps of decoding will be the inverse of the encoding steps but in reverse order (see Figure 1 and Figure 2).

The human eye is more sensitive to brightness then colors [33]. Almost no loss in visual perception quality can be achieved if chrominance component is stored in half resolution then luminance component [33]. JPEG images are stored in YCbCr color space rather then RGB. CCIR Rec 601 [6] defines the method of conversion between RGB and YCbCr.

Most JPEG encoders reduce the chrominance components to half of the resolution in both dimensions by taking the mean value of each 2x2 block. This sampling method is called “4:2:0”. Another sampling method evolved from analog television signals [33] is “4:2:2” where chrominance components are reduced only in the horizontal dimension. For completeness the “4:4:4” method should be mentioned it does not reduce any component’s resolution. For grayscale (“4:0:0”) images only the Y component is processed. Figure 3 illustrate the described sampling methods. If the “4:2:0” or “4:2:2” sampling method is used this is one of two steps in the compression process where information is lost.

![Figure 1. JPEG Encoder](image1)

![Figure 2. JPEG Decoder](image2)

![Figure 3. Sampling](image3)
dimensional DCT follows straightforward form the one dimensional DCT. A one-dimensional DCT is performed along the rows and then along the columns, or vice versa.

JPEG uses a zero-shift in the input samples to convert 8-bit image data from the range 0 to 255 to the range of -128 to +127. This is done by subtracting 128 before DCT is calculated. DCT is defined in equation (1) and IDCT is defined in equation (2)

FDCT: 
\[
S_{uv} = \frac{1}{4} C_u C_v \sum_{x=0}^{7} \sum_{y=0}^{7} s_{xy} \cos \left( \frac{(2x+1)\pi}{16} \right) \cos \left( \frac{(2y+1)\pi}{16} \right) 
\]

IDCT: 
\[
S_{uv} = \frac{1}{4} \sum_{x=0}^{7} \sum_{y=0}^{7} C_u C_v s_{xy} \cos \left( \frac{(2x+1)\pi}{16} \right) \cos \left( \frac{(2y+1)\pi}{16} \right) 
\]

\( C_u C_v = \frac{1}{\sqrt{2}} \) for \( u, v = 0 \)

The “Quantization” is a key step in the compression process since less important information is discarded. The advantage of the representation in the frequency domain is that, unlike in spatial domain before the DCT, not every dimension has the same importance for the visual quality of the image. Removing the higher frequencies components will reduce the level of detail but the overall structure remains, since it is dominated by the lower frequency components.

The 64 values of a 8x8 block will be divided according to the 64 values of an 8x8 matrix called the quantization table. There is no information lost in the division of the coefficients itself, but the result is then rounded to the next integer afterwards. The higher the divisor, the more information about the coefficient will be positioned after the decimal point hence lost in the rounding operation.

The two dimensional order of the DCT coefficients refers to the two dimensions that the 8x8 block has in spatial domain. After the quantization step most of the coefficients towards the lower right corner are zero. The Zigzag-Mapping - as shown in Figure 5(d) - rearranges the coefficients in a one dimensional order, so that most of the zeroes will be placed at the end. This array with many consecutive zeroes at the end is now optimized to achieve high compression in entropy encoding.

The final step is a combination of three techniques: run length encoding, variable length encoding, and Huffman encoding.

The first coefficient is called “DC” (#0) all other coefficients are called “AC” (#1 till #63).

The first coefficient (DC) is the mean value of the original 8x8 block. There is a correlation between the DC coefficients of neighboring blocks.

It is very likely that the first coefficient has the largest value. This is the most significant coefficient and therefore usually the least reduced one in the quantization step.

Most zero coefficients appear at the end. The chance to find some consecutive zeroes followed by a non-zero component is good as well. Most non-zero coefficients have very small values.

The DC coefficient will be decoded slightly different than the AC coefficients. Respecting the correlation to the neighboring blocks, just for the first block the whole DC coefficient is processed. Later blocks will only encode the difference to the preceding block’s DC component; this applies for each component separately. AC and DC coefficients have different Huffman tables.

Let’s look at an example block of coefficients (the one from Figure 5(e)):
Let’s assume that the previously decoded block of the same component had the DC coefficient 22, therefore we decode the difference 29 - 22 = 7.

So now we take care of the zeroes using run length encoding. The tailing zeroes will be combined in one code, called “EOB”. To each non-zero code we will stick the information about preceding zeroes, so we can remove the rest of the zeroes. For the DC coefficient there will be no preceding zeroes, however, unlike for the AC coefficients, “zero” is still a valid value that has to be concerned.

The remaining coefficients will probably be very small so that variable length approach seems feasible. Therefore we switch to binary representation and add the minimum number of bits needed to represent the coefficients value to the information part. Negative values will be represented by negating every bit (one’s complement). This can be done because we have the information about the length, so that every positive value starts with a 1.

The [EOB] is coded as [0 0], [ZRL] as [15 0]; there is no other code with the structure [X 0].

Since the coefficients are usually very small there is not much gain in compressing them further. However we have not thought about the information we attached to the coefficients yet. We use 4 bits for the preceding zeroes and 4 bits for the number of bits used to store the value. These 8 bits are compressed using a Huffman table which maps the frequently occurring values to shorter bit strings and the rarely occurring values to longer bit strings. How to choose the table is left to the encoder.

Let’s assume we have built the Huffman table and find the following tables:

Now we can construct the final bit stream:

The final bit stream:

So we compressed the 64 bytes of input data down to less than five bytes.

3. JPEG stream

JPEG standard has many parts, only parts which are compliance with applicable parts of DCF [2], Exif [4] and JFIF [3] are implemented. The resulting stream is shown in Figure 6. The variables and parameters are defined in JPEG [1].

4. Hardware implementation

The system consists of different blocks as shown in the block diagram in Figure 8. The interface of the JPEG decoder is shown in Figure 7. The design features are:

- Efficient Design.
- Minimum clock speed of 100 Mhz.
- Target Independent design.
- Synchronous design.

The system works on 8-bit data input. When the start signal is asserted parser start to read data. If first two bytes are not Start of Image Marker parser generates the error. Otherwise it starts searching for the next marker.

JPEG stream parser, parse the input stream. Huffman tables, quantization tables and other information of the image is extracted and stored in the memory by this module. At the start of the entropy coded segment the control of the JPEG decoder is transferred from parser to Huffman decoder controller.
Dequantization is one process where we lose information, this loss can be reduced by using other techniques [26, 27, 28, 29, 31], but these implementations are not the part of this project.

Dequantization and inverse-zigzag is done by one block. Inverse-zigzag was implemented by using simple lookup table. Dequantization was done by multiplying the value with the quantization table.

There are few efficient Huffman decoder architectures [24, 25, 30]. We have implemented our own architecture. Our Huffman decoder decodes the JPEG stream and generates the value in 11-bit. The input of the Huffman decoder is a 8-bit data, Huffman codes from Huffman memories block and image parameters from parser. The controller is inside the Huffman decoder top. Therefore, no extra controller is needed. Inside Huffman decoder top there are 4 sub blocks Huffman_decoder_controller, ecs_filter, value_wrapper and huffmandecoder. Huffmandecoder has more blocks including Huffmanstreamer, decisionmaker and compareblock as shown in Figure 9.

FIFO stores the decoded codes from the Huffman decoder before it is dequantized and inverse zigzagged. JPEG decoder can also downscale image in size by the factor of 8 in both vertical and horizontal direction. Therefore in downscale by 8 mode the IDCT is bypassed. Bypassing IDCT increases the throughput of the decoder.

The 2D DCT/IDCT is based on the 1D fast DCT algorithm first described by Vetterli and Ligtenberg [5]. The input is 8x8 blocks of data in frequency domain and output is 8x8 block of data in time domain.

5. RESULTS/SYNTHESIS REPORT

Synthesis report of JPEG decoder is shown in Table 4 and 5.

Table 1 Synthesis report (Altera)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device</td>
<td>Altera cyclone II EP2C8T144C6</td>
</tr>
<tr>
<td>Synth/P&amp;R Tool</td>
<td>Altera Quartus II v8.1</td>
</tr>
<tr>
<td>Synth/P&amp;R settings</td>
<td>Timing constraint 100 MHz</td>
</tr>
<tr>
<td>Fmax</td>
<td>102.45</td>
</tr>
<tr>
<td>LEs</td>
<td>3996</td>
</tr>
<tr>
<td>Memory</td>
<td>15</td>
</tr>
<tr>
<td>Hard Multipliers</td>
<td>3</td>
</tr>
<tr>
<td>Clocks</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2 Synthesis report (Xilinx)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device</td>
<td>Xilinx Spartan-3E xc3s500e-5-vq100</td>
</tr>
<tr>
<td>Synth/P&amp;R Tool</td>
<td>Xilinx ISE v10.1</td>
</tr>
<tr>
<td>Synth/P&amp;R settings</td>
<td>Time performance with physical synthesis</td>
</tr>
<tr>
<td>Fmax</td>
<td>101.8 MHz</td>
</tr>
<tr>
<td>Slices</td>
<td>3187</td>
</tr>
<tr>
<td>Memory</td>
<td>7 Block RAMs</td>
</tr>
<tr>
<td>Hard Multipliers</td>
<td>2</td>
</tr>
<tr>
<td>Clocks</td>
<td>1</td>
</tr>
</tbody>
</table>
6. CONCLUSION

The goal of this project was to design an efficient JPEG decoder. The design was generic so it can be implemented on any FPGA. The project has four major modules: Parser, Huffman decoder, dequantizer/inv-zig-zag, IDCT. During the project it was noticed that the bottleneck for the throughput is IDCT. Therefore more efficient design of IDCT can increase the throughput. However in downscaling by “8” mode, IDCT is bypassed and throughput increases but in this case Huffman decoder or inverse zig-zig block can be possible bottlenecks. Efficient design and pipelined implementation resulted in 100 MHz operating frequency and small size on silicon. The decoder can decode 6 mega pixel image in 200 msec to 600 msec depending upon image.

7. Acknowledgments

We are thankful to BitSim AB, Stockholm Sweden and Iqra University, Karachi Pakistan, for providing us a research environment and development facilities.

8. References