Optimal Compression of Medical Images

Rafi Ullah Habib
Department of Computer Science and Information
College of Science, Majmaah University, Majmaah 11952, Saudi Arabia

Abstract—In today’s healthcare system, medical images are playing a vital role in the diagnosis. The challenges arise to the hospital management systems (HMS) are to store and communicate the large volume of medical images generated by various imaging modalities. Efficient compression of medical images is required to reduce the bit rate to increase the storage capacity and speed-up the transmission without affecting its quality. Over the past few decades, several compression standards have been proposed. In this paper, an intelligent JPEG2000 compression scheme is presented to compress the medical images efficiently. Unlike the traditional compression techniques, genetic programming (GP)-based quantization matrices are used to quantize the wavelet coefficients of the input image. Experimental results validate the usefulness of the proposed intelligent compression scheme.

Keywords—Medical images; wavelet transform; JPEG2000; genetic programming; compression; quantization

I. INTRODUCTION

Images can represent many things e.g., medical, military television, satellite or any other computer storage pictures [1]. Sampling and quantization of light intensity for creating the digital images, a massive data is produced and hence its storage and transmission becomes impractical. The solution to this problem is to compress the image and make it more practical for storage and transmission. The redundant information is reduced so that the owners and other contributors can easily increase the storage capacity and speed-up the transmission over a wired/wireless network. Medical imaging has gained immense importance in the last decades. The most appropriate solution to store and transmit medical images is to apply the lossless compression techniques that guarantee the exact reconstruction [2]. Several compression schemes have been proposed to compress digital images [3] [4, 5, 6]. The images contain three different types of redundancies: spatial, coding and psycho-visual redundancies. These redundancies are used in image compression [7]. In [8], the details of spatial, coding and psycho-visual redundancies are given. In spatial redundancy, the intensity of one pixel is calculated by the values of other pixels. In coding redundancy, the variable length code is used to match the statistics of the input image. In psycho-visual redundancy, the focus is on visual perception of the compressed images.

Two types of image compression techniques are used: 1) lossless compression, 2) lossy compression. In lossless compression, the original digital image can be obtained back without any loss of information. There are many applications e.g., medical, business, and military where any loss of information is not acceptable. Medical images are more critical and loss of information may lead to incorrect diagnosis and can be life-threatening [9, 10, 5, 11]. In 2002, in North America, an annual meeting held by the Radiology Society, Digital Imaging and Communications in Medicine (DICOM) working group-4-compression, where they announced an extension to the JPEG (Joint Photographic Expert Group) compression named as JPEG2000. The JPEG2000 compression is an extension to the JPEG compression to overcome its shortcomings. JPEG2000 is a wavelet-based compression that can highly compress images with less distortion [12]. Lossy compression used for general-purpose images where a minor loss of information is acceptable [13]. In lossy compression, the original image can be obtained from the encoded image with the loss of information where the quantization losses occur during the encoding stage [14, 15, 16]. The ratio of the original image and compressed image referred to as compression ratio (CR). The compression ratio is given in Equation 1.

\[
CR = \frac{\text{Uncompressed Image (Bytes)}}{\text{Compressed Image (Bytes)}}
\]  

(1)

The performance of the compression is measured by the difference between the input image and the reconstructed image and this difference is referred to as distortion [3]. High fidelity of the reconstructed image means the difference between the original and reconstructed image is small and vice versa. Mean square error (MSE) is the most popular method to calculate the difference between the input image and the reconstructed image [3, 17]. MSE is given in Equation 2.

\[
MSE = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (f(x, y) - g(x, y))^2
\]  

(2)

where \( M \times N \) is the size of the image, \( f(x, y) \) is the original/input image and \( g(x, y) \) is the reconstructed image. The MSE is sometimes called quantization error variance. The images with the same type of degradation are highly observed by the human eye when the MSE is smaller [18]. However, in some applications, smaller MSE does not work when different types of degradation are compared. Mostly, the researchers use PSNR (peak signal to noise ration) that is based on MSE [19]. The SNR is expressed in Equation 3.

\[
\text{PSNR} = 10 \log_{10} \left( \frac{1}{\sum_{i=1}^{M} \sum_{j=1}^{N} (f(x, y) - g(x, y))^2} \right)
\]  

(3)

The PSNR is measured in dB and show better indication of degradation in the compressed image. In this paper, an intelligent compression of medical images has been proposed. The GP based module is applied to the JPEG2000. JPEG 2000, developed in 2002 is a wavelet-based efficient compression method as compared to the DCT based JPEG compression. It was initiated in 1996 where some compression algorithms were emerged to improve the compression performance. After introducing some verification models and some other technical
contributions, JPEG2000 become an international standard [20]. A fast approach of the wavelet transform, which is also called a second-generation transform, is used. It is an integer wavelet transform (iWT), using a lifting scheme [21]. Integer discrete cosine transform (iDCT) based JPEG lossless compression of medical images. In iDCT-based compression, Watson’s standard quantization matrix of size $8 \times 8$ is used to quantize the iDCT coefficients [22, 23]. Fig. 1 shows the Watson’s perceptual quantization matrix. iDCT is a fast and efficient transformation where we do not lose any of the information. Watson’s quantization matrix although good enough for quantizing the image blocks and gives us imperceptible alteration. However, it does not provide optimum results. In the proposed work, the featured based quantization matrices are used that are generated by using the GP module. GP module is discussed in detail in Section 4.

16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\
12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\
14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\
14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\
18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\
24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\
49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\
72 & 92 & 95 & 98 & 112 & 100 & 103 & 99

Fig. 1. Watsons’ $8 \times 8$ Standard Quantization Matrix.

Rest of the paper is summarized as follow: Literature survey is presented in Section II. In Sections III and IV, the Image file formats and GP module are discussed respectively. The proposed method and experimental results are discussed in Sections V and VI respectively. Conclusion and future directions are presented in Section VII.

II. LITERATURE SURVEY

In 2014, SVD (Singular Value Decomposition) and wavelet-based lossy compression have been proposed [24]. Low singular values are neglected by using SVD and then restore the image. The restored image is then compressed again by applying the WDR (wavelet difference reduction) and an improved result in terms of visual perception, are obtained.

In 2016, Kozhemiakin et al. proposed a lossy compression of Landsat multispectral images [25]. Two facts, degree of correlation and the inherent noise along with its properties have been taken into account to compress the image. Similarly, in [26, 27], the DTT (Discrete Tchebichef Transform) based lossy compression is proposed under the JPEG standard. It provides similar performance for lossy compression like DCT based JPEG compression but it does not work for lossless. This issue has been raised by Xiao et al. in [28].

Several lossless image compression schemes have been proposed especially for medical image applications. Unlike DCT that is used in JPEG lossy compression, a fast transformation, integer DCT (iDCT) has been used to reconstruct the original image without any loss of information [29]. In [28], a lossless image compression based on DTT has been proposed that reduce the computational complexity and improve the compression rate.

In [30], a lossless compression based on segmentation-based compression is introduced. Instead of compressing the whole image, the regions of interest (ROI) zone is extracted from the image and then apply the lossless compression. This improves the compression rate without losing much information. Lossless compression of medical images has been proposed in [31]. The scheme is based on HEVC (High-Efficiency Video Coding) intra coding. The anatomical medical images are characterized by large-scale edges. HEVC intra coding compression scheme is applied to different types of medical image like CT, MRI, X-Ray images.

A compression technique for telemedicine images using the DICOM format has been proposed by [32]. The delimiter based lossless compression is applied to telemedicine images. In the encoding side, an image is converted to a row vector where the number of continuous unique elements with repetitions is evaluated. The same process is reversed and a better quality reconstructed image is obtained. They compare their proposed scheme with [33, 34], in terms of compression ratio.

Compression of CT and MRI images is proposed in [35]. The author evaluates the perceptual quality of the compressed image. In image-based diagnosis, it is very important to understand the human perception of medical image quality. The authors are focusing on the measurement of the visual perception of the compressed CT and MRI images. Similarly, a block-based lossless compression for medical images has been introduced [36]. Before applying the Huffman encoder on the coefficient to compress the medical image, the authors applied DCP (DC prediction), effective NTB (Non-transformed block) validation and truncation method.

Wavelet-Based medical image compression has been proposed [37]. The authors investigate the improvement in the JPEG2000 for volumetric medical image compression. The authors tested their technique on CT scan, MRI and ultrasound images. They develop a generic codec framework, which supports the JPEG2000 compression with its volumetric extension.

A block-based lossless compression of medical images has been proposed [38]. After decomposing the image using integer wavelet transform (IWT), the low level i.e. approximation subband is passed through the lossless Hadamard transform. Correlations inside the blocks are removed by using the Hadamard transformations. The compression ratio of medical images, as well as general-purpose images, is improved. Similarly, several schemes have been proposed in the last decade for the lossless compression of medical images [39, 40, 41].

III. IMAGE FILE TYPES

Image formats are the standards for organizing and storing images in a computer system. Images contain digital data that can be rasterized for display on the computer or any other digital display. The format specifies how the information is encoded to be stored in the storage devices. The images may store information in compressed, uncompressed or vector format. The images can be stored in so many formats. Each image file type has a specific, yet different purpose and has
advantages and disadvantages. Details of some of the most popular file formats that are used in different areas are given in Table II. The general description, proc, cons, and some of the features are explained in [42, 43].

IV. GP MODULE

Genetic programming is an intelligent search technique that is used in numerous applications. GP is a machine learning technique based on natural selection and genetics [44]. It is based on the stochastic method, where randomness plays an important role in searching and learning [45]. The quantization step in JPEG2000 compression, GP-based matrices that are based on Watsons’ standard quantization matrix, are used to quantize the wavelet coefficients of the image to be compressed. These matrices are referred to as Genetic Quantization Matrices (GQMs). Compression measurement is used in the fitness function. The focus is on the compression ratio using GP-based JPEG2000 compression for the most sensitive images i.e. medical images. An initial random population is evaluated. The best individuals are reserved and all others are deleted. The retained children make a new generation and the process continues until the termination criterion is satisfied. The block diagram for the GQM is given in Fig. 2. Different functions of the GP module used in the proposed method are as follows:

A. GP Function Set

It is the set of operators to be used in the GP module. In the proposed work, plus, minus, kozadivide, mylog, sin, and some constants are used. The matrix Wat_st is used as one of the operands. The generated GP-based matrices are based on the Watsons’ standard quantization matrix.

B. Fitness Function

The individuals are evolved by using the fitness function. The performance is evaluated by the compression ratio i.e. bits per pixel (bpp), PSNR and SSIM. The PSNR is the visual perception of the compressed image. In addition, the structural similarity index module (SSIM) has been used as a perceptual measurement where the structural similarity is measured. The bit rate is the bits produced by the encoder. Feedback is provided to the GP module that represents the fitness of the individual. The best score of the individual is indicating the best performance. Four basic arithmetic operators along with the Log and exponent are used in the fitness function. The measurement used to evaluate the performance of the proposed approach is given in the fitness function. The formula for the fitness function is given in Equation 4.

\[
\text{fitness} = \alpha_1 \times \text{bpp} + \alpha_2 \times \frac{\text{PSNR}}{50} + \alpha_3 \times \text{SSIM}
\]

The constants \(\alpha_1, \alpha_2, \alpha_3\) are the weighting parameters that are decided according to the application. The compression ratio is measured by bpp. This measurement means how many bits are required for each pixel in the compressed image. In grayscale images, one byte (8 bits) is required to store each pixel. After compression, the bits per pixel are highly reduced to improve the storage cost and transmission time. For example, bpp = 0.50 means that only 0.5 bits are required for each pixel of a compression image. Although, in medical image applications, the focus is on visual perception so the constants associated with PSNR and SSIM are given a high weight. PSNR is divided by a constant number in order to scale its value.

C. Population Initialization

In GP random population of the individuals are generated. The most commonly used methods for initializing the population are “full – and – grow” and “ramped half – and – half” methods. In the proposed scheme, the “ramped half – and – half” method is used.

D. Termination Criteria

After fulfilling one of the following criteria, the simulation is terminated.

- The fitness/target score is achieved. The fitness score is application dependent. In medical image application, the fitness score must be high.
- The fitness value repeats.
- The number of generations completed.
- The error becomes less than a pre-defined threshold. Same as fitness score, this termination criterion is also critical. In this research work, the error parameter is not used as a part of the fitness function.

E. GP Operators

In the proposed GP module, the most common GP operators, Crossover, mutation and replication are used to produce the new generation. Crossover creates the offspring by exchanging the genetic material of two individual parents. It searches for the best solution. Rapid exchanges in populations are introduced by mutation. In replication, the individual in a population is copied to the next generation. Generally, the crossover operator is kept with a high ratio. Table I shows all of the GP module settings.

<table>
<thead>
<tr>
<th>TABLE I. CGP PARAMETER SETTING</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objective</strong></td>
</tr>
<tr>
<td>Fitness</td>
</tr>
<tr>
<td>Operators</td>
</tr>
<tr>
<td>GP Operators</td>
</tr>
<tr>
<td>Operands</td>
</tr>
<tr>
<td>Selection</td>
</tr>
<tr>
<td>Expected Offspring</td>
</tr>
<tr>
<td>Initial Population</td>
</tr>
<tr>
<td>Termination Criteria</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Sampling</td>
</tr>
<tr>
<td>Survival</td>
</tr>
<tr>
<td>Short Description</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Lossless</td>
</tr>
<tr>
<td>Animation</td>
</tr>
<tr>
<td>Developed by</td>
</tr>
<tr>
<td>Common use</td>
</tr>
<tr>
<td>Proc</td>
</tr>
<tr>
<td>Cons</td>
</tr>
</tbody>
</table>
V. PROPOSED ALGORITHM

In this paper, an intelligent JPEG2000 compression has been introduced to compress medical images. In digital image compression, either DCT based JPEG compression or wavelet-based JPEG2000 compression; the coefficients are quantized by using Watson’s standard perceptual quantization matrix of size $8 \times 8$. This standard quantization matrix provides much better results for all types of images but not the optimum one. In the proposed work, the featured based quantization matrices are generated using genetic programming according to the required compression ratio. The block diagram of the GP based JPEG2000 compression scheme (encoding and decoding) are shown in Fig. 3.

In JPEG2000, to avoid the artefacts appeared in the compressed image, lifting based integer wavelet transform is used to transform the image before quantization [21].

The input image is pre-processed before decomposition. The pre-processing is image tiling, DC level shifting and colour transformation if it is a coloured image. The colour is transformed from RGB channels to Y Cr Cb channels. The input image is decomposed by using lifting based integer wavelet transform. By wavelet decomposition, both the approximation and frequency details of the image are obtained. The frequency details are horizontal, vertical and diagonal details. A reversible biorthogonal CDF 5/3 wavelet transform is used where only integer coefficients are used to avoid the quantization noise. Fig. 4 shows a two-level wavelet decomposition of an image.

The coefficients obtained by decomposing the input image, are quantized similar to the JPEG lossy compression. In this step, instead of using Watsons’ perceptual quantization matrix, GP based generated matrices (GQMs) are used to quantize the coefficients. These matrices are featured based and provide the best results in terms of compression ration without affecting the perceptual quality of the compressed image. The input parameters were tuned For example, the maximum tree level is 31, the maximum The PSNR, bpp and SSIM were also set to the maximum values. Some of the parameters were used with their default values. The numerical pre-fix expression has been generated while generating GQM. The quantization step is repeated until a fitness criterion is fulfilled. Once, GQM is generated and checked by using the fitness function, it is selected for quantizing the corresponding block of wavelet coefficients. The quantization function is to map the floating-point values to integer values that are efficiently processed by the entropy encoder.

For entropy encoding, code-blocks are the fundamental objects. After decomposition of the image, all of the subbands are divided into rectangular blocks called the precinct and then further divided into non-overlapping blocks called code blocks. Each block is then entered to the entropy encoder that encodes the blocks independently.

In entropy encoding, the coefficients in the code-block are separated into bit-planes. An example entropy encoder is shown in Fig. 5.

The contextual information of the bit-plane data is collected by coding passes. The arithmetic encoder then receives the collected data to generate compressed bit-stream. There are three types of coding passes 1) Significance propagation pass, 2) Magnitude refinement pass, 3) Clean-up pass. For each code block, a separate bit-stream in packet form is generated. If multiple layers are used to encode, the code-block bit-streams are distributed across different packets corresponding to different layers. Therefore, each layer consists of a number of bit-plane coding passes from each code-block in the tile.

![Block Diagram for GQM](image)

![Block Diagram of Compression and Reconstruction of the Image using a GP-Based JPEG2000 Scheme](image)

![Two Level Wavelet Decomposition of an Image](image)
Once the image is compressed, it can be stored and transmitted any without perceptual degradation. The reconstruction process is the inverse of the compression encoding process. The input image to the decoding side is the compressed image. First entropy decoding is applied and then de-quantized. For de-quantization, the function expression is used that was generated on the encoding side. The function expression is given before the conclusion section. The same GQMs are generated to de-quantize the coefficients. Further, the inverse lifting based integer wavelet transform is applied. The coefficients are then post-processed to get the reconstructed image.

VI. EXPERIMENTAL RESULTS

MRI image for brain tumours of size 256 × 256 has been used to test the proposed scheme [46]. The algorithm is also tested on the CT and Ultrasound scan. Experimental work has been carried in a MATLAB environment. For the GP simulation, GPLAB toolkit has been used [47, 48]. As compared to the other proposed lossless compression schemes, the proposed technique provides better results in terms of visual perception and compression ratio. The performance comparison of the proposed approach in terms of PSNR and SSIM values with [32, 33, 37] is shown in Table III. The algorithm provides a high compression ratio without affecting the perceptual similarity of images. The compression percentage is considerably increased as compared to others. The PSNR and the compression percentage depend on the compression ratio (bpp). When the bit rate is improved then PSNR, SSIM and compression percentage is degraded accordingly. The compression percentage is calculated in Equation 5.

\[
CL = (1 - \frac{\text{CompressedFileSize}}{\text{OriginalFileSize}}) \times 100
\]

The sample MRI, CT and ultrasound images used in the experimental work are given in Fig. 6.

![Figure 5: Entropy Encoder.](image)

![Figure 6: The Sample Medical Images: Brain MRI, Brain CT and Liver Ultrasound Images.](image)

Quantization matrices generated through GP module gives the optimum compression ratio. After applying the techniques on the selected set of images, the compression ratio is calculated. Decomposition of one of the images using lifting based integer wavelet transform is given as:

\[
\text{waveName} = \text{liftwave}'(\text{cdf5.3}', \text{int2uint}')
\]

\[
\begin{align*}
\text{LL1}, \text{HL1}, \text{LH1}, \text{HH1} & = \text{lwt2} (\text{double(CT Image)}, \text{waveName}) \\
\text{LL2}, \text{HL2}, \text{LH2}, \text{HH2} & = \text{lwt2} (\text{double(HH1)}, \text{waveName}) \\
\text{LL3}, \text{HL3}, \text{LH3}, \text{HH3} & = \text{lwt2} (\text{double(VV1)}, \text{waveName})
\end{align*}
\]

Where LL is the approximation subband and others are frequency subbands.

Normally, the performance of the compression technique is measured by the compression ratio. However, other measurements like perceptual similarity and compression level are also measured in the proposed approach. In Table III, it can be seen the performance of the proposed approach with the compression schemes. The available measurements with previous techniques are considered. The CL i.e. the percentage of the compression is much better as compared to [32]. The average CL of JPEG2000 compression for CT, MRI and ultrasound images is 50%. In [33], the compression is object-based, therefore, the compression rate varied according to the size of the object i.e. region of interest (ROI). SSIM, which is the structural similarity measurement, is given only in the proposed technique.

VII. CONCLUSIONS

The standard JPEG compression scheme uses a DCT transformation to describe the image. This transformation has a disadvantage non-locality that is overcome in the JPEG2000 compression scheme. JPEG2000 is wavelet-based compression where the fast approach of discrete wavelet transform has been used. The standard JPEG2000 uses the Watsons’ perceptual quantization matrix to quantize the wavelet coefficients. This standard matrix provides reasonably good results but not the optimum one. Hence in this paper, a GP based quantization matrices (GQMs) are generated before quantization. These matrices are featured based matrices where the quantization
values vary according to the input image block (region). The compression is totally application dependent where the requirements can be set as a fitness function. Usually, in medical image application, the focus is on the visual perception of the compressed image. The proposed approach can be used for any other sensitive application i.e. military application.

An exemplary numerical expression in prefix form, developed by GP, is given as:

\[
\text{`\text{minus(kozadivide(Wat_st,0.13117),kozadivide(kozadivide(minus(kozadivide(Wat_st,0.13117),kozadivide(kozadivide(Wat_st,0.82031,0.70212),kozadivide(Wat_st,mylog(sin(0.34645))),0.70212),kozadivide(Wat_st,mylog(sin(0.34645)))))0.70212),kozadivide(Wat_st,mylog(sin(0.34645))))})
\]

ACKNOWLEDGMENT

The author would like to thank Deanship of Scientific Research at Majmaah University for supporting this work under Project No. 1440-92.

REFERENCES


