Lossless Compression of Medical Images based on the Differential Probability of Images

Cheng Yao, Shuchao Chen, Jiawen Fu, Shuai Ren, Lizhi Liu, and Hongbo Chen

Abstract-Lossless compression is crucial in the remote transmission of large-scale medical image and the retainment of complete medical diagnostic information. The lossless compression method of medical image based on differential probability of image is proposed in this study. The medical image with DICOM format was decorrelated by the differential method, and the difference matrix was optimally coded by the Huffman coding method to obtain the optimal compression effect. Experimental results obtained using the new method were compared with those using Lempel-Ziv-Welch, modified run-length encoding, and block-bit allocation methods to verify its effectiveness. For 2-D medical images, the lossless compression effect of the proposed method is the best when the object region is more than 20% of the image. For 3-D medical images, the proposed method has the highest compression ratio among the control methods. The proposed method can be directly used for lossless compression of DICOM images.

Keywords—Medical image, Lossless compression, Huffman coding, DICOM.

I. INTRODUCTION

omputer tomography (CT), magnetic resonance imaging (MRI), ultrasound imaging, and other medical imaging technologies are crucial medical diagnostic technologies and means. A considerable amount of medical data is generated in a hospital. Dozens of TB medical data increments are generated in a hospital yearly, and approximately 20 TB of data account for medical images, thus giving significant importance to data storage for hospitals.

This work was supported in part by the research grants from the National Natural Science Foundation of China (Nos. 81760322 and 81460273) and the Innovation Program for College Student (No. 201810595010).

Cheng Yao is with the School of Life and Environmental Sciences, Guilin University of Electronic Technology, Guilin 541004, People's Republic of China (email: <u>2860964144@qq.com</u>)

Shuchao Chen is with the School of Life and Environmental Sciences, Guilin University of Electronic Technology, Guilin 541004, People's Republic of China (email: <u>374849631@qq.com</u>)

Jiawen Fu is with the School of Life and Environmental Sciences, Guilin University of Electronic Technology, Guilin 541004, People's Republic of China (email: <u>2451901128@qq.com</u>)

Shuai Ren is with the School of Life and Environmental Sciences, Guilin University of Electronic Technology, Guilin 541004, People's Republic of China (email: ren_shuai6@qq.com)

Lizhi Liu is with the Department of Medical Imaging, Sun Yat-Sen University Cancer Center, Guangzhou 510060, People's Republic of China (email: <u>liulizh@sysucc.org.cn</u>)

Hongbo Chen is with the School of Life and Environmental Sciences, Guilin University of Electronic Technology, Guilin 541004, People's Republic of China (corresponding author to email: <u>hongbochen@163.com</u>)

Meanwhile, with the rapid development of big data, artificial intelligence, internet and cloud technologies, telemedicine and remote transmission of medical data are crucial. The remote transmission of considerable medical image data will lead to increased requirements for network transmission performance. Hence, compressing medical image data effectively is necessary.

The main methods for image compression are lossy and lossless compressions. The lossless compression has the advantage of no image loss after decoding. Although the compression ratio of lossy compression is much higher than that of lossless compression, the original image cannot be restored error-free due to the distortion between the restored image from lossy compression and the original image [1-2]. The medical image is the original evidence of medical diagnosis, and image distortion may affect the doctor's diagnosis of disease. For medical images, the lossless compression is a basic requirement [3-4].

The principle of lossless compression is to reduce the statistical redundancy of the image data that can completely recover the original image without causing distortions. Encoding and transform are the commonly used methods for lossless compression [5]. Encoding method includes Huffman encoding, Lempel-Ziv-Welch (LZW), and run-length encoding (RLE) methods. The Huffman encoding is implemented on the basis of constructing the optimal binary tree, which can obtain the minimum average code length to obtain lossless compression [6]. The LZW method is also known as the string table compression algorithm, which is similar to the principle of colour palette. A compiled table is created based on the different grey values in the original image, and the grayscale index in the compilation table is used instead of that in the original image to reduce the size of each image pixel [7-9]. The RLE process is a series of consecutive and repeated image pixel grey values that are replaced with a grey and count values. RLE can be applied to medical images with large-area continuous shadows [10-12]. The improved RLE method (modified RLE, MRLE) [13] groups the image data several times and performs RLE to improve the compression effect. Block-bit allocation (BBA) [14] is another method for lossless compression that can obtain better performance than Huffman encoding and LZW method. A Hilbert space-filling curve was used to divide the image into several blocks. In each block, a bit number was reassigned according to its grey value.

The existing image lossless compression methods are

conducted on 8-bit grayscale images or true colour images. However, for medical images, the storage format varies with different imaging modes. For example, microscopic and colour ultrasonic images are stored in true colour format, whereas MRI and CT images are generally stored in the form of grayscale images, and the grayscale length of each pixel (voxel) is not 8 bits, but 12- bits or higher [15-16]. In this study, a difference method is proposed to remove the correlation of the image, and then the optimal encoding of the difference matrix is obtained by using Huffman encoding based on the probability of the difference image to obtain the lossless compression of the 12-bit CT or MRI images. To verify the performance of the algorithm in this study, we selected the LZW, MRLE, and BBA methods for the control experiments.

II. RELATED WORKS

imaging application demands Medical lossless or high-fidelity image compression because the lossy compression might affect the diagnostic value of the image and possibly contribute to diagnostic errors [17]. The method based on the principle of minimum rate predictors was proposed for lossless compression of volumetric sets of medical images, such as CTs or MRIs and obtained gains above 20% and 12% for 8- and 16-bit-depth contents compared with the traditional method, respectively [18]. The near-lossless compression algorithm based on adaptive spatial prediction was proposed for medical sequence images for possible diagnostic use [19] and used adaptive block size-based spatial prediction to predict blocks directly in the spatial domain as lossless Hadamard transform was used before quantization to improve the quality of the reconstructed images.

The lossless image compression method for region of interest (ROI) in medical images has attracted considerable attention from researchers recently because the diagnostically useful information called ROI is localized in a small area in the medical images. ROI was processed using the lossless compression algorithm and lossy compression elsewhere in the image [20-22]. Song et al. [23] proposed an algorithm for the lossless compression of medical images based on irregular segmentation and region-based prediction that exploited not only spatial correlation between pixels but also utilized local structure similarity, thereby resulting in efficient compression performance. Kumarganesh et al. [24] detected and segmented brain tissues with and without tumour using mathematical morphological operations. Furthermore, the brain tissues with and without tumour were compressed using lossless and lossy compression techniques, respectively. Aneja et al. [25] segmented the image into three regions, such as ROI, non-ROI, and background, using histogram-based thresholding. Moreover, the ROI and non-ROI regions were compressed by Huffman and SPIHT coding methods, respectively. The nonrelevant regions were directly converted to zero. Automatic segmentation is the key technology for lossless image compression based on ROI. Kumar et al. [26] separated the ROI (foreground) and background using the region-growing

algorithm. The ROI was encoded with low compression ratio and high bit rate, whereas the background region was encoded with high compression ratio and low bit rate.

The medical image lossy compression method can obtain high compression ratio. The lossy compressed medical images with high fidelity do not affect the diagnostic accuracy in a statistically significant way. Juliet et al. [27] used the wavelet transform to represent singularities along arbitrarily shaped curves and set partitioning in hierarchical tree encoder to encode the significant coefficients. Selvi et al. [28] presented a rapid 2-D lossy compression technique for CT and MRI images based on wavelet-based contourlet transform and binary array techniques. The diagnostic features of CT and MRI images could be precisely reproduced, as evaluated by radiologists. Parikh et al. [29] studied the method of diagnostically acceptable lossy compression and the complexity of high bit-depth medical image compression. Sophia et al. [30] presented an enhanced method for the compression of medical images using wavelet transformation, normalization, and prediction that could be adjusted to reproduce good quality images close to the original for the selected contextual area. An algorithm of multilevel compressive sensing was presented for MRI image lossy compression [31-32].

III. PROPOSED METHOD

Fig. 1 shows the flowchart for the medical image lossless compression method. The three steps of the compression process are as follows. First, differential calculation is performed on the original image to obtain decorrelation and the difference matrix. Second, the Huffman coding method is implemented on the difference matrix to obtain the optimal coding. Third, the lossless compressed data of the medical image can be obtained based on its Huffman code and grey value.



Fig. 1. Flowchart for the lossless compression

Fig. 2 shows that the differential calculation process is performed to remove the correlation of the image. Differential

calculation can reduce the coding redundancy and improve the compression ratio. Let the original image matrix be matrix I.

0	0	0	0	0	0	2
0	0	0	0	4	10	25
0	0	0	8	19	25	37
0	0	8	23	30	37	34
0	6	21	32	38	34	33
3	20	25	40	42	29	46
9	20	28	41	33	34	55



The original image matrix I

Differen	ce ca	alcul	atio	n 🖌			
	0	0	0	Ö	0	0	2
	0	0	0	0	4	6	15
	0	0	0	8	11	6	12
	0	0	8	15	7	7	-3
	0	6	15	11	6	-4	-1
	3	17	5	15	2	-13	17
	9	11	8	13	-8	1	21

Difference matrix T

			\checkmark				·				\neg			
0	0	0	0	0	0	0		0	0	0	0	0	0	2
0	0	0	0	0	0	0		0	0	0	0	4	6	15
0	0	0	0	0	0	0		0	0	0	8	11	6	12
0	0	0	0	0	0	1		0	0	8	15	7	7	3
0	0	0	0	0	1	1		0	6	15	11	6	4	1
0	0	0	0	0	1	0		3	17	5	15	2	13	17
0	0	0	0	1	0	0		9	11	8	13	8	1	21
									-					

Difference sign matrix S

Difference value matrix V



Volume 14, 2020

Each element I(m, n) represents the grey value of the m row and the n column elements in *I*. First, a new column is added at the leftmost of matrix *I* and set to zero. The difference matrix *T* is obtained by the difference calculation. The elements of *T* is calculated by T(m, n) = I(m, n + 1) - I(m, n). The difference matrix *T* is divided into value matrix *V* and sign matrix *S*. The value matrix *V* is the absolute value of the difference matrix *T*, which is encoded by Huffman coding to obtain the lossless compression. The matrix *S* is the sign of matrix *T*, which is used in the image reconstruction via image compression coding. For 3-D medical images, the same differential calculation is performed for each layer of the images.

B. Huffman Encoding

Huffman encoding is a lossless compression method that is mainly achieved by constructing the optimal binary tree and obtaining the optimal coding. The four main steps for constructing the Huffman optimal binary tree are as follows.

Step 1: The probability of all values in the difference value matrix V is calculated and organized in descending order, and then a binary tree set BT is constructed, in which a value can be considered an independent binary tree with only one node.

Step 2: Two binary trees (n0, n1) with minimum probabilities are selected and a new binary tree whose leaf nodes are the selected binary trees (n0, n1) is constructed. The leaf nodes with high and low probabilities are considered the left and right children of the new binary tree, respectively.

Step 3: The probability of the new binary tree is the sum of the probabilities of the two selected binary subtrees, which are deleted in the set *BT*. The new binary tree will be added into the set *BT*.

Step 4: Repeat steps 2 and 3 until only one binary tree remains in the set BT that can be considered the Huffman optimal binary tree.

In the Huffman optimal binary tree, each leaf node represents a value in the difference value matrix. A path from the root node to each leaf node is the coding of the corresponding value of the leaf. The encoding is 0 or 1 when the path passes through the left or right subtree, respectively. Each value in the difference value matrix V corresponds to a leaf of the Huffman optimal binary tree with a unique Huffman code. The compressed data for a given medical image can be obtained based on the Huffman coding table and the difference value matrix V.

C. Decompression

The decompression process is an inverse process of lossless

compression that includes two steps. The first step is Huffman decoding, whereby the difference value matrix V is reconstructed using the Huffman coding table and the compressed data. The second step is the inverse differential calculation, whereby the difference value and difference sign matrices are combined into the difference matrix, and then the original medical image can be reconstructed using the difference matrix by inverse difference calculation.

IV. EXPERIMENTAL RESULTS

A. Experimental Data

The experimental data in this study are provided by the Cancer Center of Sun Yat-sen University. To verify the effectiveness of the algorithm, we selected seven different types of 3-D medical images for the experiments, and Table 1 summarizes the detailed information. To evaluate the performance of the new compression algorithm, we compared the results from the proposed method with those from MRLE, BBA, and LZW under the same experimental conditions.

B. Performance Evaluation

The lossless compression method is proposed in this study. The original medical image can be restored error-free. Therefore, the image compression ratio (*CR*) is used to measure the performance of the algorithm. The CR refers to the ratio of the space size of the original image and that of the compressed image. CR = S1/S2, where S1 and S2 represent the sizes of the original image and the compressed data, respectively. In this study, compressed data includes the difference sign matrix, Huffman coding table, and compressed value. *CR* can be used as an evaluation method to objectively evaluate the performance of the proposed method in this study. The enlarged *CR* enhances the lossless compression effect.

C. Experimental Results

The experiments have been conducted on the 2-D CT and MRI images. Fig. 3 (a) shows the results of 12 representative 2-D CT images that were randomly selected from the chest CT (CT01). Fig. 3(b) shows the results of 12 representative 2-D MRI head images that were randomly selected from the head MRI (MR04). The CR of the new method is the highest for chest CT images, followed by that of MRLE, BBA, and LZW. However, for 2-D head MRI images, the results of BBA and LZW are relatively poor. The CR of the new method is higher and sometimes lower than that of the MRLE method.

	Table 1. Image information for lossless compression of medical images							
	Imaging Method	Body Part	Image Size (Voxel)	Bits of Voxel				
CT01	СТ	Chest	$512\times512\times371$	12				
CT02	CT	Abdomen	$512\times512\times201$	12				
MR01	MRI (BOLD)	Head	$68 \times 68 \times 40$	12				
MR02	MRI (DTI)	Head	$256\times256\times30$	12				
MR03	MRI (DWI)	Head	$256\times 256\times 18$	12				
MR04	MRI (T1)	Head	$256 \times 256 \times 176$	12				
MR05	MRI (T2)	Head and neck	$512 \times 512 \times 36$	12				

Table 1. Image information for lossless compression of medical images



Fig. 3. Comparison of the lossless compression ratios of 2-D medical images. (a) the results for 2-D chest CT images; (b) the results for 2-D head MR images

	MRLE	LZW	BBA	Proposed method
CT01	1.711	1.172	1.554	1.800
CT02	1.711	1.435	1.579	1.940
MR01	1.276	1.120	1.229	1.448
MR02	1.667	1.464	1.757	1.863
MR03	2.457	2.191	2.455	2.422
MR04	1.648	1.385	1.516	1.800
MR05	1.297	1.177	1.418	1.579

Table 2. Comparison of the results of lossless compression of 3-D medical images

Table 2 shows the results of the CRs for the 3-D medical images. The CR of the new method is significantly better than that of the three other methods which means that the new method is effective for lossless compression of 3-D images.

V. DISCUSSION

The increased number of values in the difference value matrix V deepens the constructed binary tree and lengthens the process of Huffman coding. The difference calculation is helpful to reduce the value used in the Huffman coding. The values that do not appear in the matrix do not participate in Huffman coding. For example, 2,223 grey values exist in the original CT image with DICOM format, thus indicating that 2,223 values would be inputted by Huffman encoding. However, only 805 values exist in the difference value matrix V, thus significantly reducing the amount of values used for encoding, shortening the length of the Huffman code, and subsequently improving the compression ratio.

For 3-D medical images, the compression ratio produced by the proposed method is significantly higher than that obtained by MRLE, BBA, and LZW coding techniques. However, for

CRNew Method=2.479

2-D head MRI images, although the compression effect of this method is better than that of BBA and LZW, sometimes the results of the MRLE method are better than that of the proposed method. The MRLE method is extremely effective for lossless compression of the images with the same grey value for large areas because the grey and numerical values replace the continuous repetition of the image grey value. However, for the image with rich details, MRLE method is not so effective. Fig. 4a shows that the CR of MRLE is higher than that of the new method for the head MRI image because of its few details and large blank areas. Fig. 4b shows that the CR of the proposed method is similar to that of the MRLE method because the object region accounts for approximately 20% of the image. However, Fig. 4c shows that the compression ratio of the proposed method is higher than that of the MRLE method when the object region accounts for more than 20% of the image. In this study, the CR of MRLE is higher than that of the new method in 48 2-D images, which is only 5.5% of the 872 2-D images used in the experiment. Hence, the object region is often large for medical images, and the lossless compression effect of the new method is better than of the MRLE method.



CRMRLE=1.639 CRNew Method=1.724



CRNew Method=1.940

VI. CONCLUSIONS

In this study, a lossless compression method for medical images based on the differential probability of the images is proposed. Decorrelation by differential method can reduce the number of values used for the Huffman encoding, thereby shortening the code length of the compression encoding and obtaining the optimal compression effect. The experimental results of the lossless compression were obtained directly in the CT and MRI images with DICOM format. The experimental results show that the proposed method can obtain better results than the LZW, MRLE, and BBA techniques. The proposed method can be applied to the compression of actual medical images to improve the transmission speed of medical images in the network and reduce the storage pressure of hospitals for enlarged medical image data.

The main works in the future include two aspects. First, we will expand the data set, especially the multi-center medical image data to verify the generalization ability of the algorithm. Second, we will study the medical image lossless compression method based on convolutional neural network to obtain better compression performance.

REFERENCES

- J. Liu, Y.H. Wang, G. Liu, et al., "Image Compression Based on Geometrical Direction," Acta Electronica Sinica, vol. 39, no. 7, PP. 1693 – 1697, 2011.
- [2] K.S. LI, Y.S. Wei, and W.S. Zhang, "WDCT Compression Based on Niching Evolutionary Algorithm," Acta Electronica Sinica, vol. 42, no. 4, pp. 809 – 814, 2014.
- [3] A.P. Made, A.A. Media, and M. Teddy, "Performance Analysis of Lossless Compression Algorithms on Medical Images," Recent Advances in Information and Communication Technology, vol.769, pp. 177 – 186, 2018.
- [4] G. Nagashree, and V. Gadre, "Lossless Medical Image Compression," Int. J. Engineering Research and Applications, vol. 4, no. 6, pp. 7 – 11, 2014.
- [5] I. Mohd, J. Tasleem, H. Misbahul, et al., "Comparative Analysis of Lossless Image Compression techniques SPHIT, JPEG-LS and Data Folding," Int. J. Engineering and Technology, vol. 19, no.3s, pp. 375 – 379, 2017.
- [6] L.S. Nashar, W.P. Tito, and L.P. Anggunmeka, "Comparative Analysis of Image Compression Using Huffman and DCT," J. Engineering and Applied Sciences, vol. 13, no.12, pp. 447 – 4452, 2018.
- [7] G. Tejeshwar, "Colour Image Steganography Using LZW Compression and Fisher-Yates Shuffle Algorithm," Int. J. Innovative Research and Development, vol. 3, no.6, pp. 54 – 61, 2014.
- [8] C.A. Sherin, P. Ketki, and J.P. Jigna, "Compressive Sensing Based Image Reconstruction," Future Internet Technologies and Trends, vol. 220, pp. 97 – 105, 2018.
- [9] P.T. Anitha, "Hybrid SOFM–MLP Neural Network for Steganalysis to Detect Stego-Contents in Corporate Emails," Automation and Autonomous Systems, vol. 10, no.1, pp. 5 – 9, 2018.

- [10] Y. Qin, Z.P. Wang, H.J. Wang, et al., "Binary image encryption in a joint transform correlator scheme by aid of run-length encoding and QR code," Optics and Laser Technology, vol. 103, pp. 93 – 98, 2018.
- [11]B. Dai, S.C. Yin, Z.S. Gao, et al, "Data Compression for Time-Stretch Imaging Based on Differential Detection and Run-Length Encoding," J. Lightwave Technology, vol. 35, no. 23, pp. 5098 – 5104, 2017.
- [12] E. Aldemir, G. Tohumoglu, and M.A. Selver, "Binary medical image compression using the volumetric runlength approach," Imaging Science Journal, vol. 67, no.3, pp. 123 –135, 2019.
- [13] J. Gao, Y. Chen, and D. Liu, "Image Lossless Compression Algorithm Based on Data Multi-Packet,", Computer Systems & Applications, vol. 19, no. 12, pp. 64 – 68, 2010.
- [14] W. Wu, Z.Z. Chen, W.S. Yu, et al., "Image Lossless Compression Algorithm Based on Block Bit Allocation," Jilin Normal University Journal (Natural Science Edition), vol. 36, no. 03, pp. 139 – 143, 2015.
- [15] S. Saurin, D.R. Parikh, K. Hari, et al., "High Bit-Depth Medical Image Compression With HEVC," IEEE J. Biomed. Health Inf., vol. 22, no. 2, pp. 552 – 560, 2018.
- [16] A. Avramović, and G. Banjac, "On predictive-based lossless compression of images with higher bit depths," Telfor Journal, vol. 4, no.2, pp. 122 – 127, 2012.
- [17] A.J. Hussain, A. Al-Fayadh, and N. Radi, "Image compression techniques: A survey in lossless and lossy algorithms," Neurocomputing, vol. 300, pp. 44-69, 2018
- [18] L. Lucas, N. Rodrigues, L. Cruz, et al., "Lossless Compression of Medical Images Using 3D Predictors," IEEE Trans. Med. Imaging., vol. 36, no. 11, pp. 2250-2260, 2017.
- [19] X. Song, Q. Huang, S. Chang, et al., "Novel Near-Lossless Compression Algorithm for Medical Sequence Images with Adaptive Block-Based Spatial Prediction," J. Digit Imaging, vol. 29, no. 6, pp. 706-715, 2016
- [20] K. Vidhya, T.R.G. Babu, and S.S. Devi, "Extraction of Abnormalities in MRI, CT, X-ray and Ultrasound Images Towards Development of Efficient Compression Algorithm," J. Scientific & Industrial Research, vol. 77, no.1, pp. 29-34, 2018.
- [21] A. Erguzen, and E. Erdal, "Medical Image Archiving System Implementation with Lossless Region of Interest and Optical Character Recognition," J. Med. Imaging Health Inf., vol. 7, no. 6, pp. 1246-1252, 2017.
- [22] Z.Y. Zuo, and X. Lan, "Deng, L.H. et al.: An Improved Medical Image Compression Technique with Lossless Region of Interest," Optik - International Journal for Light and Electron Optics, vol. 126, no. 21, pp. 2825-2831, 2015.
- [23] X. Song, Q. Huang, S. Chang, et al., "Lossless medical image compression using geometry-adaptive partitioning and least square-based prediction," Medical & Biological Engineering & Computing, vol. 56, no. 6, pp. 957-966, 2018.
- [24] S. Kumarganesh, and M. Suganthi, "An efficient approach for brain image (tissue) compression based on the position of the brain tumor," In. J. Imaging Systems and Technology, vol. 26, no.4, pp. 237-242, 2016.

- [25] R. Aneja, and A.H. Siddiqi, "Hybrid Image Compression Using Shearlet Coefficients and Region of Interest Detection," J. Med. Imaging Health Inf., vol. 6, no.2, pp. 506-517, 2016.
- [26] S.N. Kumar, F.A. Lenin, and V.P. Sebastin, "Compression of CT Images using Contextual Vector Quantization with Simulated Annealing for Telemedicine Application," J. Med. Systems, vol. 42, no.11, p. 218, 2018.
- [27] S. Juliet, E.B. Rajsingh, and K. Ezra, "A novel medical image compression using Ripplet transform," J. Real-Time Image Processing, vol. 11, no. 2, pp. 401-412, 2016.
- [28] G.U.V. Selvi, and R. Nadarajan, "CT and MRI image compression using wavelet-based contourlet transform and binary array technique," J. Real-Time Image Processing, vol. 13, pp. 261-272, 2017.
- [29] S. Parikh, D. Ruiz, H. Kalva, et al., "High Bit-Depth Medical Image Compression with HEVC," IEEE J. Biomed. Health Inf., vol. 22, no.2, pp. 552-560, 2018.
- [30] P.E. Sophia, and J. Anitha, "Contextual Medical Image Compression using Normalized Wavelet-Transform Coefficients and Prediction," IETE J. Research, vol. 63, no. 5, pp. 1-13, 2017.
- [31] T. Tashan, and M. Alazawi, "Multilevel Magnetic Resonance Imaging Compression Using Compressive Sensing," IET Image Processing, vol. 12, no.12, pp. 2186-2191, 2018.
- [32] S.P. Raja, "Multiscale transform-based secured joint efficient medical image compression encryption using symmetric key cryptography and ebcot encoding technique," International Journal of Wavelets Multiresolution and Information Processing, vol. 17, no. 5, pp. 1950034, 2019