Lossless Compression of Digital Mammography Using Fixed Block Segmentation and Pixel Grouping

Ravi kumar.M.S.
Department of IS&E.
KVGCE,Sullia, Karnataka,
perajeravi@yahoo.com

Shivaprakash Koliwad.
Department of E&C.
MCE,Hassan,Karnataka,
spksagar2006@yahoo.co.in

Dwarakish.G.S.
Department of applied mechanics & hydraulics,
NITK,Suratkal,Karnataka
dwarki_gs@yahoo.com

Abstract

A mammography is a specific type of imaging that uses low-dose x-ray system to examine breasts. This is an efficient means of early detection of breast cancer. High resolution is a common characteristic of such images. Archiving and retaining these data for at least three years is expensive, difficult and requires sophisticated data compression techniques. In this paper an efficient method is proposed for lossless compression of mammography images. After performing de-correlation of the image using two efficient predictors, the residue image is divided into 4x4 blocks. The blocks with all-zero pixels are identified using one bit code. Later, Second order of pixel grouping is employed to the remaining blocks to increase the coding efficiency. Such blocks are coded using Base offset method. Special techniques are used to save the header information.

The method is tested using 25 mammograms from the MIAS database, each having a resolution of 1024x1024 pixels with 8 bits/pixel. Experimental results indicate better compression ratio when compared to JPEG 2000, JPEG-LS, PNG and JBIG.

Keywords: JPEG 2000, JPEG-LS, Lossless compression, mammography image compression, prediction

1. Introduction

Breast cancer is the most frequent cancer in the women worldwide with 1.05 million new cases every year and represents over 20% of all malignancies among female. In India, 80,000 women affected by breast cancer in 2002. In the US, alone in 2002, more than 40,000 women died of breast cancer. 98% of women survive breast cancer if the tumor is smaller than 2 cm [1]. One of the effective methods of early diagnosis of this type of cancer is non-palpable, non-invasive mammography. Through mammogram analysis radiologists have a detection rate of 76% to 94%, which is considerably higher than 57% to 70% detection rate for a clinical breast examination [2].

A mammography is a low dose x-ray technique to acquire an image of the breast. By digitizing mammograms and applying a series of signal processing techniques to them, it is possible to diagnose breast cancer. Every year millions of mammograms are generated world-wide. Screening of mammograms in rural clinics is a growing concern, especially due to the scarcity of radiologists, infrequent visits of radiologists to rural clinics. Tele-radiology of mammography could significantly alleviate this problem, and may facilitate an early diagnosis and reduce the incidence of this disease. High resolution is a common characteristic of mammography images. They require very large storage space and transmission bandwidth. Mammography images commonly have a spatial resolution of 1024x1024 and use 16, 12 or 8 bits/pixel. An image with a resolution of 1024x1024 and using 16 bits/pixel require 2 Mbytes of storage space.

Basically image compression techniques have been classified into two categories namely lossy and lossless methods. Lossy compression methods cannot achieve exact recovery of the original image, but achieves significant compression ratio. Lossless compression techniques, as their name implies, involve no loss of information. The original data can be recovered exactly from the compressed data. If data of any kind are to be processed later to yield more information, it is important that the integrity be preserved. For example, suppose we compressed a radiological image in a lossy fashion, and the difference between the reconstruction

978-0-7695-3476-3/08 $25.00 © 2008 IEEE
DOI 10.1109/ICVGIP.2008.54

Y and the original X was visually undetected. If this image is later processed for diagnosis using computer software, the previously undetectable differences may cause the wrong diagnosis [3-5].

Lossless image compression can be always modeled as a two-stage procedure: de-correlation and entropy coding. The first stage removes spatial or inter-pixel redundancy and the second stage removes the coding redundancy. Now a days the performance of entropy coding techniques are very close to its theoretical bound, and thus more research activities concentrate on de-correlation stage. Many algorithms are used for the lossless compression of the images [6-13].

Several techniques have been proposed for the lossless compression of the digital mammography [14-18]. K.R.Namuduri et al. proposed a method using slicing of the image with variable block size segmentation. It exploits the two basic image characteristics: similarity and smoothness to achieve high compression efficiency. When compared to Lossless JPEG, 6% to 9% average improvement in the compression efficiency is obtained. Marwan Y et al. proposed fixed block based (FBB) lossless compression methods for the digital mammography. The algorithm codes blocks of pixels within the image that contain the same intensity value, thus reducing the size of the image substantially while encoding the image at the same time. FBB method alone gives small compression ratio but when used in conjunction with LZW it provides better compression ratio. A. Neekabadi et al. uses chronological sifting of prediction errors and coding the errors using Arithmetic coding. For the 50 MIAS images CSPE gives better average compression ratio than JPEG-LS and JPEG-2000. Xiaoli Li et al. uses grammar codes in that the original image is first transformed into a context free grammar from which the original data sequence can be fully reconstructed by performing parallel and recursive substitutions and then using an arithmetic algorithm to compress the context free grammar. Compression ratio achieved is promising but it involves more complicated processing and large computation time. Delaunay triangulation method is another approach. It uses geometric predictor based on irregular sampling and the Delaunay triangulation. The difference between the original and the predicted is calculated and coded using the JPEG-LS approach. The method offers lower bit rate than the JPEG-LS, JPEG-lossless, JPEG2000 and PNG. A limitation is the slow execution time. Lossless JPEG2000 and JPEG-LS are considered as the best methods for the mammography images. Lossless JPEG 2000 methods are preferred due to the wide variety of features, but are suffered from a slightly longer encoding and decoding time. Although JPEG-LS is slightly faster than JPEG 2000, it is very susceptible to channel errors since it does not perform arithmetic coder termination like the JPEG termination. This would make JPEG-LS a poor choice for telemedicine based applications [14].

This paper presents a method based on block segmentation. Specific characteristics of mammography images are well suited for our suggested methods. These characteristics include low number of edges and vast smooth regions. Initially in the de-correlation stage two predictions are used. The two error images are segmented into 4x4 blocks. Best prediction for every 4x4 block is selected. A binary file stores bit 1 or 0 to indicate which of the two predictors is used. In most of the 4x4 blocks all the 16 pixels will be zeroes. Our method encodes such blocks of 0's better than Run Length Encoding. The algorithm stores bit 1 to a binary file if the block is an all zero block; otherwise stores a 0. The Coding of 4x4 blocks with mixed gray levels is done subsequently using the Base-offset method. To improve the compression ratio a pixel grouping technique is used before encoding. To efficiently save the header information, they are first categorized and then Huffman encoded. Finally the two binary files are run-length encoded and compressed using Huffman coding method.

The organization of the paper is as follows. The proposed method is detailed in section 2. Simulation results are presented in section 3 and conclusion of the paper is given in section 4.

2. Proposed Method

The proposed method takes advantage of the smoothness characteristic of the mammography images. In the first step a prediction is made based on the two prediction rules given below in equations 1 and 2. Here A, B, C, D, E, F&G are the neighboring pixels involved in the prediction of pixel X as shown in figure 1. The coefficients of predictor 2 are empirical.

If the pixels are in first column or first row of the image, the immediate previous value is the prediction for them.

	C	B	F	G
	A	X		
	D			
	E			

Figure 1. Neighbors of pixel involved in prediction

Having two predictions is always advantageous. In spite of giving same smoothness, a predictor giving smaller minimum value will be preferred since it can be compressed better in our algorithm.

$$Pr1 = \begin{cases} \min(A, B) & \text{if } C>=\max(A, B). \\ \max(A, B) & \text{if } C<=\min(A, B). \\ A+B-C & \text{otherwise.} \end{cases} \quad (1)$$

$$Pr2=\lfloor(A*0.1+B*0.2+C*0.1+D*0.2+E*0.1+F*0.2+G*0.1)\rfloor \quad (2)$$

After the de-correlation, prediction errors are computed for the above two prediction rules. The produced errors have both positive and negative values. We found that most of the gray values in the error images are very small. In the second step, the two error images are segmented into blocks of size 4x4. Mammography images are highly correlated so that pixels inside most of the blocks are similar. This property is called Smoothness. The difference between maximum and minimum gray values d within a block is a suitable parameter to measure the smoothness. For example if d=0, then all the pixels of the block are same. If d=128, it means that the difference between maximum and minimum pixel values is 128. For every block, best prediction is selected depending on the smoothness. A binary file bin2 saves the predictor information by storing a binary number in it. If all the pixels of any of the blocks happen to be zero, those blocks are marked as all-zero blocks. A binary file bin1 saves this information. All the nonzero blocks are collected in a separate file and sent to the next stage.

At the end of the second step, the nonzero file will be an array of 4x4 blocks. We need to encode them efficiently. Any such block can be represented by the minimum gray value in the block and the difference value corresponding to the other pixels in the block with respect to the minimum value. Such a scheme is referred to as Base-offset scheme [11, 13]. The parameter d directly affects the compression ratio.

In the original Base-offset scheme, a block is coded in the following format.

Min. value in 4x4 block.	Number of bits/pixel used for coding.	16 pixels coded using $B=\lceil\log_2(d+1)\rceil$ bits.
(9 bits)	(4 bits)	(Bx16 bits)

For every block Maximum and Minimum gray values are identified and the difference d between them is computed. A modified block is obtained by subtracting minimum value from every pixel of the block. Then, every pixel in the modified block is coded using $\lceil\log_2(d+1)$ bits$\rceil$. During decoding, decoder first reads the minimum value and the number of bits/pixel used. Modified blocks are obtained by reading the number of bits as mentioned in the format. They are added to the minimum value to get the original block. Example below shows the encoding of a block.

6	6	7	3
6	7	3	3
7	6	6	3
6	6	7	3

Figure 2. A sub-block

3	3	4	0
3	4	0	0
4	3	3	0
3	3	4	0

Figure 3. Modified sub-block

For the block in figure 2 the maximum value is 7, minimum value is 3 and the difference d=4. Figure 3 shows the modified sub-block. This block is coded using $\lceil\log_2(d+1)\rceil=3$ bits. Therefore the encoding of the block requires (9+4+16x3) = 61 bits.

In our method minimum values and the number of bits/pixel used are separately stored in two files and coded using the method given in the sub-section 2.2. The two binary files bin1 and bin2 are first run length encoded and then compressed using Huffman coding.

2.1 Pixel Grouping

Higher compression ratio is achieved if d is small. Number of bits used for encoding the pixels of a block is an integer if (d+1) is an integer power of 2. Otherwise we round the number to the next higher integer. Thus, in such cases we use higher number of bits than absolutely required. This redundancy can be partially removed by grouping of adjacent pixels and then encoding. It is possible to reduce the source entropy by grouping of adjacent pixels [3]. We use the second order of grouping.

For the purpose of grouping the pixels into second order, we transform a 4x4 block into a 2x2 block in two steps using the transformations given in equations 3 and 4. The maximum value in the 4x4 block is d. In the first step a 4x2 block is obtained by multiplying one of the pixels of the group by (d+1) and adding it to the other pixel in the group. The expected maximum value for the block is $(d+1)^2-1$. In the second step a 2x2 block is obtained by second order of grouping. Here we can observe that, one of the pixels in the group is multiplied by the factor $(d+1)^2$ to account for the maximum value expected in the 4x2 block.

$$B(i, j) = A(i, j)*(d+1)+A(i, j+1) \quad \text{if } j=1$$
$$B(i, j-1) = A(i, j)*(d+1)+A(i, j+1) \quad \text{if } j=3 \quad (3)$$

Where i=1,2,3,4.

$$C(i, j) = B(i, j)*(d+1)^2+B(i+1, j) \quad \text{if } i= 1$$
$$C(i-1, j) = B(i, j)*(d+1)^2+B(i+1, j) \quad \text{if } i= 3 \tag{4}$$

Where j=1,2.

The resulting 2x2 block C is coded in the following format.

Min.value in the4x4 block. (9 bits)	No.of bits/pixel used for coding. (5 bits)	4 pixels coded using $B= \lceil \log_2(d+1)^4 \rceil$ bits. (Bx4 bits)

Figure 4 and Figure 5 shows the transformation for the block given in figure 3.

18	20
19	0
23	15
18	20

Figure 4. Sub-block after the first transformation.

469	500
593	395

Figure 5. Sub-block after the second transformation

For the block given in Fig.5 total number of bits required is 9+5+10*4=54. We can find that this algorithm saves 7 bits for the given block.

2.2 Compression of the Minimum value

The error image will have both negative and positive pixel values. The range of the minimum value is changed from the range [0, 255] to [-255, 255]. Therefore, 9 bits are required to record the minimum value.

By studying various images, we found that most of the minimum values are greater than -128 and less than 64. More concentration is observed between -8 and 8. This statistics is supported by table 1.To exploit this redundancy, we made the minimum values to belong to one of the 16 categories. For each minimum value we use 4 bit category address followed by the code for the value using the number of bits required for that category.

Table 1. Distribution of minimum values for different images.

Image #	-8 to 8	Others
001	65515	21
018	65495	41
019	65508	28
025	65501	35
027	65513	23
030	65496	40
033	65473	63
034	65478	58
036	65468	68
037	65508	28
038	65511	25
039	65470	66
040	65474	62
042	65397	139
043	65521	15

Table 2 below shows the various categories, category addresses and bits required for the elements of each category.

Table 2. Coding arrangement for the minimum Pixel values.

Category no	Category address	Elements of the category	Bits required
1	0	0	Nil
2	1	-1	Nil
3	2	+1	Nil
4	3	-2,-3	1
5	4	+2,+3	1
6	5	-4,-5	1
7	6	+4,+5	1
8	7	-6,-7	1
9	8	+6,+7	1
10	9	-8to -11	2
11	10	-12 to -15	2
12	11	-16 to -23	3
13	12	-24to -31	3
14	13	-32to-63	5
15	14	-64to-127	6
16	15	else	7

For example if the minimum value is 2, it belongs to the category 5. There are 2 elements i.e., 2 and 3 in this category. It requires only one bit for coding. Address for this category is 0100 and the code for 2 is 0. Only 5 bits are thus required instead of 9 bits. Also, if the minimum value is -120, the code uses 10 bits. From the table 1 we can see that such cases are very rare.

3. Simulation Results

In order to test our algorithm, 25 mammography images are selected from the MIAS (Mammographic Image Analysis Society) database. The MIAS is a European Society that researches mammograms and supplies over 11 countries with real world clinical data. Films taken from the UK national Breast Screening Program have been digitized to 50 micron pixel edge and representing each pixel with an 8 bit word [18]. MATLAB is the tool used for simulation. All the simulation was conducted on a 1.7GHZ processor and was supplied with the same set of 25 mammography. Each mammogram has a resolution of 1024x1024 and uses 8 bits/pixel.

Simulation results of our method are compared with that of the popular methods. The average compression results are shown in table 3. Here, compression ratio is obtained by dividing the original image size by the compressed size. The results confirm that our method gives reasonably good result compared to other methods.

4. Conclusions

By exploiting the typical characteristics of mammography images like smoothness and large number of zero blocks we have achieved better compression ratio. Comparison with other proposed approaches are given for a set of 25 high resolution digital mammograms. The preliminary simulation results are encouraging. By grouping of pixels and then coding, significant improvement in the compression ratio has been obtained. For the selected images our method performs better than JPEG-2000 & JPEG-LS. We used Second order of grouping. Higher order of grouping is computationally expensive. The difference d between the maximum and minimum pixel values in a block decides the compression ratio. By using better prediction d can be reduced. More study on this area is required.

Table 3. Comparison of Compression results

IMAGE#	PNG	JBIG	JPEG2000	JPEGLS	PROP-OSED
001	4.37:1	5.89:1	6.28:1	6.47:1	7.06:1
005	3.45:1	4.29:1	4.71:1	4.76:1	4.75:1
016	4.78:1	6.01:1	6.68:1	6.77:1	6.77:1
017	5.45:1	8.16:1	8.45:1	8.58:1	8.58:1
018	5.30:1	7.19:1	8.22:1	8.23:1	8.32:1
019	3.42:1	4.17:1	4.67:1	4.79:1	4.78:1
022	4.27:1	5.64:1	6.05:1	6.04:1	6.02:1
025	3.29:1	4.05:1	4.46:1	4.52:1	4.50:1
027	3.42:1	4.39:1	4.71:1	4.76:1	4.77:1
030	4.20:1	5.54:1	6.05:1	6.02:1	6.02:1
033	5.41:1	8.11:1	8.35:1	8.45:1	8.40:1
034	5.20:1	7.70:1	8.09:1	8.11:1	8.25:1
035	5.43:1	8.09:1	8.37:1	8.46:1	8.44:1
036	5.32:1	7.84:1	8.20:1	8.24:1	8.31:1
037	5.41:1	7.98:1	8.36:1	8.46:1	8.48:1
038	5.94:1	8.42:1	9.21:1	9.45:1	9.42:1
039	6.24:1	9.32:1	9.91:1	10.0:1	10.07:1
040	8.22:1	12.5:1	13.7:1	14.1:1	14.26:1
042	3.09:1	5.42:1	5.91:1	5.98:1	5.94:1
043	5.31:1	7.86:1	8.14:1	8.22:1	8.30:1
044	5.02:1	7.22:1	7.69:1	7.70:1	7.79:1
045	4.17:1	5.60:1	6.05:1	6.07:1	6.07:1
046	4.11:1	5.31:1	5.84:1	5.82:1	5.82:1
047	4.39:1	6.06:1	6.44:1	6.46:1	6.41:1
049	3.66:1	4.55:1	5.07:1	5.13:1	5.11:1
Aver-age	**4.75:1**	**6.69:1**	**7.18:1**	**7.28:1**	**7.31:1**

References

[1] Sunitha Saxena, Bharath Rekhi, Anju Bansal, Ashok Bagya, Chintamani & Nandagudi S. Murthy., " Clinico-morphological patterns of breast cancer including family history in a New Delhi Hospital, India- A Cross sectional study ",*World Journal of Surgical Oncology,* 2005, pp 1-8.

[2] Cahoon T.C., Sulton M.A, Bezdek J.C. “Breast Cancer Detection Using Image Processing techniques ”, *The 9th IEEE International Conference on Fuzzy Systems 2000*, Vol.2, pp 973-976.

[3] Rafael C.Gonzalez, Richard E.Woods, *Digital Image Processing,* second edition, Pearson Education Asia 2002.

[4] D.Solomon, *Data compression: The Complete Reference*, Third Edition,Springer,2004

[5] K.Sayood, *Introduction to Data Compression,* Second Edition, Morgan Kaufmann, 2000

[6] Bruno Carpentieri,Marcelo J.Weinberger,and Dadiel Seroussi,“Lossless Compression of continuos-tone images” *Proceedings of the IEEE*., Vol.88. No11, November 2000, pp 1797-1809.

[7] Marcelo.J.Weinberger, Jorma J.Rissanen, and Ronald.B.Arps, “Applications of universal Context Modeling to Lossless Compression of Gray scale images ”, *IEEE transactions on image processing,*Vol.5,No. 4, April 1996,pp 575-586.

[8] T.J. Chuang and J.C. Lin, “New approach to image encryption*,” Journal of electronic imaging*, vol.7, no.2, April 1998, pp 350-366.

[9] K.S.Ng and L.M.Cheng., “Sub-block interchange for lossless image compression ”*, IEEE transactions on Consumer electronics*, Vol. 45, No.1, Feb.1999, pp 236-241.

[10] Min shi, Shengli Xie, “A lossless image compression algorithm by combining DPCM with integer wavelet ” *IEEE 6th CAS symposium on emerging technologies,* Shanghai, China, May 31-June 2, 2004.

[11] K.R.Namuduri, N.Ranganathan, Hooman Rashedi, “SVBS: A High-resolution medical image compression algorithm using slicing with variable block size segmentation”*, IEEE proceedings of ICPR*,1996, pp 919-923.

[12] Chin-Chen Chang, Chih-Ping Hsieh, and Ju-Yuan Hsiao “A new approach to lossless image Compression ”. *Proceedings of ICCT*2003, pp 1734-38.

[13] N.Ranganathan, Steve G.Romaniuk,& Kameswara Rao Namuduri., “A lossless Image compression algorithm using variable block segmentation ”*, IEEE Transaction on Image processing*, Vol.4, N0. 10, October 95. pp 1396-1406.

[14] April Khademi and Sridhar Krishnan, “Comparison of JPEG2000 and other Lossless Compression Schemes for Digital Mammograms ”, *Proceedings of the 2005 IEEE, Engineering in medicine and Biology* 27th annual Conference, Shanghai, China, September 1-4, 2005, pp 3771-3774.

[15] Min-Mo Sung, Hee-Joung Kim, Eun-Kyung Kim, Jin-young Kwak, Jae-Kyung, and Hyung-Sik Yoo, “ Clinical Evaluation of JPEG 2000 Compression for Digital Mammography”, *IEEE Transactions on Nuclear Science,* Vol.49,No3, June 2002, pp 827-832.

[16] Marwan Y.Al-saiegh and Sridhar Krishnan , “ Fixed block-based Lossless compression of Digital Mammograms”, *Canadian conference on Electrical and Computer Engineering*, Vol.2, 2001, pp 937-942

[17] L.S. da Silva and J.Scharcanski, “A lossless compression approach for mammographic digital images based on the Delaunay triangulation”, *International conference on Image processing,* Vol.2, September 2005, pp 758-761.

[18] X.Li, S.Krishnan and N.W.Ma, “A Novel way of Lossless compression of Digital Mammograms using Grammar Codes”, *Canadian Conference on Electrical and Computer Engineering,* Vol. 4, May 2004, pp 2085-2088.

[19] A.Neekabadi, S. Samavi, N. Karimi et al., “Lossless Compression of Mammographic Images by Chronological Sifting of Prediction Errors”, *IEEE Pacific Rim Conference on Communications, Computers,& signal processing* 2007, pp 58-61

[20]www.wiau.man.ac.uk/services/MIAS/MIASweb.html