

# Breast Lump Recognition Algorithms Based on Ultrasound Radio-Frequency Signals

YAN Yu<sup>1,2</sup>, CAI Xiaowei<sup>2</sup>, ZHU Wei<sup>2</sup>, CAI Runqiu<sup>2</sup>, WU Yiyun<sup>2\*</sup>

1. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, P.R. China;

2. Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing 210029, P.R. China

(Received 22 June 2018; revised 7 August 2018; accepted 9 January 2019)

**Abstract:** A method for evaluating the benign and malignant breast tumors based on radio-frequency (RF) data was explored by extracting the characteristic parameters of breast ultrasound RF signals. The breast biopsy data were used as the reference data for judging the lump benign or malignant. The extracted ultrasound RF data were reconstructed and segmented by computer aided method to obtain the breast tumor region of interest (ROI) and its characteristic parameters (entropy and standard deviation). The characteristic parameters were statistically analyzed to evaluate the relationship between characteristic parameters and benign or malignant breast tumors. The results indicate the entropy and standard deviation of normal region is much higher than that of lump region, which shows that the standard deviation and entropy characteristic parameters of ultrasonic RF signals are meaningful in the diagnosis of breast tumors. The proposed method provides a new direction for computer-aided diagnosis of benign and malignant breast tumors.

**Key words:** ultrasound radio-frequency signal; breast lump; signal processing; standard deviation; entropy

**CLC number:** TN957.52    **Document code:**    **Article ID:** 1005-1120(2019)04-0635-06

## 0 Introduction

Breast tumor is an extremely common tumor lesion that jeopardizes the health of women. So far, no studies have shown the exact cause of breast tumors. Early diagnosis and grading are crucial for breast tumors. Breast ultrasound and breast biopsy are the main means to evaluate the benign and malignant breast tumors. Although breast biopsy is the gold standard for tumor evaluation, positive predictive values (PPVs) are achieved lowly (38.0%<sup>[1]</sup>) when using the BI-RADS classification criteria to identify benign and malignant, while the sensitivity is very high (98.4%<sup>[1]</sup>). This situation results in a large number of benign biopsy results (60%—80%<sup>[1]</sup>), and unnecessary biopsy increases the pain and expense of the patient. Therefore, it is necessary to improve the ultrasound diagnosis for enhancing specificity and reducing unnecessary biopsy.

Ultrasound imaging uses the ultrasonic beam to scan the human body. It distinguishes different tissues by receiving, processing and reflecting the difference in acoustic characteristics of the reflected signals. Images of various tissue components of the human body are obtained in this way. Ultrasound images also show the organ interface and the internal tissue. With the use of ultrasound images, doctors can identify the lesions from normal tissues, and conduct further diagnosis<sup>[2-3]</sup>. The diagnosis of breast tumor is based on B-mode ultrasonic images, which are transformed from original radio-frequency signals. To get a clearer image, the region of interest (ROI) area is enhanced and the background noise is suppressed during the transforming period, but some usable information is lost inevitably<sup>[4]</sup> (as shown in Fig. 1). On the other hand, breast ultra-

\*Corresponding author, E-mail address: wuyi425@sina.com.

**How to cite this article:** YAN Yu, CAI Xiaowei, ZHU Wei, et al. Breast Lump Recognition Algorithms Based on Ultrasound Radio-Frequency Signals [J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2019, 36(4): 635-640.

<http://dx.doi.org/10.16356/j.1005-1120.2019.04.010>

sound diagnosis mainly comes from the doctor's qualitative evaluation of breast tissue and muscle tissue structure around the breast. The diagnosis results are related to the doctor's experience, level, status and other factors. So it is likely to be influenced by subjective factors<sup>[5]</sup>. According to statistics, the accuracy of conventional two-dimensional ultrasound diagnosis is only from 69% to 75%<sup>[6]</sup>. In this case, an auxiliary diagnosis based on the complete original signal is urgently needed. In order to minimize the influence of doctors' subjectivity and imaging factors, we should find a more accurate and quantitative method for ultrasonic diagnosis. It is necessary to perform characteristic analysis from the source of the signal, that is, the ultrasound radio-frequency (RF) signal.

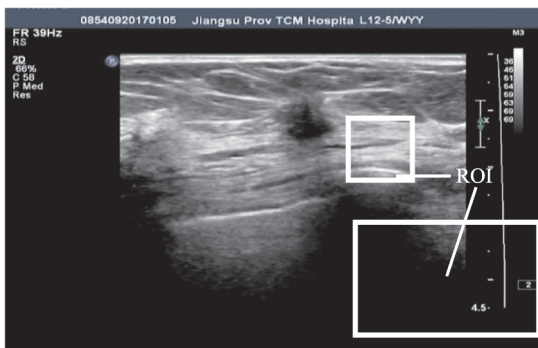


Fig.1 A typical ultrasonic image of breast tumor

RF signal is the original signal that reflects the tissue irradiated by ultrasound<sup>[7]</sup>. By extracting and processing RF signals, the differences between normal and diseased areas can be quickly and accurately summarized, making diagnosis easier. Masotti et al.<sup>[8]</sup> used RF ultrasound local characterization parameter analysis method to determine the prostate tissue in vitro. Moradi et al.<sup>[9]</sup> proposed a new method for the diagnosis of prostate cancer using ultrasound RF signals. Liu et al.<sup>[10]</sup> used time domain, frequency domain analysis method and pattern recognition technology to explore a new quantitative method for fatty liver grading. The recognition rate of fatty liver reached 86.7%. However, it is rarely involved in the identification of benign and malignant breast tumors from an RF perspective. Considering that the breast ultrasound diagnosis relies on the naked eye and it is strongly affected by subjective

judgment, breast biopsy data are used as reference data for judging benign and malignant tumors in this paper. The characteristic parameters of breast ultrasound RF signal are extracted by computer aided method. We quantitatively evaluate the relationship between characteristic parameters and benign or malignant lesions, with analyzing and evaluating the algorithms based on the experimental results.

## 1 Methods

### 1.1 Feature algorithm extraction

Processing the ultrasound RF signals requires extraction of the RF signal in a readable manner. The extraction of the ultrasound RF signals is carried out with the use of the open background database of the ultrasonic diagnostic system VINNO 70. We wrote a RF data function program and ran the program with MATLAB. The derived hexadecimal data are converted and rearranged into a decimal numerical matrix. Each of the numerical points corresponds to the pixel of the same row and column coordinates in the ultrasound image, and the magnitude of the numerical point reflects the intrinsic gray value of the corresponding ultrasound image pixel without image enhancement, that is, the formed numerical matrix is a two-dimensional representation of the ultrasound image without processing<sup>[11]</sup>. Since there is no image enhancement, the ultrasound image indicated by the numerical matrix is rarely used in the current ultrasound clinical diagnosis, especially the physician's ultrasound image visual judgment. But for signal analysis and processing, the numerical matrix, containing the most comprehensive and undistorted ultrasound information, is suitable for algorithm compilation to get the clinical significance of the parameters. The subject of ultrasound feature extraction of breast tumors is the original RF data of normal breast tissue and breast tumor. In this study, the RF signal and image are combined to extract the characteristic parameters of the RF signal in the breast tumor ROI. The specific process is shown in Fig.2.

First, the ROI is selected. To ensure its accuracy, the artificial segmentation method used in this

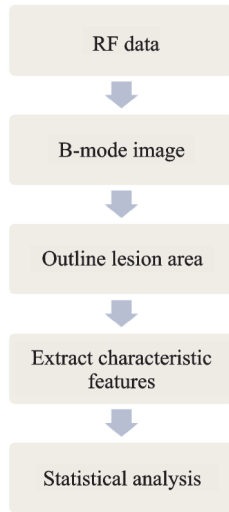


Fig.2 Schematic diagram of RF data processing

study is a relatively simple target object-based contour extraction method. The image segmentation tool used is the ImageCorp software, and it is written by the research group based on MATLAB R2016b, as shown in Fig.3.

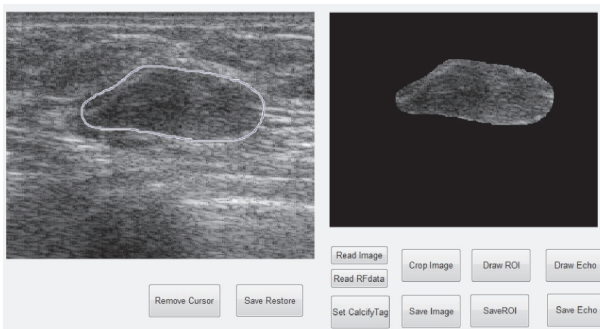


Fig.3 Interface of software “ImageCorp”

The ROI is saved into a 100\*100 standard size for later feature extraction. The sampling error is  $\sqrt{1/n}$ , where  $n$  is the sampling number. In this case  $n$  is  $100*100=10\ 000$ . The sampling error is 1%, making the statistical reliability 99%. Then we use MATLAB to extract the standard deviations and entropies of the lump area and normal area for statistical analysis.

## 1.2 Standard deviation and entropy

The standard deviation  $\sigma$  represents the degree of dispersion between the sample elements in the mathematical statistics, and is defined as the arithmetic mean square of the differences of the elements  $x_i$  and the sample mean  $\mu$  in the sample. The size of

the standard deviation  $\sigma$  reflects the uniformity of the sample elements, shown as

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

The entropy of the statistical sample  $S$  is defined as the sum total of all possible states of the sample elements of the system, the entropy value  $N_n$  is the number of possible states in the  $n$ th state, and for signal analysis and processing the entropy reflects the number of elements with the same numerical value in the numerical matrix. The entropy values reflect the degrees of confusion in statistical samples, shown as

$$S = \sum_n N_n \ln(N_n)$$

In the analysis of the standard deviation and entropy of the sample, the distal region and the region adjacent to the skin surface are removed. Due to the severe attenuation of the ultrasonic signal at the distal end of the far field, the signal noise accumulation, and the acoustic interference in the vicinity of the skin surface area and the actual incomplete coupling, the data of the far field area and the skin adjacent area will increase the uncertainty of the statistical fluctuation. In this paper, the far field region end and the adjacent region are defined as the values of 1 500 and 100, so the value of the region locates in the vertical coordinate range  $[100, 1\ 500]$ , as shown in Fig.4.

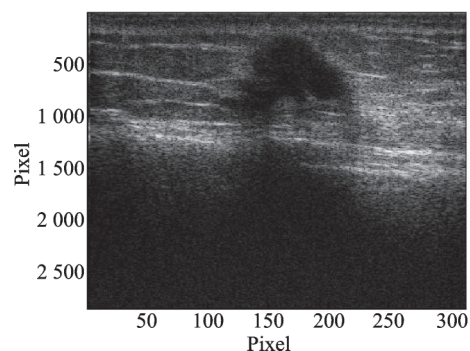


Fig.4 Image representation of the ultrasound RF data matrix

## 2 Results

In this study, the RF data of 100 patients with breast tumors are collected. All patients had biopsy diagnosis as gold standard, including 50 benign tu-

mors and 50 malignant tumors, respectively. Fig. 5 shows a typical malignant breast tumor and its pathological photomicrograph.

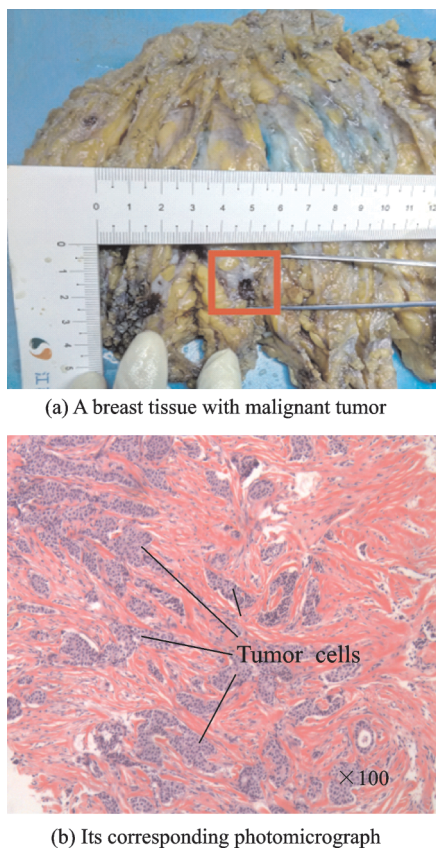


Fig.5 A typical malignant breast tumor and its photomicrograph

The benign mass area (normal area) is determined by the boundary recognition algorithm<sup>[12]</sup>. The algorithm outlines the mass, uses the judgment program to read the boundary coordinates, and intercepts the lump area. The standard deviation (Stdv) of benign and malignant tumors is then calculated, respectively. The standard deviation and entropy of benign lump and malignant tumor areas are shown in Figs.6, 7.

From Fig.6 we can see that the standard deviations of normal tissues are higher than those of tumor areas. The standard deviations of the normal areas are concentrated in the range of [163, 178], and the standard deviations of the tumor areas are concentrated in the range of [60, 143]. From Fig.7 we can see clearly that the entropies of normal tissues are higher than those of tumor areas. The entropies of the normal areas are concentrated in the range of

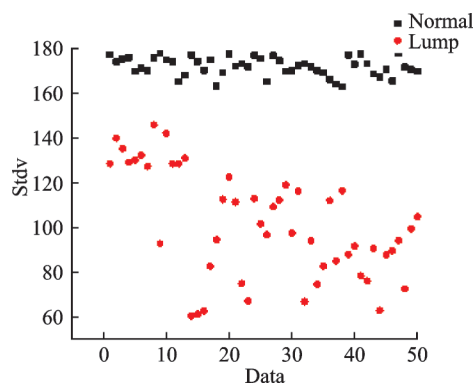


Fig.6 Standard deviations of the ultrasound RF data

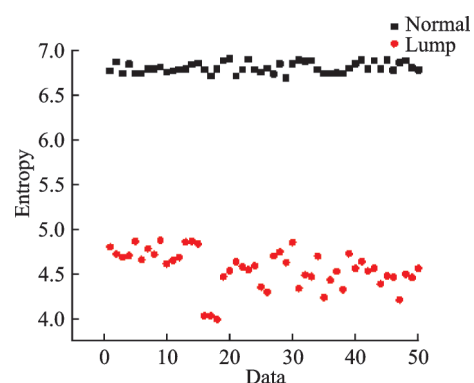


Fig.7 Entropies of the ultrasound RF data

[6.7, 6.9], and the entropies of the tumor areas are concentrated in the range of [4.0, 4.9].

In summary, when analyzing the ultrasonic RF signal, it is found that by comparing the standard deviation and entropy of the normal tissue and the lesion area, the difference between the benign and malignant tumors can be clearly distinguished. Both benign and malignant tumors have a certain interval concentration on the standard deviation and entropy parameters. The value of the normal area is higher than that of tumor area, which is consistent with the morphological characteristics of the cells. At the same time, we found that when comparing the standard deviation between the tumor area and the normal area, the average difference is 40%, which has a high clinical value. In the entropy algorithm, the values of normal regions are more concentrated and the values of tumor regions are more dispersed. Both algorithms have their own advantages and disadvantages, all of which help to aid diagnosis.

### 3 Discussion

Considering that the structure of human body is



extremely complex and different organs and different tissues have different acoustic characteristics, there is also a significant difference between the acoustic impedances of the normal cell areas and the lesion areas. So when the ultrasound is exposed to human body, the two areas will have different refraction, scattering and attenuation coefficients. These lead to the differences in the images, which mainly come from two points: (1) The normal cells have regular morphologies and sizes. The ratio of the nucleus to the whole cell is stable, and in each nucleus, chromatins distribute evenly. In contrast, tumor cell volume significantly increases and the morphologies of tumor nucleuses vary from each other. There may be multinuclear, binuclear, macronucleus and other circumstances, because tumor cells are out of normal system regular control. So the nucleo-cytoplasmic ratio of tumor cells is much larger (Fig.5(b)). (2) The normal cells have contact inhibition. Normal cells locate inside human bodies neatly, once they contact with each other to a dense structure they will no longer split. But tumor cells have no contact inhibition, and they amplify disorderly and rapidly in vivo. Even if tumor cells accumulate into groups, they can still continue to split, leading to the formation of tumors.

These characteristics of tumors are consistent with the characteristics of the data presented in this paper. Because of the loss of contact inhibition of the tumor cells, they show a more agglomerated state. Therefore the tumor area occupies larger space than the normal area, which makes the response of the whole region to the ultrasound signal more consistent, and the standard deviations and the entropies of tumor regions much smaller in the images.

## 4 Conclusions

In this study, we propose a method to identify the benign and malignant tumors by extracting the characteristic parameters: standard deviation and entropy of the ultrasound RF signal of the breast tumor. The experimental results show that the standard deviation and entropy can clearly distinguish be-

nign and malignant tumors. It can be used as an early screening and auxiliary judgment for breast tumors, because the ultrasound RF signal is quite different between normal tissue and diseased area. In the subsequent practical clinical application, RF parameter generation program is considered to be added to the ultrasound image scanning module. The ROI region is manually selected during the image scanning process, and the RF parameter is directly generated by the ultrasound diagnostic instrument to assist the doctor in diagnosis.

## References

- [1] STAVROS A T, THICKMAN D, RAPP C L, et al. Solid breast nodules: Use of sonography to distinguish between benign and malignant lesions[J]. *Radiology*, 1995, 196(1): 123-134.
- [2] TZAGKARAKIS G, ACHIM A, TSAKALIDES P, et al. Joint reconstruction of compressively sensed ultrasound RF echoes by exploiting temporal correlations [C]// *IEEE 10th International Symposium on Biomedical Imaging: From Nano to Macro*. [S.l.]: IEEE, 2013: 632-635.
- [3] LI G K, LIN Y, DENG W, et al. Computer aided diagnosis of fatty liver ultrasonic images based on support vector machine [C]// *30th Annual International IEEE EMBS Conference*. [S.l.]: IEEE, 2008: 4768-4771.
- [4] XIE X Q, LUO Y, QUAN J R, et al. SD rats' fatty liver tissue classification based on ultrasound radiofrequency signal [J]. *Advances in CSIE*, 2012, 168: 643-647.
- [5] LIU L B, LANE J W, QUAN Y L. Radar attenuation tomography using the centroid frequency downshift method [J]. *Journal of Applied Geophysics*, 1998, 40: 105-116.
- [6] CHENG P W, SHEN C C, LI P C. Ultrasound RF channel data compression for implementation of a software-based array imaging system [C]// *IEEE International Ultrasonics Symposium Proceedings*. [S.l.]: IEEE, 2011: 1423-1426.
- [7] BERG W A, COSGROVE D O, DORÉ C J, et al. Shear-wave elastography improves the specificity of breast US: the BE1 multinational study of 939 masses[J]. *Radiology*, 2012, 262(2): 435.
- [8] MASOTTI L, BIAGI E, GRANCI S, et al. Tissue differentiation based on radiofrequency echographic signal local spectral content[C]// *Ultrasonics, 2003 IEEE Symposium on*. [S.l.]: IEEE, 2003, 1:

1030-1033.

- [9] MORADI M, ABOLMAESUMI P, ISOTALO P A, et al. A new feature for detection of prostate cancer based on RF ultrasound echo signals[C]// 2006 IEEE Ultrasonics Symposium. Vancouver, BC, Canada: IEEE, 2006: 10.1109/ULTSYM.2006.530.
- [10] LIU Z D, LUO Y, LIN J L. Quantizing and grading of fatty liver based on ultrasonic RF signals[J]. Journal of Sichuan University (Engineering Science Edition), 2011, 43:160-164.
- [11] American College of Radiology. Breast imaging reporting and data system: BI-RADS: ultrasound[M]. 4th ed. Reston, VA, USA: American College of Radiology, 2003.
- [12] HUANG Q, YANG F, LIU L, et al. Automatic segmentation of breast lesions for interaction in ultrasonic computer-aided diagnosis [J]. Information Sciences An International Journal, 2015, 314(C): 293-310.

**Acknowledgements** This work was financially supported by the National Natural Science Foundation of China (No. 61703201), and the National Natural Science Foundation of Jiangsu Province (No. BK20170765).

**Authors** Mr. YAN Yu received the M. S. degree in software engineering from Nanjing University of Science & Technology, Nanjing, China, in 2009. He is pursuing the Ph.D. degree in the School of Computer Science and Engineering, Nanjing University of Science & Technology, Nanjing, China. His current research is focused on artificial intel-

ligence and pattern recognition.

Mr. CAI Xiaowei received the M. S. degree in biomedical engineering from Southeast University, Nanjing, China, in 2012. His research is focused on the bioeffects of microbubbles.

Mr. ZHU Wei received the M. S. degree in biomedical engineering from Nanjing Medical University, Nanjing, China, in 2013. His research is focused on medical electronics.

Mr. CAI Runqiu received the M. S. degree in physical and medical engineering from University College London, London, England, in 2013. His research is focused on ultrasound signal processing.

Mr. WU Yiyun received the M. S. degree in combination of traditional Chinese and western medicine from Nanjing University of TCM, Nanjing, China, in 2014. He is also the director of Ultrasound Medical Department of Affiliated Hospital of Nanjing University of TCM. His research is focused on clinical ultrasound diagnosis.

**Author contributions** Mr. YAN Yu designed the study and compiled the models. Mr. CAI Xiaowei interpreted the results and wrote the manuscript. Mr. ZHU Wei conducted the analysis. Mr. CAI Runqiu contributed data for the analysis. Mr. WU Yiyun contributed to the discussion and background of the study. All authors commented on the manuscript draft and approved the submission.

**Competing interests** The authors declare no competing interests.

(Production Editor: Zhang Huangqun)