IMPLEMENTATION METHOD ON MEDICAL IMAGE COMPRESSION SYSTEM: A REVIEW

Azlan Muharam^{a*}, Afandi Ahmad^b

^aReconfigurable Computing for Analytic Acceleration Focus Group (ReCAA), Microelectronics and Nanotechnology – Shamsuddin Research Centre (MiNT-SRC), Universiti Tun Hussein Onn Malaysia (UTHM) Beg Berkunci 101 Parit Raja Batu Pahat Johor 86400, Malaysia

^bReconfigurable Computing for Analytic Acceleration Focus Group (ReCAA), Microelectronics and Nanotechnology – Shamsuddin Research Centre (MiNT-SRC), Department of Computer Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia (UTHM) Beg Berkunci 101 Parit Raja Batu Pahat Johor 86400, Malaysia Article history Received 16 March 2016 Received in revised form 6 March 2017 Accepted 10 September 2017

*Corresponding author afandia@uthm.edu.my

Abstract

The rapid development of medical imaging and the invention of various medicines have benefited mankind and the whole community. Medical image processing is a niche area concerned with the operations and processes of generating images of the human body for clinical purposes. Potential areas such as image acquisition, image enhancement, image compression and storage, and image based visualization also include in medical image processing analysis. Unfortunately, medical image compression dealing with three-dimensional (3-D) modalities still in the pre-matured stage. Along with that, very limited researchers take a challenge to apply hardware on their implementation. Referring to the previous work reviewed, most of the compression method used lossless rather than lossy. For implementation using software, MATLAB and Verilog are the famous candidates among researchers. In term of analysis, most of the previous works conducted objective test compared with subjective test. This paper thoroughly reviews the recent advances in medical image compression mainly in terms of types of compression, software and hardware implementations and performance evaluation. Furthermore, challenges and open research issues are discussed in order to provide perspectives for future potential research. In conclusion, the overall picture of the image processing landscape, where several researchers more focused on software implementations and various combinations of software and hardware implementation.

Keywords: Medical image processing, hardware, software, compression

Abstrak

Perkembangan pesat imej perubatan dan penciptaan pelbagai ubat telah memberi manfaat kepada manusia dan seluruh masyarakat. Pemprosesan imej perubatan adalah bidang khusus yang berkaitan dengan operasi dan proses menghasilkan imej badan manusia bagi tujuan perubatan. Bidang yang berpotensi seperti pemerolehan imej, penambahbaikan imej, pemampatan imej dan penyimpanannya, serta imej berasaskan visual juga termasuk dalam analisis pemprosesan imej perubatan. Malangnya, pemampatan imej perubatan yang berurusan dengan modaliti tiga-dimensi (3-D) masih dalam tahap pra-matang. Seiring dengan itu, bilangan penyelidik yang minima dalam menggunakan perkakasan pada pelaksanaannya. Merujuk kepada hasil penyelidikan lepas, kebanyakan kaedah pemampatan digunakan tanpa kehilangan dan bukan lossy. Untuk pelaksanaan menggunakan perisian, MATLAB dan Verilog adalah perisian yang popular dan menjadi pilihan dalam kalangan penyelidik. Dari segi analisis, sebahagian besar penyelidik hanya menjalankan ujian objektif berbanding dengan ujian subjektif. Kertas ini secara menyeluruh mengkaji perkembangan terkini dalam mampatan imej perubatan terutamanya dari segi jenis pemampatan, pelaksanaan perisian dan perkakasan dan penyelidikan yang berpotensi di masa akan datang. Kesimpulannya, landskap keseluruhan bagi pemprosesan imej lebih tertumpu pada pelaksanaan perisian dan perkakasan.

Kata kunci: Pemprosesan imej perubatan, perkakasan, perisian, pemampatan

© 2017 Penerbit UTM Press. All rights reserved

Full Paper

1.0 INTRODUCTION

Recently, the most widespread use of threedimensional (3-D) imaging modalities, such as magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), and ultrasound (US) have generated massive amounts of volumetric data [1]–[3]. These have provided an impetus to the development of other applications [4], [5], in particular telemedicine and teleradiology [6]–[8]. In these fields, medical image compression is important since both efficient storage and transmission of data through high-bandwidth digital communication lines are of crucial importance [9]–[18].

Despite their advantages, most 3-D medical imaging algorithms are computationally intensive with matrix transformation as the most fundamental operation involved in the transform-based methods. Therefore, there is a real need for high-performance systems, whilst keeping architectures flexible to allow for quick upgradeability with real-time applications [19], [20]. Moreover, in order to obtain efficient solutions for large medical data volume, an efficient implementation of these operations is of significant importance [21], [22].

A reconfigurable hardware, in the form of field programmable gate arrays (FPGAs), appears as a viable building block in the construction of highperformance systems at an economical price [23]– [26]. Consequently, FPGAs are ideal candidates for their inherent advantages such as massive parallelism capabilities, multimillion gate counts, and special low-power packages [27], [28]. A-part from that, it surely understood digital signal processor (DSP) offer low data rates over the FPGAs processor. Meaning that, in high data throughput implementation such as image and video processing, FPGAs outperform over DSPs [29], [30].

A close examination of the algorithms used in real-time medical image processing applications has revealed that many fundamental actions had involved matrix or vector operations [9], [28], [31], [32]. Most of these operations are matrix transforms that include fast Fourier transform (FFT), discrete wavelet transforms (DWT)[33], and other recently developed transforms such as finite Radon, as well as curvelet and ridgelet transforms which are used in either two-dimensional (2-D) or three-dimensional (3-D) medical imaging. Unfortunately, computational complexity for the matrix transform algorithms is in the order of O (N × logN) for FFT to O (N² × J) for the curvelet transform (where N is the transform size and J is the maximum transform resolution level) which are computationally intensive for large size problems. Thus, efficient implementations for these operations are of interest not only because matrix transforms are important in their own right, but also because they automatically lead to efficient solutions to deal with massive medical volumes.

Currently, 3-D transforms are being implemented with other methods, such as in a network of

computers. Additionally, having a chip that is solely dedicated to this transform will give tremendous results [34], [35]. Despite its complexity, there has been an interest in 3-D DWT implementation on various platforms. Previous surveys have classified this research field into three categories: architecture development [36], architecture with FPGAs implementation [37]–[39], and finally architecture that has been implemented on other silicon platforms [40].

Referring to the existing implementation as reported by [40], a huge gap stills remain for further research in exploiting reconfigurable computing for 3-D medical image compression in the form of two major limitations :

- Medical image compression has not been intensively addressed in the existing 3-D DWT implementation. Only the Daubechies filter has been extensively used in several implementations [36], [41], [42], whilst the Haar, Symlet, Coifflet and Biorthogonal filters remain available for further experimentation.
- 2. Image compression is a well-established research area. However, medical image compression researchers, especially dealing with 3-D modalities, are still considered as in stage. Although the infancy newer compression methods algorithm optimizations have been proposed, a very limited hardware implementation of 3-D medical image compression is discussed [9], [19].

The rest of the paper is organized as follows. An overview of the classification of compression method is presented in Section 2. Section 3 explains the hardware-based implementation of the compression systems and architecture. Software and hardwarebased implementation of the compression systems are described in Section 4. Finally, concluding remarks are given in Section 5.

2.0 CLASSIFICATION OF COMPRESSION METHODS

The rapid development of computer applications has come with the enormous increase of diaital images manipulation's, mainly in the domain of multimedia, games, satellite transmissions and medical imaging. This situation has prompted more researchers on effective compression algorithms. The basic idea of image compression is to reduce the middle number of bits per pixel (bpp) necessary for image representation [43]-[45]. Image compression approaches can be divided into two (2) methods: lossless and lossy [46]-[50]. Lossy compression is a data encoding method that compresses data by discarding (losing) some data in the image [51], [52]. Meanwhile, lossless compression is used to compress the image without any loss of data in the image [53]-[55].

A variety of methods has been developed for compressing images and videos. However, if a compression method does not take into account the contents of an image, important information may be lost due to the high compression. We can transfer images and videos more efficiently if we extract important regions for humans and make the compression rate of other regions as low as possible.

The performances of several medical image compression techniques are presented in [43]. Three types of entropy encoding have been used, which include set partitioning in hierarchal trees (SPIHT) encoding, Huffman encoding and run length encoding (RLE). For a JPEG based image compression, the RLE and Huffman encoding techniques were used by varying the bit pixel. Meanwhile, for a JPEG 2000 based image compression, the SPIHT encoding method was used. SPIHT is a very effective and computationally simple technique for image compression. A-part from that, this method provides a way to generate consistent quality images at a lower bit rate compared to JPEG. In this work, the Cohen-Debaucheries-Feauveau Biorthogonal wavelet was used along with SPIHT to provide a high compression ratio and good resolution. Figure 1 shows a JPEG block diagram for full cycle compression system.



Figure 1 JPEG block diagram [43].

In this implementation, DWT has become an important method image compression. for Moreover, wavelet based coding has provided substantial improvements in picture quality at high compression ratios, mainly due to the better energy compaction property of the wavelet transforms. The standard JPEG is only capable of providing excellent performances at rates of higher than 0.25 bits per pixel. In order to overcome this situation, the JPEG committee had initiated the development of another standard, commonly known as JPEG 2000, JPEG 2000 is a standard that was based on wavelet decomposition. To evaluate the compression performance, magnetic resonance image (MRI) data with pixel resolution 512×512 and in 8-bits was used, whilst the peak signal to noise ratio (PSNR) and compression ratio (CR) was deployed for objective assessments.

Results have shown that for CT scan image JPEG compression method outperforms the PSNR and degree of compression than wavelet compression method. For a lower compression ratio, JPEG has

yielded higher quality images than the wavelet. While the DCT based image JPEG coders perform very well at moderate bit rates, at higher compression ratios, image quality was degraded due to the artefacts resulting from the block-based DCT scheme.

On the other hand, wavelet based coding has led to substantial improvements in picture quality at low bit rates because of the overlapping basis functions and better energy compaction property of wavelet transforms. SPIHT is the most efficient method with respect to compression ratio and PSNR value. With the increasing degree of compression SPIHT was capable of maintaining the image quality quite well. The encoding and decoding time had increased when the bit rate increased.

The implementation of image compression technique of a new transformation method called the enhance DPCM transformation (EDT) was presented in [56]. The EDT is based on predictive models and more energy will be obtained by improving prediction ability. The proposed new transform method engine is illustrated as a block diagram in Figure 2.

Huffman encoding was chosen as the entropy encoder for this lossless compression method. The proposed algorithm was validated with non-medical images and real medical images. Results from the various test images have shown that this new compression method has improved the lossless ratio for JPEG and JPEG 2000 to approximately 7% and more than 3.5% respectively.



Figure 2 Block diagram of new transformation method [56]

Another issue related to the compression methodology was reported in [57]. This work proposed a hybrid approach that can efficiently compress 3-D medical images and optionally embeds a digital invisible watermark inside them. Moreover, this proposed hybrid approach for the lossless compression of 3-D medical images, was based on the predictive coding model (PCM). A-part from that, transform domain techniques, usually permit the embedding of watermark string inside the whole image by using spectral coefficient, particularly the DCT and DWT. Two modalities of medical imaging have been used in this work; the CT, and the MRI. Additionally, this hybrid approach has enabled the efficient lossless compression of 3-D medical images while simultaneously embedding a

digital invisible watermark string. Although hardware implementations were not discussed, interestingly, this work has provided significant information on the hybrid approaches for compression techniques. The proposed hybrid approach is illustrated as a block diagram in Figure 3.



Figure 3 Block diagram of hybrid approach [57]

A new compression scheme based on symmetry in the field of prediction-error was reported in [58]. In order to provide an efficient method, both the intra and inter-slice correlations have been considered. The existing correlations in the slice and the third dimension redundancies between consecutive slices were respectively extracted for the intra-slice and inter-slice compression purposes. Generally, the main idea of this proposed method was to use the symmetrical nature of the brain to predict the pixels of half of the image, using the corresponding pixels in the other half of the image. Therefore, this proposed method has six stages of compression algorithm as shown in Figure 4.

These stages include the prediction stage; splitting of the result of the predictor; two separate block matching methods (symmetry-based and inter-slice); three-dimensional context modeling; regression; and entropy coding. Moreover, the implementation of the proposed compression algorithm; attached with the activity level classification model (ALCM), block matching and 3-D context modeling, has revealed that bit rates of the proposed method were better than for the JPEG-LS and JPEG-3D. Therefore, on average, this proposed algorithm has superiority over both standards.



Figure 4 Block diagram of proposed method of six stages of compression algorithm [58]

An analysis of the discrete fractional Fourier transform (DFT) based medical image compression using a hybrid encoding technique was discussed in [59]. This system comprises of the following phases namely encryption; domain transformation; SPIHT algorithm with Huffman compression; decoding through SPIHT with Huffman encoder and inverse fast two-dimensional fractional Fourier transform; and finally, the quasi-group decryption with Hadamard and number theoretic transform. Hadamard transform is a generalized class of Fourier transforms, created either recursively, or through binary representation. The overall process diagram is shown in Figure 5.

A-part from that, the Hadamard and number theoretic transform act as hash functions which produce diverse hash values for different input values [60]. Whereas, there is a huge difference in the generated random sequence if there is a one-bit change in the input sequence. In brief, this proposed approach was observed as being capable of producing the best results for all the test images taken for consideration. The experimental results have shown higher PSNR values and lesser means squared error (MSE) values when compared with other existing techniques.



Figure 5 The overall flow of the proposed image compression approach [59]

A variety of method and approach were conducting to archive the better results for image compression. Lossless compression techniques more famous among researcher due to the algorithm that allows perfectly reconstructed the original data from the compressed data. Wavelet-based one of the popular method of improvement the picture quality at high CR. Unfortunately, for lower CR JPEG is better than wavelet. Weaknesses of standard JPEG is only greatly performed at a rate more than 0.25bpp. In order to overcome that issues, JPEG 2000 were replaced with based on wavelet decomposition. On the other side of image compression, most of the researchers were focussing on objective tests such as CR, MSE, and PSNR. The objective test successfully archived the proposed algorithm and architecture, where to outperform with the previous method. Furthermore, in order to evaluate the best duration of image compression, measuring of latency and throughput were conducted. A-part from that, three (3) types of popular transform were used DCT, DWT, and DFT. The main distinctive feature of transforms that makes them so efficient in digital image processing is their energy compaction capability. For the encoding process, which is final block of image compression. RLE, Huffman and SPIHT were used in previous research. Furthermore, future potential research would explore on context-adaptive binary arithmetic coding (CABAC) and context-adaptive variable-length coding (CAVLC).

3.0 HARDWARE-BASED IMPLEMENTATION OF THE COMPRESSION SYSTEMS

A key aspect of the image compression was the implementation of hardware-based. Architecture or algorithm was implemented in selected board depend on the applications. The performances of their architecture in the hardware were evaluated.

Another issue regarding medical image compression was presented by [61]. A curveletbased medical image compression using system error compensation (SEC) was proposed. The quantisation and curvelet transform was used as the SEC algorithm. Output image and the system error can be combined to create output images, as demonstrated in Figure 6.



Figure 6 Block diagram of vector quantisation [61]

Additionally, the curvelet transform was used to improve the PSNR while the encoding process was used to produce the image with high PSNR at low bit rate. This experiment was conducted using Intel (R) Core(TM) i5 CPU M520 @ 2.40GHz. In order to test this algorithm, four (4) standard images the size of 512*512 pixels and block size of 4*4 pixels (Airplane, Girl, Lena, and Pepper) and three (3) medical images (MRI cardiac tract, MRI skull, and Ultrasound Liver Cyst) have been used. The experimental results have shown that the proposed method can improve both the bit rate and PSNR by 40.48% and 9.69% respectively, when compared to conventional method.

The main advantage of distributed arithmetic (DA) is its efficiency of mechanization, which is in contrast with the multiply-accumulate (MAC) structure. The MAC computes bits of a single input vector in parallel, while the DA serially computes individual bits of multiple inputs. Figure 7 shows the basic block diagram of the DA. The basic idea of the DA is to replace all multiplications and additions by applying a look-up-table (LUT) and shifter accumulator. In addition, this technique relies on the fact that one of the input coefficients is constant [62]. Which important difference and a prerequisite for DA techniques.



Figure 7 Basic block diagram of LUT-based DA [62]

In addition, DA can also be used to compute sum of product and the implementation of DA- based multiplier exploits the LUT architecture. This design was targeted on Xilinx Virtex-2 pro and the result of using DA architecture has shown speed (throughput) increment by 25%.

According to [63], DA provides an efficient method of computing vectors or matrix multiplication by means of bit level rearrangement of the multiplyaccumulate process. This process distributes arithmetic operations rather than grouping them as multipliers would normally do. Implementation of FPGAs using DA architecture will speed up the efficiency of mechanization, which unfortunately will become slower, particularly when deadline with large transformation because of its bit serial nature. However, some modification such as portioning and bit pairing can significantly increase its performance. Figure 8 illustrates the proposed architecture for 1-D HWT (N=8) based on the DA principle.



Figure 8 Proposed architecture (Arch 2) for 1-D HWT (N=8)[63]

On the other hand, systolic array (SA) design techniques are an array of processing elements (PE), called cells. Each cell is connected to a small number of their nearest neighbours in a mesh-like topology. Each cell performs a sequence of operation on the data that flows between them [64]. PE at each step takes input data from one or more neighbours (Left and Top), processes it and, in the next step, output results in the opposite direction (Right and Bottom). Figure 9 illustrates, the proposed two-dimensional systolic architecture for 3 by 3 matrices. The proposed architecture was designed for matrix multiplication and targeted for FPGAs devices. The parallel processing and pipelining were introduced into the proposed systolic array to enhance the speed and reduce the complexity of the matrix multiplier.



Figure 9 Two-dimensional (2-D) systolic array [64]

The 2-D processor array and a smart schema for communication and data storage were proposed by Saldana and Arias [65] as shown in Figure 10. The purpose of this architecture is to allow process chaining that can be replicated inside the same FPGAs several times in order to process different algorithms independently. Each PE has been specially designed to support the operations involved in ME. In addition, to provide more capacity to the architecture, the PE's functionality has been extended to support most window-based operation in image processing; multiplication, addition, subtraction, accumulation, maximum, minimum and absolute value. This research had also implemented the Handle-C, DK4 and synthesis technology (XST) tools.



Figure 10 The proposed architecture [65]

Referring to the current research trends regarding hardware implementation [61]–[65], most of the researcher's tendency to explore in term of throughput (speed) and PSNR, instead of power consumption and maximum frequency. These because of researchers concentrate on the quality of images and the possible time to process an architecture on the selected hardware. Ultimately, SA and DA architectures frequently used for the hardware implementation. More importantly, DA was the efficiency of mechanization, which is in contrast with the multiply-accumulate (MAC) structure. Important to realize a characteristic of FPGAs, when dealing with large transformation it's become slower as a nature of bit serial hardware. In order increased its performance, some modification such as portioning and bit pairing were applied [63]. Table 1 shows the summary of DA and SA implementation.

 Table 1
 Summary of DA and SA implementation

Reference	DA	SA	Application
[66]	\checkmark		FIR filters
[67]	\checkmark		Entropy encoding
[68]	\checkmark		Multiplierless filter
[69]	\checkmark		Algorithm
[14]	\checkmark		Modified DA based DWT
			architecture
[70]		\checkmark	Scalable core architecture
[71]		\checkmark	SA-DWT architecture
[72]		\checkmark	Full Search Block Matching
			Algorithm

4.0 SOFTWARE AND HARDWARE-BASED IMPLEMENTATION OF THE COMPRESSION SYSTEMS

The nature of image compression is focused on software implementations and showing results. Thus, several contemporary researchers related to image compression in integrating of software and hardware will be explored.

An efficient implementing both sequential and parallel versions of fractal image compression algorithms using compute unified device architecture (CUDA) programming model for parallelizing the program in graphical processer unit (GPU) for medical images is described in [73]. Several improvements in terms of an algorithm for the implementation have been made. Fractal image compression is based on self-similarity of an image, which means that the image has similarity in a majority of its regions. Original implementation with three kernel black layouts in 8 x 8, 16 x 16, and 32 x 32 were conducted. Ultimately, the 16 x 16 blocks had performed better with the best results of execution time for different image sizes using global memory.

Other researcher proposed a method of parallel processing of Sobel edge detector and homomorphic filtering [74]. Both GPU parallel code and the CPU serial code were designed for each image-processing algorithm and the calculation was based on single precision floating point values to ensure comparability of the results. Results have shown that for speed comparison between the Sobel edge detectors, the GPU has significantly improved computing speed when the image size increased. The speedups archived range from 3.6 (128 x 128) to 25.3 (2048 x 2048). On the other hand, homomorphic filtering has shown approximately 40 times speedup.

Thus, it can be concluded that the GPU has significantly improved approximately 25 times and 49 times as fast as CPU [74].

As we know, ultrasound (US) images are captured in real-time, means all the interaction and movement through the blood vessels of the internal body. According to [75], the architecture of medical Bmode for the backend was implemented in Kintex-7 FPGAs platform and MATLAB Simulink. In order to achieve the ideal dynamic range for display and image enhancement, fixed filter coefficient for transformation and compression Hilbert log techniques was used. Results show, the resolution of display was improved and the hardware resource utilization is minimized leading to compact design for portable ultrasound systems.

The comparisons between CUDA, C, MATLAB and FPGAs for the convolution of greyscale images are reported in [76]. The C implementation was based on an equation which h was followed by a sequentially separable algorithm. Next, for MATLAB implementation, a built in conv2() function was used to perform the convolution. A-part from that, CUDA was implemented through two (2) different kernels; the first part was through the line kernels and the second part was through the column kernel. For FPGAs implementation, the architecture was developed with the assistance of SOPC Builder and Verilog HDL coding.

Based on the results, it is inferable that CUDA has presented the best performance in execution time, a number of clock cycles and speedup in comparison to C, MATLAB and the implementation of FPGAs architecture and growth in image resolution. The overall CUDA performance has archived speedups of roughly 200 times in comparison to C, 70 times in comparison to MATLAB and 20 times in comparison to FPGAs.

CABAC was used in H.264/AVC and it has offered compression results that are 10%-15% better than those obtained with the baseline CAVLC entropy coder [77]. CABAC consists of a two-tier process. Firstly, events produced during the encoding process are converted into sequences of binary symbols, sometimes people known as the binarization process. Then, a binary arithmetic coder was used to perform the actual compression. Figure 11 shows the general procedure for binary AC and CABAC procedures. This design was implemented using 0.35µ standard cells libraries to obtain the maximum speed of 186 MHz with the moderate area. A FPGAs based implementation has been considered using the Virtex-II 2000 device and the maximum speed had reached 92Mhz, with only 10% of the slices were used.

193



Figure 11 (a) General procedure for binary AC. (b) CABAC procedure [77]

The CABAC encoding process consisted of three (3) elementary steps; context model selection, binarization, and self-adaptive binary arithmetic encoding, while the last step consisted of probability estimation and binary arithmetic encoder reported in [78]. Figure 12 illustrates the CABAC block diagram. In CABAC, the probability state will be updated after every binary symbol's coding, which allowed the probability estimation in the context model to keep up with the real video stream's statistical properties thus improving the coding efficiency. Results from the proposed CABAC have shown, that CABAC algorithm can tremendously decrease computations. At the same time, by adaptively changing the packet size, a better coding speed for CABAC can be obtained.



Figure 12 General block diagram of CABAC [78]

For integration approach, shows a variety of architecture, algorithm, software, and hardware had been used. CUDA is a new software as a platform to the programming model for parallelizing in order to communicate with GPU. CUDA is created by NVIDIA and based on C language, where ISE based on VHDL and Verilog. GPU shows the performance of computing speed increased proportionally with increasing of image size. Meaning that GPU outperforming CPU approximately 25 and 49 times fast [74]. A-part from that, CUDA had revealed the best performance in execution time, a number of clock cycles and speedup in comparison to C, MATLAB and implementation of FPGAs [76]. Entropy coding is the last stage of image compression blocks, CABAC one of the familiar arithmetic codinas had been used. It's used H.264/AVC standard with lossless compression techniques. Finding of research shows, it has offered 10-15% better compression results than obtained with the CAVLC.

Another study by Alex et al. [65] proposed a technique to investigate and compare some of these basic operations (direct transformations, indirect transformations, and computation of partial derivatives) as follow: (1) recursive filter based implementations in MATLAB and C++ and (2) GPUaccelerated implementations in CUDA. All operations were compared to a variety of resolution levels on a 2-D panoramic image as well as a 3-D magnetic resonance (MR) image. Results indicated significant improvements in efficiency for the CUDA implementations. Their experiment was conducted using Intel Core i7-2600 CPU at 3.40GHz and 16GB of RAM. The GPU used was a NVIDIA Tesla C2070 with 4GB of global memory at 1494 MHz and a core clock of 1.147 GHz.

Pingfan et al. [79] presented the technique of GPU implementation of the multi-frequency bi-planar inter-imaging (MBI) algorithm. In order to execute this technique, an efficient memory access arrangement and a balanced parallel thread assignment were required. Furthermore, GPU would provide threads as the parallel processing unit. Different sets of data were processed but threads concurrently shared the same instruction. Once MBI algorithm was implemented in GPU, every three element estimation operation would be assigned to one thread. Global memory was generally a major bottleneck issue for GPU implementations. In order to minimize this impact on the performance, a memory coalescina technique was proposed. The input data and the intermediate data were stored in the registers as opposed to reading them again from the global memory. The experiments were performed using Intel i7 guad-core 3.4GHz workstation equipped with GTX590 GPU.

Hui et al. [80] implemented the OS-EM iterative image reconstruction algorithms for a clinical SPECT system. The Monte Carlo software SIMSET was used and CUDA was adopted for the GPU programming of the OS-EM algorithm with high complexity. Two (2) implementations in this experiment, accessed every element of the pre-stored system matrix and the system matrix was calculated on-the-fly. Finally, it was evident that, for both GPU implementations, there were no significant differences in images reconstructed by the GPU and CPU implementations of the OS-EM algorithm but for the FBP reconstruction image, the OS-EM reconstruction image had a better performance. The experiments were conducted on a system with Intel [®] Core [™] i7 CPU and NVIDIA Tesla C1060 GPU.

Shoko et al. [81] presented a method for parallelization. Ray-driven and voxel-driven methods were categories of voxel searching method, in the projection calculation. For the forward projection, most of the former time were used, while the latter was used in back-projection. However, for the listmode reconstruction in GPU implementation, by using the ray-driven method was smaller the calculation cost in the back projection rather than used voxel-driven method because the ray-driven method searched only the voxels traversed by the measured line of response (LORs). Therefore, to address these issues, the ray-driven method was implemented in both forward and back projections. The projection value of each LOR was calculated in each thread for the forward projection. For the back projection, the ray-driven was used, a memory access confliction had to be prevented because at the same time some threads may access the same voxel address. Finally, calculated voxels for each LOR were grouped by slice number and the calculations for grouped voxels were included in the same thread in the back projection. This evaluation was performed using 2.67GHz Intel Quad-Core Xeon CPU and NVIDIA Tesla C1060 GPU.

In order to provide a brief snapshot of existing techniques, Figure 13 and Table 2 lists representative samples of 3-D medical image compression along with their key ideas.



Figure 13 Trend of implementation

It can be observed that various elements and techniques are used in the three components, i.e., compression, implementation, and types of evaluation. In the following, existing work is extensively reviewed from the viewpoint of each of these three components. To our best knowledge, there has been no systematic subjective test conducted a comparison of the objective and subjective test [82]. In the table, we include only the cases where used data from medical imaging such as MRI, US, CT and X-Ray. This list justifies the significant amount of work that needs to be carried out on hardware development of 3-D medical imaging.

5.0 CONCLUSION AND FUTURE POTENTIAL RESEARCH

In conclusion, all the methodologies and approaches discussed in this review have used the same sequences or blocks of image processing. Referring from previous works have shown the overall picture of the image processing landscape, where several researchers more focused on software combinations of implementations and various and software hardware implementation. Furthermore, it has been discussed that the GPU was more outstanding compared to other hardware, such as the CPU and FPGAs [74]. Nevertheless, the certain hardware is capable of similarly excellent performance depending on applications and algorithms.

Reference	Data	Compression		Implementation		Evaluations/ Tests		
		Lossy	Lossless	Hardware	Software	Objective	Subjective	Others
[61]	MRI and		\checkmark		\checkmark	Compression		
	US					ratio and PSNR		
[83]	X-Ray		\checkmark	Virtex-5,	Verilog	Compression		
				FPGAs		rates, time and		
						decompression		
						time		
[58]	MRI		\checkmark		MATLAB	Bit rates and		
						threshold		
[57]	MRI and		\checkmark		\checkmark	Bits per pixel and		
	CT					PSNR		
[84]	MRI		\checkmark	Virtex-6-	Verilog	Frequency and		
				FPGAs	-	processing time		
[56]	MRI and	\checkmark			\checkmark	Compression		
	CT					ratio		
[43]	MRI and		\checkmark		MATLAB	MSE, PSNR and		
	CT					compression		
						ratio		
[12]	MRI and		\checkmark		\checkmark	Compression		
	CT					ratio and PSNR		
[24]	MRI		\checkmark	Virtex-6-	Verilog	Logic Utilization		
				FPGAs	-	(Register, LUTs)		
[51]	MRI and	\checkmark			\checkmark	Compression		
	CT					ratio and PSNR		
[75]	US		\checkmark	Kintex-7-	MATLAB			Image
				FPGAs				clarity

	Table 2 Summary	of 3-D medical image	compression
--	-----------------	----------------------	-------------

References

- N. H. Ja'Afar, A. Ahmad, and A. Amira. 2013. Distributed Arithmetic Architecture of Discrete Wavelet Transform (DWT) with hybrid method. Proc. IEEE Int. Conf. Electron. Circuits Syst. 501-507.
- [2] C. Dhaarani, D. Venugopal, and A. Sivanantha Raja. 2014. Medical Image Compression using Ripplet Transform. Proc. - 2014 Int. Conf. Intell. Comput. Appl. ICICA 2014. 233-238.
- [3] R. Pizzolante, A. Castiglione, B. Carpentieri, and A. De Santis. 2014. Parallel Low-complexity Lossless Coding of Three-dimensional Medical Images. Proc.-2014 Int. Conf. Network-Based Inf. Syst. NBiS 2014. 91-98.
- [4] T. Li, Y. Wang, C. Chang, N. Hu, and Y. Zheng. 2014. Colorappearance-model Based Fusion of Gray and Pseudocolor Images for Medical Applications. *Inf. Fusion*. 19(1): 103-114.
- [5] J. Zhang, G. Lin, L. Wu, C. Wang, and Y. Cheng. 2015. Wavelet and Fast Bilateral Filter Based De-speckling Method for Medical Ultrasound Images. *Biomed. Signal Process. Control.* 18: 1-10.
- [6] T. Nadu and S. S. Lakshmi. 2014. Segmentation and Compression of Medical Image Using MSPIHT in Telemedicine Application. ICICES. 978.
- [7] M. Razaak and M. G. Martini. 2014. Rate-distortion and Rate-quality Performance Analysis of HEVC Compression of Medical Ultrasound Videos. *Procedia Comput. Sci.* 40: 230-236.
- [8] R. Sumalatha and M. V. Subramanyam. 2015. Hierarchical Lossless Image Compression for Telemedicine Applications. Procedia Comput. Sci. 54: 838-848.
- [9] A. Arthur and V. Saravanan. 2012. Efficient Medical Image Compression Technique for Telemedicine Considering Online and Offline Application. 2012 Int. Conf. Comput. Commun. Appl. ICCCA 2012.
- [10] D. U. Shah and C. H. Vithlani. 2011. Efficient implementations of Discrete Wavelet Transforms using FPGAs. Int. J. Adv. Eng. Technol. 1(4): 100-111.
- [11] M. Firoozbakht, J. Dehmeshki, M. Martini, Y. Ebrahimdoost, H. Amin, M. Dehkordi, A. Youannic, and S. D. Qanadli. 2010. Compression of Digital Medical Images based on Multiple Regions of Interest. 4th Int. Conf. Digit. Soc. ICDS 2010, Incl. CYBERLAWS 2010 1st Int. Conf. Tech. Leg. Asp. e-Society. 260-263.
- [12] V. Sanchez, R. Abugharbieh, and P. Nasiopoulos. 2010. 3-D Scalable Medical Image Compression with Optimized Volume of Interest Coding. *IEEE Trans. Med. Imaging*. 29(10): 1808-1820.
- [13] P. Suapang, K. Dejhan, and S. Yimmun. 2010. Medical Image Archiving, Processing, Analysis and Communication System for Teleradiology. IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON. 339-345.
- [14] M. Nagabushanam, 2011. Design and FPGA Implementation of Modified Distributive Arithmetic based Dwt-Idwt Processor for Image Compression. International Conference on Communications and Signal Processing (ICCSP). 1-4.
- [15] Q. Min and R. J. T. Sadleir. 2012. Medical Image Compression Using Region-based Prediction. IEEE EMBS Int. Conf. Biomed. Eng. Sci. 677-682.
- [16] G. Ukrit, Mferni. Suresh. 2013. Effective Lossless Compression for Medical Image Sequences Using Composite Algorithm. Int. Conf. Circuits, Power Comput. Technol. 1122-1126.
- [17] J. A. Sophia. P Eben. 2014. Implementation of Region Based Medical Image Compression for Telemedicine Application. IEEE Int. Conf. Comput. Intell. Comput. Res. 1-4.
- [18] S. A. Rein, F. H. P. Fitzek, and T. Sikora. 2015. Evaluation of the Wavelet Image Two-line Coder: A Low Complexity Scheme for Image Compression. Signal Process. Image Commun. 37: 58-74.

- [19] A. Ahmad, B. Krill, A. Amira, and H. Rabah. 2010. Efficient Architectures for 3D HWT using Dynamic Partial Reconfiguration. J. Syst. Archit. 56: 305-316.
- [20] N. Zhou, H. Li, D. Wang, S. Pan, and Z. Zhou. 2015. Image Compression and Encryption Scheme based on 2-D Compressive Sensing and Fractional Mellin Transform. Opt. Commun. 43: 10-21.
- [21] Q. Min and Robert J.T. Sadleir. 2012. An Edge-based Prediction Approach for Medical Image Compression. IEEE EMBS Int. Conf. Biomed. Eng. Sci. 717-722.
- [22] K. G. Thanushkodi and S. Bhavani. 2013. Comparison of Fractal Coding Methods for Medical Image Compression. IET Image Process. 7(7): 686-693.
- [23] S. Kim, H. Sohn, J. H. Chang, T. Song, and Y. Yoo. 2010. A PC-based Fully-programmable Medical Ultrasound Imaging System Using a Graphics Processing Unit. IEEE Int. Ultrason. Symp. 314-317.
- [24] I. Chiuchisan. 2013. A new FPGA-based Real-time Configurable System for Medical Image Processing. E-Health Bioeng. Conf. EHB 2013. 0-3.
- [25] S. Saha, K. H. Uddin, M. S. Islam, M. Jahiruzzaman, and A. B. M. A. Hossain. 2014. Implementation of Simplified Normalized Cut Graph Partitioning Algorithm on FPGA for image Segmentation. Ski. 2014 - 8th Int. Conf. Software, Knowledge, Inf. Manag. Appl. 3.
- [26] Y. Li, W. Jia, B. Luan, Z. H. Mao, H. Zhang, and M. Sun. 2015. A FPGA implementation of JPEG baseline Encoder for Wearable Devices. 2015 41st Annu. Northeast Biomed. Eng. Conf. NEBEC 2015. 3-4.
- [27] A. Ahmad, B. Krill, A. Amira, and H. Rabah. 2009. 3D Haar wavelet Transform with Dynamic Partial Reconfiguration for 3D Medical Image Compression. 2009 IEEE Biomedical Circuits and Systems Conference. 1: 137-140.
- [28] A. Ahmad and A. Amira. 2009. Efficient Reconfigurable Architectures for 3D Medical Image Compression. Int. Conf. Field-Programmable Technol. 472-474.
- [29] N. Kehtarnavaz and S. Mahotra. 2011. FPGA Implementation Made Easy for Applied Digital Signal Processing Courses. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2892-2895.
- [30] K. Benkrid, A. Akoglu, C. Ling, Y. Song, Y. Liu, and X. Tian. 2012. High Performance Biological Pairwise Sequence Alignment: FPGA Versus GPU Versus Cell BE Versus GPP. Int. J. Reconfigurable Comput. 2012.
- [31] G. Grossi, R. Lanzarotti, and J. Lin. 2015. High-rate compression of ECG signals by an accuracy-driven sparsity model relying on natural basis. *Digit. Signal Process. A Rev. Journal.* vol. 45: 96–106.
- [32] A. H. Abouali. 2014. Object-based VQ for Image Compression. Ain Shams Eng. Journal. 6(1): 211-216.
- [33] Z. Liu, H. Yin, Y. Chai, and S. X. Yang. 2014. A Novel Approach for Multimodal Medical Image Fusion. Expert Syst. Appl. 41(16): 7425-7435.
- [34] T. Chen, Y. Wang, H. Zhang, and C. Xiao. 2010. An Embedded 3D Medical Image Processing and Visualization Platform Based on Dual-core Processor. Proc. World Congr. Intell. Control Autom. 2007. 2936-2941.
- [35] N. L. Bian, Jing, Yinghong Dong, Kejing Li. 2013. Research and Application of 3D Interactive Processing System for Medical Image based on MITK Algorithm. Int. Conf. Mechatron. Sci. Electr. Eng. Comput. 1951-1954.
- [36] M. Weeks and M. Bayoumi. 1998. 3D Discrete Wavelet Transform Architectures. ISCAS '98. Proc. 1998 IEEE Int. Symp. Circuits Syst. 4.
- [37] B. Das and S. Banerjee. 2003. A Memory Efficient 3-D DWT Architecture. 16th International Conference on VLSI Design, Proceedings.
- [38] Y.-H. B. Wu, L.-X. Jin, and H.-J. Tao. 2010. An Improved Fast Parallel SPIHT Algorithm and its FPGA Implementation. Proc. 2010 2nd Int. Conf. Futur. Comput. Commun. ICFCC 2010. 1(3): 1191-1195.
- [39] R. Gavvala, M. M. Gopal, S. S. Chandra, and S. S. Rao.

2012. Pass-parallel VLSI Architecture of BPC for Embedded Block Coder in JPEG2000. Asia Pacific Conf. Postgrad. Res. Microelectron. Electron. 111-117.

- [40] S. M. Ismail, A. E. Salama, and M. F. Abu-ElYazeed. 2007. FPGA Implementation of an Efficient 3D-WT Temporal Decomposition Algorithm for Video Compression. 2007 IEEE International Symposium on Signal Processing and Information Technology.
- [41] Z. Shi, L.-Y. Deng, and Q.-J. Chen. 2007. Numerical Solution of Differential Equations By Using Haar Wavelets. 2-4.
- [42] R. Singh and A. Khare. 2014. Fusion of Multimodal Medical Images using Daubechies Complex Wavelet Transform-A Multiresolution Approach. Inf. Fusion. 19(1): 49-60.
- [43] S. J. Pinto and Jayanand P.Gawande. 2012. Performance Analysis of Medical Image Compression Techniques. 3rd Asian Himalayas International Conference on Internet (AH-ICI). 5-8.
- [44] Z. Zuo, X. Lan, L. Deng, S. Yao, and X. Wang. 2015. An improved Medical Image Compression Technique with Lossless Region of Interest. Optik (Stuttg). 126(21): 2825-2831.
- [45] M. Sabarimalai Sur and S. Dandapat. 2014. Waveletbased Electrocardiogram Signal Compression Methods and Their Performances: A Prospective Review. Biomed. Signal Process. Control. 14(1): 73-107.
- [46] P. Suapang, M. Thongyoun, and S. Chivapreecha. 2013. Medical Image Compression and Quality Assessment. Proc. SICE Annu. Conf. 841-846.
- [47] H. Zaineldin, M. A. Elhosseini, and H. A. Ali. 2015. Image Compression Algorithms In Wireless Multimedia Sensor Networks: A Survey. Ain Shams Eng. Journal. 6(2): 481-490.
- [48] N. Karimi, S. Samavi, S. Shirani, A. Banaei, and E. Nasr-Esfahani. 2015. Real-time Lossless Compression of Microarray Images by Separate Compaction of Foreground and Background. Comput. Stand. Interfaces. 39: 34-43.
- [49] X. Cheng, H. Long, W. Chen, J. Xu, Y. Huang, and F. Li. 2015. Three-dimensional Alteration of Cervical Anterior Spinal Artery and Anterior Radicular Artery in Rat Model of Chronic Spinal Cord Compression by micro-CT. Comput. Methods Programs Biomed. 37(2): 838-848.
- [50] B. Koc, Z. Arnavut, and H. Koçak. 2015. The Pseudodistance Technique for Parallel Lossless Compression of Color-Mapped Images. Comput. Electr. Eng. 46: 456-470.
- [51] A. M. Rufai, G. Anbarjafari, and H. Demirel. 2013. Lossy Medical Image Compression Using Huffman Coding and Singular Value Decomposition. 21st Signal Process. Commun. Appl. Conf.
- [52] Y. Nian, M. He, and J. Wan. 2014. Distributed Near Lossless Compression Algorithm for Hyperspectral Images. Comput. Electr. Eng. 40(3): 1006-1014.
- [53] M. Sundaresan and E. Devika. 2012. Image Compression Using H.264 and Deflate Algorithm. Int. Conf. Pattern Recognition, Informatics Med. Eng. 242-245.
- [54] T. G. Shirsat and V. K. Bairagi. 2013. Lossless Medical Image Compression by IWT and Predictive Coding. 2013 International Conference on Energy Efficient Technologies for Sustainability. 1279-1283.
- [55] Y. Nian, M. He, and J. Wan. 2015. Lossless and Near-lossless Compression of Hyperspectral Images Based on Distributed Source Coding. Journal. Vis. Commun. Image Represent. 28: 113-119.
- [56] F. Sepehrband, M. Mortazavi, S. Ghorshi, and J. Choupan. 2011. Simple Lossless and Near-lossless Medical Image Compression based on Enhanced DPCM Transformation. IEEE Pacific RIM Conf. Commun. Comput. Signal Process. 66-72.
- [57] R. Pizzolante, B. Carpentieri, and A. Castiglione. 2013. A Secure Low Complexity Approach for Compression and Transmission of 3-D Medical Images. Proc.-2013 8th Int. Conf. Broadband, Wirel. Comput. Commun. Appl. BWCCA 2013. 387-392.
- [58] S. Amraee, N. Karimi, S. Samavi, and S. Shirani. 2011. Compression of 3D MRI Images based on Symmetry in

prediction-error Field. Proc.-IEEE Int. Conf. Multimed. Expo.

- [59] K. T. Kumari, P Vasanthi. 2013. A Secure Fast 2D-Discrete Fractional Fourier Transform Based Medical Image Compression Using Hybrid Encoding Technique. Int. Conf. Curr. Trends Eng. Technol.
- [60] M. Satti and S. Kak. 2009. Multilevel Indexed Quasigroup Encryption for Data and Speech. IEEE Trans. Broadcast. 55: 270-281.
- [61] T. Phanprasit. 2013. Compression of Medical Image Using Vector Quantization. 0-3.
- [62] S. L. Pinjare, K. Mudnaf, and S. Kumar. 2012. Distributed Arithmetic Multiplier based Artificial Neural Network Architecture for Image Compression. 2nd IEEE Int. Conf. Parallel, Distrib. Grid Comput. 135-140.
- [63] A. N. Sazish and A. Amira. 2008. An Efficient Architecture for HWT using Sparse Matrix Factorisation and DA Principles. APCCAS 2008-2008 IEEE Asia Pacific Conf. Circuits Syst. 1308-131.
- [64] M. Vucha. 2011. Design and FPGA Implementation of Systolic Array Architecture for Matrix Multiplication. 26(3): 18-22.
- [65] G. Saldana and M. Arias-Estrada. 2007. Compact FPGAbased Systolic Array Architecture Suitable for Vision Systems. J. High Perform. Syst. Archit. 3-8.
- [66] Y. Zhou and P. Shi. 2011. Distributed Arithmetic for FIR Filter Implementation on FPGA. 2011 Int. Conf. Multimed. Technol. ICMT 2011. 1(4): 294-297.
- [67] L. Wenna, G. Yang, Y. Yufeng, and G. Liqun. 2011. Medical Image Coding based on Wavelet Transform and Distributed Arithmetic Coding. 2011 Chinese Control Decis. Conf. 4159-4162.
- [68] M. Martina, G. Masera, M. R. Roch, and G. Piccinini. 2015. Result-biased distributed-arithmetic-based filter Architectures for Approximately Computing the DWT. IEEE Trans. Circuits Syst. I Regul. Pap. 62(8): 2103-2113.
- [69] A. M. Al-Haj. 2003. Fast Discrete Wavelet Transformation Using FPGAs and Distributed Arithmetic. Int. J. Appl. Sci. Eng. 1(2): 160-171.
- [70] A. Otero, Y. E. Krasteva, E. De La Torre, and T. Riesgo. 2010. Generic systolic Array for Run-time Scalable Cores. Lect. Notes Comput. Sci. 5992: 4-16.
- [71] P. C. Chandrasekhar and S. N. Reddy. 2012. FPGA Implementation of Systolic Array Architecture for 3D- DWT Optimizing Speed and Power. 2(10): 39-50.
- [72] M. M. Azadfar. 2008. Implementation of A Optimized Systolic Array Architecture for FSBMA using FPGA for Realtime Applications. IJCSNS. 8(3): 46.
- [73] M. E. Haque, A. Al Kaisan, M. R. Saniat, and A. Rahman. 2014. GPU Accelerated Fractal Image Compression for Medical Imaging in Parallel Computing Platform. 1-7.
- [74] N. Zhang, J. Wang, and Y. Chen.2010. Image parallel Processing based on GPU. 2nd International Conference on Advanced Computer Control. 367-370.
- [75] V. Akkala, P. Rajalakshmi, P. Kumar, and U. B. Desai. 2014. FPGA based ultrasound Backend System with Image Enhancement Technique. ISSNIP Biosignals Biorobotics Conf.
- [76] L. M. Russo, E. C. Pedrino, E. Kato, and V. O. Roda. 2012. Image Convolution Processing: A GPU versus FPGA Comparison. SPL 2012 - 8th Southern Programmable Logic Conference.
- [77] R. R. Osorio and J. D. Bruguera. 2006. High-throughput Architecture for H.264/AVC CABAC Compression System. IEEE Trans. Circuits Syst. Video Technol. 16(11): 1376-1384.
- [78] W. Wang, B. Guo, S. Zhang, and Q. Ye. 2009. A CABAC Accelerating Algorithm based on Adaptive Probability Estimation Update. 2nd Int. Congr. Image Signal Process. CISP'09.
- [79] P. Meng, G. R. C. Jr, R. Kastner, and D. A. Demer. 2013. GPU Accelerated Post-processing for Multifrequency Biplanar Interferometric Imaging.
- [80] H. Liu, T. Ma, S. Chen, Y. Liu, S. Wang, and Y. Jin. 2012. Development of GPU based Image Reconstruction Method For Clinical SPECT. IEEE Nucl. Sci. Symp. Med.

Imaging Conf. Rec. 3415-3418.

- [81] S. Kinouchi, T. Yamaya, E. Yoshida, H. Tashima, H. Kudo, H. Haneishi, and M. Suga. 2012. GPU-based PET Image Reconstruction Using an Accurate Geometrical System Model. *Ieee Trans. Nucl. Sci.* 59(5): 1977-1983.
- [82] J. S. Lee and T. Ebrahimi. 2012. Perceptual Video Compression: A Survey. IEEE J. Sel. Top. Signal Process. 6(6): 684-697.
- [83] A. Amselem, T. Hatsui, and M. Yamaga. 2011. Real-time Embedded Lossless Compression for Sparse Signal Data Optimized for X-Ray Free- Electron Laser Experiments. *IEEE Nuclear Science Symposium Conference Record*. 2180-2182.
- [84] I. Chiuchisan. 2013. Implementation of Medical Image Processing Algorithm on Reconfigurable Hardware. IEEE Int. Conf. E-Health Bioeng. 4-7.