

# A New Diagnosis Loseless Compression Method for Digital Mammography Based on Multiple Arbitrary Shape ROIs Coding Framework

Ping Xu<sup>1</sup>, Yan Zuo<sup>2</sup>, Wei-Dong Xu<sup>1</sup>, and Hua-Jie Chen<sup>2</sup>

1: College of Life Information Science & Instrument Engineering, Hangzhou Dianzi University  
Hangzhou, Zhejiang, China

E-MAIL: xuping@hdu.edu.cn, temco@hdu.edu.cn

2: College of Automation, Hangzhou Dianzi University, Hangzhou, Zhejiang, China

E-MAIL: yzuo@hdu.edu.cn, chj247@hdu.edu.cn

**Abstract**— With the rapidly growing use of digital images in medical archival and communication, image compression technology, especially diagnosis lossless compression technology, plays a more and more important role for medical applications. In this thesis, a novel diagnosis lossless compression algorithm is presented for digital mammography. The mammogram is divided into breast region, pectoral muscle and background using the CAD technology. Then multiple arbitrary shape ROIs coding framework is used to compress the mammogram in which the breast region and pectoral muscle are compressed losslessly and lossily respectively, and the background can be discarded or compressed lossily as user's will. Experimental results show that the proposed method offer potential advantage in medical applications of digital mammography compression.

**Index Terms**—digital mammography; diagnosis lossless compression; CAD; ROI

## I. INTRODUCTION

In the latest twenty years, the mammogram cancer has become one of the most dangerous malignant tumors of women whose death ratio has been more than forty percent. With the rapid development of living standard and requirements of health care, doctors advise women to take mamograms twice every year. It is estimated that approximately 10%-30% of breast cancer cases are missed by radiologists[1][2]. Computer-aid diagnosis (CAD) systems have been widely used to help improve the detection precision[5][6]. Diagnosis information of digital mammography need be saved losslessly in CAD. Without compression, the size of each digital mammogram can be more than 4 MB. This brings a huge challenge to the current medical system.

Several lossless and lossy compression method have been presented to resolve this problem[7-18]. However, lossless compression has brought about at most only 4:1 compression ratio. Most lossy compression algorithms

need take into account the special clinical issue to be dealt with[8-17]. Receiver operation characteristic(ROC) analysis on lossy compression show that it is promising to use lossy techniques in medical image compression[8].

In recent years, digital mammography compression has become a research focus in the field of medical image processing. Previous studies have evaluated, with CAD systems or observer's performance studies, lossy compression in digital mammography. Good and Zheng assessed the detection of masses and clustered microcalcifications in JPEG mammograms by means of ROC study[9][10]. Kocsis found that compressed mammograms with wavelet transform algorithm at 40:1 compression ratio provided perceptually lossless compression[11]. Perlmuter et al. found no significant differences between original and compressed images at 80:1 compression ratio of 57 digital mammograms compressed with the set partitioning in hierarchical trees (SPIHT) algorithm[12]. Sung, Suryanarayanan and Penedo compared the performance of detection of primary signs of breast cancer using original images and images reconstructed at 20:1 compression ratio after JPEG2000 compression[13-15]. In a similar study, Suryanarayanan et al. obtained that JPEG2000 does not affect the CAD scheme for detecting masses, but detection cluster of microcalcifications is affected with compression at 30:1 compression ratio[16]. Idris found that false positive rate of microcalcifications was as high as 66 percent and some details with lessons ignored at 100:1 compression ratio after JPEG2000 compression, which may lead to unnecessary misdiagnosis[17]. Penedo found that detection of microcalcification clusters and masses was not failure when segmented breast region is compressed by JPEG2000 and OB-SPIHT (object-based set partitioning in hierarchical trees) at the compression ratio of 40:1 and 80:1[18].

In this paper, a new diagnosis lossless compression algorithm based on CAD technology and multiple arbitrary shape region of interests(ROIs) coding framework is proposed for digital mammography to

---

This work was supported by The National Natural Science Foundation of China (61004119, 60705016,30800248), and The Natural Science Foundation of Zhejiang Province (Y1080674)

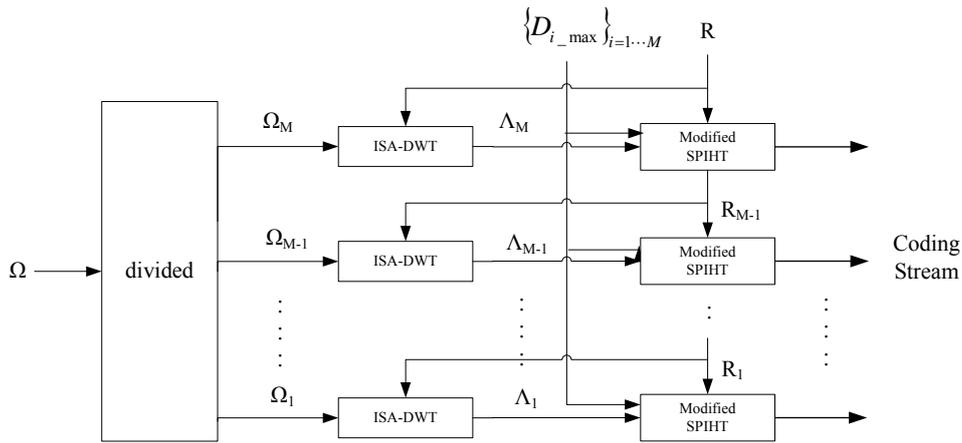


Fig.1 Flow diagram of mutiple arbitrary shape ROIs coding framework

high compression ratio while maintaining the diagnosis information losslessly.

This paper is organized with Section II presenting the proposed compression method. Section III provides the experimental results. Section IV draws the conclusion.

## II. PROPOSED METHOD

### A. Multiple, arbitrary shape ROIs coding framework

Mutiple, arbitrary shape ROIs coding framework is given as shown in Figure.1. The coding framework can be summarized as follows:

Firstly, the image plane is partitioned into multiple ROIs with different priorities which are marked by a gray mask, in which the higher gray value of the region, the higher priority of the ROIs will be;

Secondly, different ROIs are transformed into independent region bit streams by the integer-to-integer shape adaptive discrete wavelet transform(ISA-DWT)<sup>[17]</sup> in accordance with the order from highest to lowest priority respectively;

Lastly, ROIs with the same proprity are coding by the modified SPIHT algorithm<sup>[18]</sup> to generate the coding stream under their target distortion constraint and the target bit rate.

#### 1) Gray mask of ROIs:

The image plane  $\Omega$  is partitioned into  $N$  subsets  $ROI_i, i=1, \dots, N$ . The  $N$  different ROIs need to be encoded with different priorities. The set of  $N$  priorities can be expressed as

$$H = \{P_{ROI}^1, P_{ROI}^2, P_{ROI}^3, \dots, P_{ROI}^N\} \quad (1)$$

where  $P_{ROI}^i$  is the priority assigned to the  $i$ th ROI.

The priority  $P_{ROI}^i$  assigned to the  $i$ th ROI determinates the relative importance of the region in the entire image. A higher priority results in earlier isolated bit stream. Because different ROIs may have the same priority,  $M$  different priorities are produced as

$$P^1 < P^2 < P^3 < \dots < P^M \quad (2)$$

where  $P^i$  is the  $i$ th priority. A gray mask is used to show the priorities of different ROIs, in which the larger

gray value of the region means the higher priority. So the mask region of the background is black and those ROIs with the same priority have the same gray value.

#### 2) Estimation of the quality of the reconstructed ROIs:

During the encoding process, for each priority, a separate distortion constraint is employed

$$D^i \leq D_{i\_max} \quad (3)$$

Where  $D^i$  and  $D_{i\_max}$  are the distortion and the maximum allowable distortion of the  $i$ th priority, respectively. ROIs with the same priority has the same distortion constrains. ROIs with different priorities are separately represented by ISA-DWT to produce independent region bit streams. ISA-DWT maps the ROIs with the same priority into a set of transform coefficients defined over  $\Lambda_i$  with  $\Lambda_i \subseteq \Lambda$  where

$\Lambda$  is the wavelet coefficient plane corresponding to the conventional wavelet representation of image plane  $\Omega$ .

According to Parseval's theorem of energy preservation, energy in the image domain is equal to that in its wavelet domain for the orthonormal wavelet transform, and almost equal for the biorthogonal wavelet transform. So the energy in the image domain can be estimated by that in its wavelet domain. We use orthonormal (or biorthogonal) wavelet basis which yields:

$$D^i \approx D^{\Lambda_i} \quad (4)$$

where  $D^{\Lambda_i}$  is the distortion of ROIs, which can be measured by parameters such as MSE, with the  $i$ th priority measured in the wavelet domain. Therefore, we can use orthonormal (or biorthogonal) wavelet basis to obtain the estimated quality of the reconstructed ROI by measuring the region distortion without addition computational cost of obtaining the reconstructed image.

The wavelet coefficients are encoded by the modified SPITH with an additional task for the encoder to get the reconstructed wavelet coefficients. For the value of  $n$  when a coordinate is moved to the LSP, it is known that  $2^n \leq |c[i, j]| < 2^{n+1}$ , where  $c[\cdot]$  is the wavelet

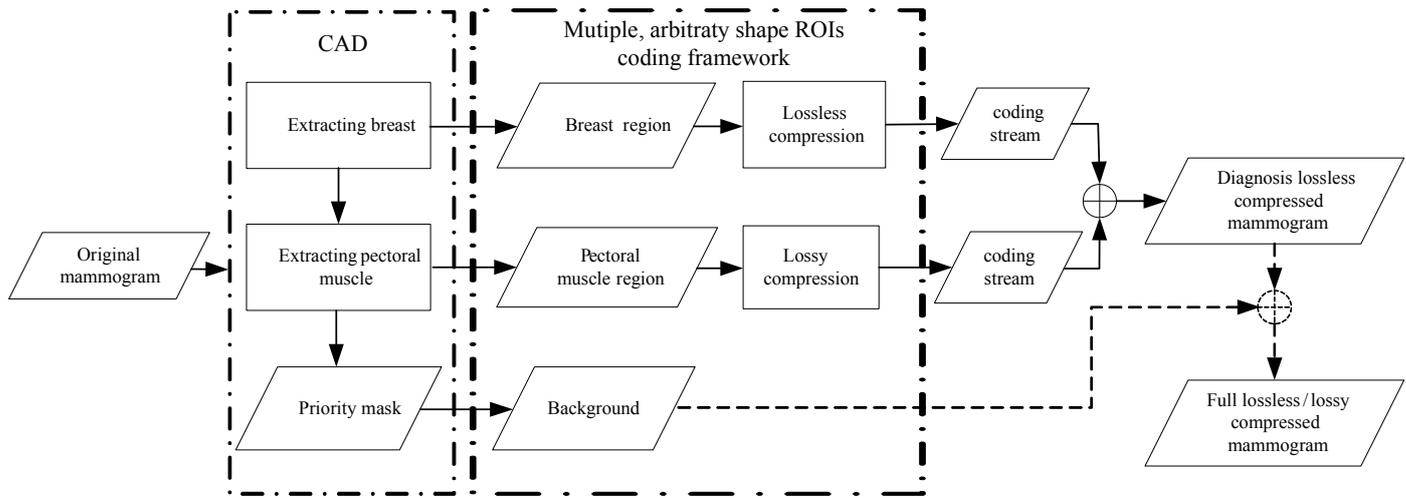


Fig.2 Flow diagram of the proposed compression scheme

coefficient of the original image.

Plus the sign bit and set the estimated reconstructed wavelet coefficient  $\hat{c}[i, j] = \pm 1.5 \times 2^n$ , where  $\hat{c}[\cdot]$  is the estimated wavelet coefficient of reconstructed image. During the refinement pass,  $\hat{c}[i, j]$  is added or subtracted by  $2^{n-1}$  when it inputs the bits of the binary representation of  $|c[i, j]|$ .

### 3) Arbitrary shape ROI lossy and lossless compression algorithm

A gray mask is induced to mark the priorities of ROIs. The priorities decide the order of the processing for coefficients of the ROIs at the encoder and the decoder. Those ROIs with the same priority are transformed as one ROI. Because of its good coding performance for lossy and lossless compression of arbitrary shape region, the ISA-DWT is used to transform arbitrary shape ROI into independent region bit streams. Then the modified SPIHT algorithm is used to coding each region to generate coding stream under its target distortion constraint and the target bit rate. So each ROI can achieve its target reconstructed quality by estimating the distortion of its wavelet coefficients in the encoder.

### B. The proposed diagnosis loseless compression method for digital mammography

The diagram of the proposed compression scheme, which mainly includes CAD and multiple arbitrary ROIs coding framework, is shown in Fig.2. The original mammogram is divided into breast region, pectoral muscle and background by the CAD technology. CAD extracts breast at first and then pectoral muscle to get the priority mask in which region of higher priority is marked with higher gray value. Then mutiple arbitrary shape ROIs coding framework is used to coding all ROIs. Specifically, the breast region is compressed losslessly by arbitrary shape ROI compression algorithm; the

pectoral muscle region is compressed lossily under given distortion constraint; the background can be discarded or compressed lossily as user's will.

### C. Extracting the breast region

The breast is extracted by a four-step method[21] as Fig.3 shows.

Step 1: Iterative thresholding is used to separate breast from the background and line scan to locate the breast;

Step 2: Column scan and utilizing the least square estimation(LSE) to detect the orientation of horizontal frame;

Step 3: Elastic thread technique is applied to cut the conglutination of the breast and the frame;

Step 4: Watershed method is used to refine the breast boundary;

### D. Segmenting the pectoral muscle

A model-based algorithm[22] is used to segment the pectoral muscle from the breast as Fig.4 shows. Two contour line models are set up to describe the intensity distribution of surrounding area of pectoral muscle. After the medialateral oblique(MLO) of each mammogram is adjusted to make pectoral muscle at the same position such as the lower right corner, a series of ROIs with different sizes are applied upon the region surrounding the pectoral muscle to obtain the optimal threshold and mean square error (MSE) curve. The model type of current pectoral muscle region can be determined according to MSE curve shape, and the optimal threshold of the pectoral muscle boundary be computed by its model's character. After thresholding, the boundary of main threshold region is extracted, then zonal hough transform is used to approach the edge of the pectoral muscle in optimal ROI. The maximum gradient of the optimal ROI can be got by line scan and the second fitting line be used to form a two-stage fitting line by zonal hough transform. At last, elastic thread and polygon approaching techniques are carried out to refine

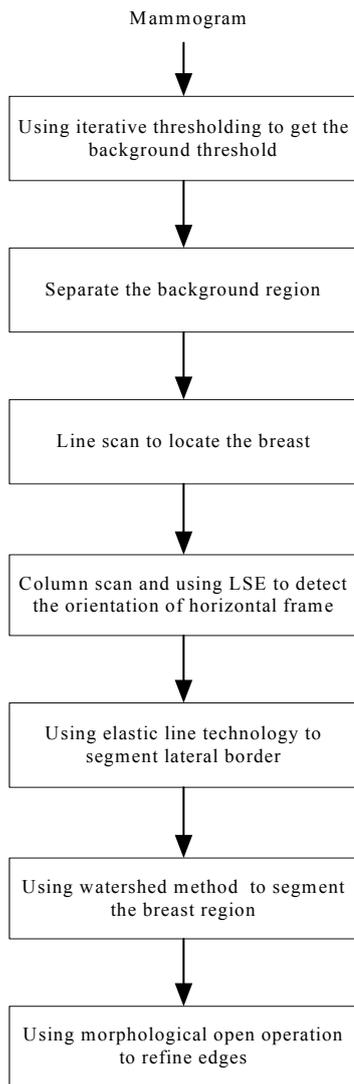


Fig.3 Flow diagram of extracting the breast region

its boundary.

The intensity distribution of breast and pectoral muscle is quite different for different mammograms. Two models with different characters are used to avoid the difficulties of segmentation pectoral muscle as the traditional approach meets.

**E. compression algorithm for breast region and pectoral muscle**

Because breast region contains important diagnosis information, it needs to be compressed losslessly. Breast region is transformed by ISA-DWT, and then encoded by the modified SPIHT to produce the ultimate coding stream, as Fig.5 shows.

For pectoral muscle region, it is not quite possible that there is diagnosis information in it. So it can be compressed lossy with relatively high compression ratio, as Fig.6 shows. It subjects to the following restrict condition:

$$D < D_{max} \tag{5}$$

Where  $D$  is the reconstructed distortion in wavelet domain and  $D_{max}$  is the target one.

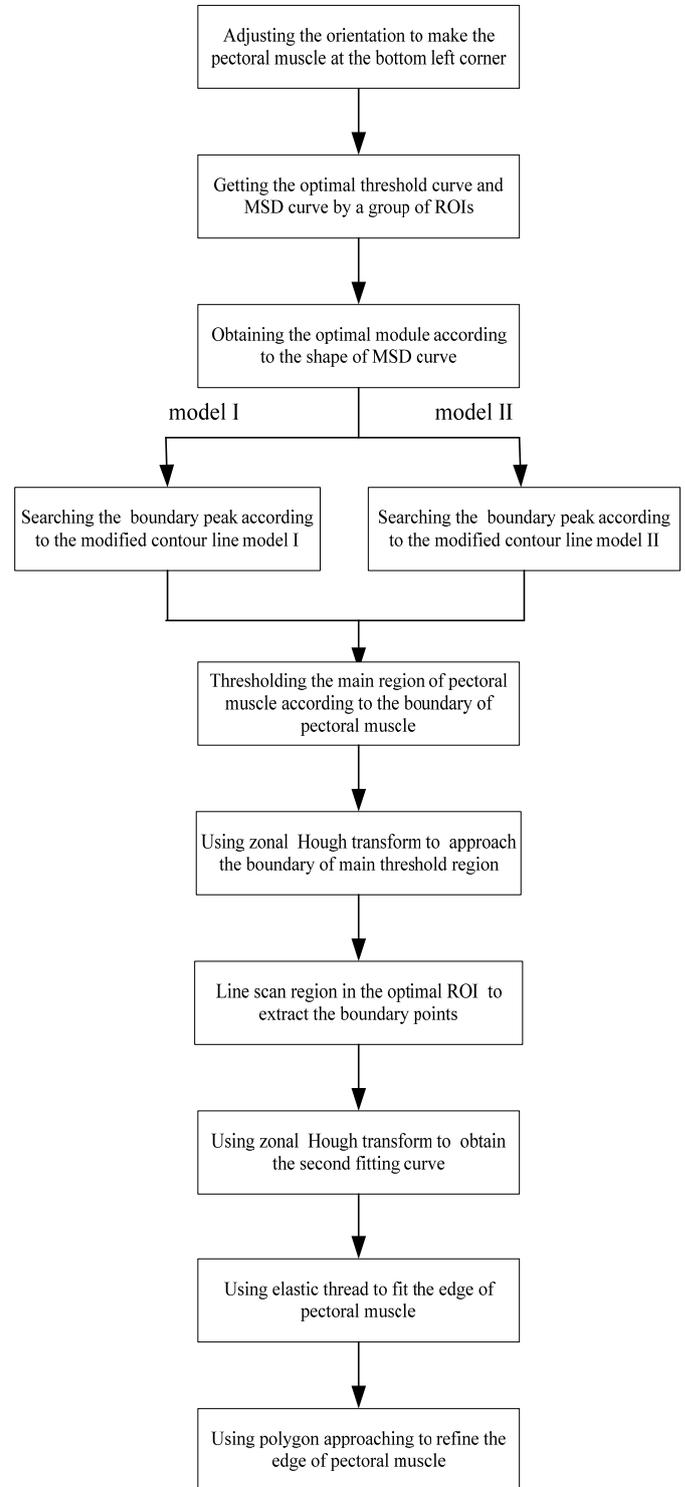


Fig.4 Flow diagram of segmenting the pectoral muscle

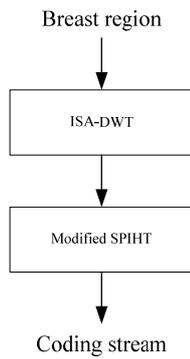


Figure 5. Flow diagram of ROI coding of breast region

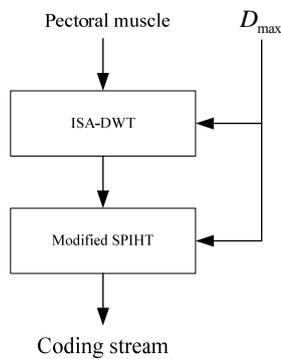


Fig.6 Flow diagram of ROI coding of pectoral muscle

### III. Experimental results

A set of mammograms are tested with the proposed method. Fig.7 shows the original mammogram of  $1760 \times 2304$  pixels at 8 bit. After extracting the breast and pectoral muscle, the mask is obtained to show different regions, in which larger gray value means higher priority as Fig.8 shows. The integer-to-integer 5/3 and 9/7-F wavelet are used in ISA-DWT. The distortion is evaluated by means of PSNR in wavelet domain. The target PSNR is set as 36dB and the background is discarded completely. Table I gives the compression ratio comparison for different compression methods, where lossless compression is adopted in JPEG2000. It is evident that integer-to-integer 5/3 and 9/7-F wavelet can achieve better compression efficiency than JPEG2000 and 5/3 wavelet shows the best one. Fig.9 shows the reconstructed mammogram for integer-to-integer 5/3 wavelet. The ROI with lessons marked in Fig.6 is magnified in Fig.10(a). The reconstructed region with integer-to-integer 5/3 wavelet shows completely the same effect as the original one in Fig.10, which means that the proposed method can achieve quite good compression performance while not bringing any diagnosis information loss.

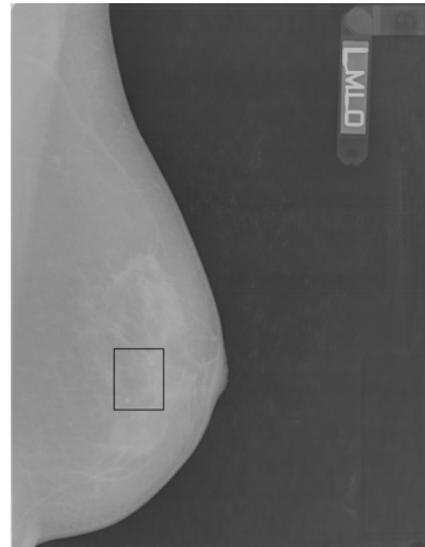


Fig.7 The original mammogram

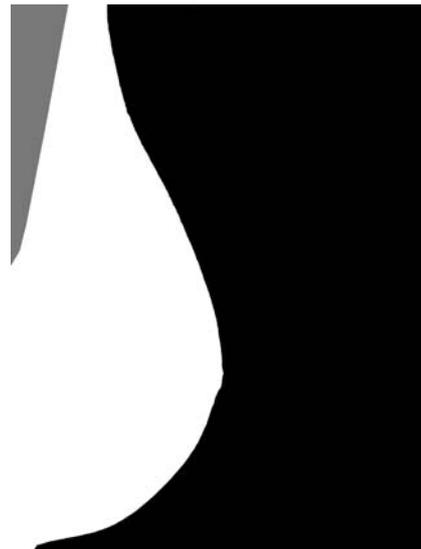


Fig.8 mask after extracting the breast and pectoral muscle

TABLE I. COMPRESSION RATIO COMPARISON

	JPEG2000	IWT-5/3	IWT-9/7-F
Compression ratio	2.838	6.412	4.641

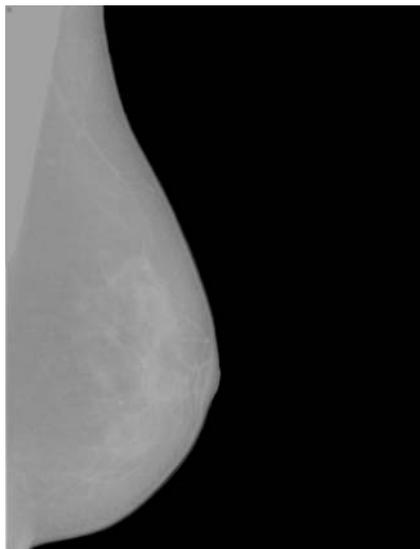
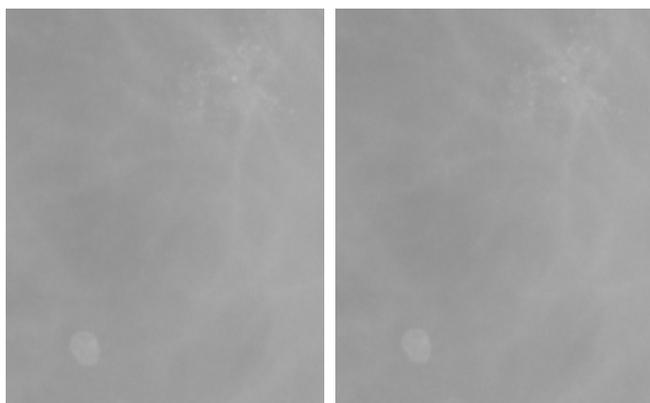


Fig.9 the reconstructed mammogram



(a) The original region

(b) the reconstructed region

Fig.10 ROI containing lessons

#### IV. CONCLUSION

In this paper, the CAD technology and multiple arbitrary shape ROIs compression framework are combined to realize diagnosis lossless compression for digital mammography. Primary experimental results show quite good compression efficiency while keeping the diagnosis information lossless, especially for integer-to-integer  $5/3$  wavelet. Motivated by the results obtained here, our next study is going to carry out the clinical evaluation of the proposed method for digital mammography.

#### REFERENCES

[1] R. G. Bird, T. W. Wallace, and B. C. Yankaskas, Analysis of cancers missed at screening mammography, *Radiology*, vol. 184, pp. 613–617, 1992.

[2] H. Burhenne, L. Burhenne, F. Goldberg, T. Hislop, A. J. Worth, P. M. Rebbeck, and L. Kan, Interval breast cancers in the screening mammography program of British Columbia: Analysis and classification, *Am. J. Roentgenol.*,

vol. 162, pp. 1067–1071, 1994.

- [3] K. Doi, H. MacMahon, S. Katsuragawa, R. M. Nishikawa, and Y. Jiang, Computer-aided diagnosis in radiology: Potential and pitfall, *Eur. J. Radiol.*, vol. 31, pp. 97–109, 1999.
- [4] H. Li, K. J. Liu, and S. Lo, Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms, *IEEE Trans. Med. Imag.*, vol. 16, no. 6, pp. 785–798, Dec. 1997.
- [5] Bradley J. Erickson M D, Irreversible Compression of Medical Images, *J. Digit. Imag.*, 15(1): 5-14, 2002.
- [6] Ishigaki T, Sakuma S, Ikeda M et al, Clinical evaluation of irreversible image compression: Analysis of chest imaging with computed radiography, *J. Radiol.*, 175:739–743, 1990.
- [7] Slone R M, Foos D H, Whiting B R et al, Assessment of visually lossless irreversible image compression: Comparison of three methods by using an image comparison workstation, *Radiology-Computer Applications*, 215(2):543–553, 2000.
- [8] Penedo M, Pearlman W A, Tahoces P G et al, Region-based wavelet coding methods for digital mammography. *IEEE Trans Med Imaging*, 22:1288–1296, 2003.
- [9] Good W F, Sumkin J H, Ganott M et al, Detection of masses and clustered microcalcifications on data compressed mammograms: an observer performance study. *AJR Am J Roentgenol*, 175:1573–1576, 2000.
- [10] Zheng B, Sumkin J H, Good W F et al, Applying computer-assisted detection schemes to digitized mammograms after JPEG data compression: an assessment. *Acad Radiol*, 7:595–602, 2000.
- [11] Kocsis O, Costaridou L, Varaki L et al, Visually lossless threshold determination for microcalcification detection in wavelet compressed mammograms. *Eur Radiol*, 13: 2390–2396, 2003.
- [12] Perlmutter S, Cosman P, Gray Ret et al, Image quality in lossy compressed digital mammograms. *Signal Process*, 59:189–210, 1997.
- [13] Sung M M, Kim H J, Kim E K et al, Clinical evaluation of JPEG2000 compression for digital mammography. *IEEE Trans Nucl Sci*, 49: 827–832, 2002.
- [14] Suryanarayanan S, Karellas A, Vedantham S et al, A perceptual evaluation of JPEG 2000 image compression for digital mammography: contrast detail characteristics. *J Digit Imaging*, 17:64–70, 2004.
- [15] Penedo M, Carreira J M, Tahoces P G et al, Effects of JPEG2000 data compression on an automated system for detecting clustered microcalcifications in digital mammograms, *IEEE Trans Inf Technol Biomed*, 10(2):354–361, 2006.
- [16] Suryanarayanan S, Karellas A, Vedantham S et al, Detection of Simulated Lesions on Data compressed Digital Mammograms, *Radiology*, 236:31–36, 2005.
- [17] Idris F M, AlZubaidi N I, Detection of breast cancer in the JPEG2000 domain, *Trans Eng. Comput Technol*, 8:1305-5313, 2005.
- [18] Penedo M, Souto M, Tahoces P G et al, FROC evaluation of JPEG2000 and object-based SPIHT lossy compression on digitized mammograms, *Radiology*, 237: 450–457, 2005.
- [19] Ahmed Abu-Hajar and Ravi Sankar, “Integer-to-integer shape adaptive wavelet transform for region of interest image coding”, *Digital Signal Processing Workshop*, pp:94 – 97, Oct. 2002.
- [20] Zhongmin Liu et al., “Cascaded differential and wavelet compression of chromosome images”, *IEEE Transaction on Biomedical Engineering*, 49(4), pp. 372-383, Apr. 2002.

- [21] Xu W, Study on computer-aided diagnosis of ammograms, Ph.D. dissertation, Dept. Biomedical Engineering, Zhejiang Univ, Hangzhou, China, 2006. (in Chinese)
- [22] Xu W, Xia S, A model based algorithm to segment the pectoral muscle in mammograms, in proc IEEE Int. Neural Networks & Signal Processing Conf, pp.1163-1169, Dec 2003.

**Ping Xu** was born in 1978. He received the Ph.D degree in control theory and control engineering from Zhejiang University, Hangzhou, China, in 2006. He has been an Associate professor of College of Life Information Science & Instrument Engineering of Hangzhou Dianzi University since 2008. His current research interest include image compression, compressive sensing, and biomedical image processing.

**Yan Zuo** was born in 1980. She received the Ph.D degree in control theory and control engineering from Shanghai Jiao Tong University, Shanghai, China, in 2007. She has been an Associate professor of College of automation of Hangzhou Dianzi University since 2009.

**Wei-Dong Xu** was born in 1977. He received the Ph.D degree in biomedical engineering from Zhejiang University, Hangzhou, China, in 2006. He has been an Associate professor of College of Life Information Science & Instrument Engineering of Hangzhou Dianzi University since 2008. His current research interest include computer-aided diagnosis, and biomedical image processing.

**Hua-Jie Chen** was born in 1978. He received the Ph.D degree in control theory and control engineering from Zhejiang University, Hangzhou, China, in 2006. He has been an Associate professor of College of Life Information Science & Instrument Engineering of Hangzhou Dianzi University since 2008. His current research interest include image processing and computer vision.