Automated breast cancer detection and classification using ultrasound images: A survey

H.D. Cheng\textsuperscript{a,\*}, Juan Shan\textsuperscript{a}, Wen Ju\textsuperscript{a}, Yanhui Guo\textsuperscript{a}, Ling Zhang\textsuperscript{b}

\textsuperscript{a}Department of Computer Science, Utah State University, Logan, UT 84322, USA
\textsuperscript{b}School of Mathematics and System Science, Shandong University, China

1. Introduction

Breast cancer is the second leading cause of death for women all over the world. Since the cause of the disease remains unknown, early detection and diagnosis is the key for breast cancer control, and it can increase the success of treatment, save lives and reduce cost. Ultrasound imaging is one of the most frequently used diagnosis tools to detect and classify abnormalities of the breast. In order to eliminate the operator dependency and improve the diagnostic accuracy, computer-aided diagnosis (CAD) system is a valuable and beneficial means for breast cancer detection and classification. Generally, a CAD system consists of four stages: preprocessing, segmentation, feature extraction and selection, and classification. In this paper, the approaches used in these stages are summarized and their advantages and disadvantages are discussed. The performance evaluation of CAD system is investigated as well.

Keywords: CAD (computer-aided diagnosis), Automated breast cancer detection and classification, Ultrasound (US) imaging, Feature extraction and selection, Classifiers

ARTICLE INFO

Article history:
Received 9 September 2008
Received in revised form 16 April 2009
Accepted 14 May 2009

Abstract

Breast cancer is the second leading cause of death for women all over the world. Since the cause of the disease remains unknown, early detection and diagnosis is the key for breast cancer control, and it can increase the success of treatment, save lives and reduce cost. Ultrasound imaging is one of the most frequently used diagnosis tools to detect and classify abnormalities of the breast. In order to eliminate the operator dependency and improve the diagnostic accuracy, computer-aided diagnosis (CAD) system is a valuable and beneficial means for breast cancer detection and classification. Generally, a CAD system consists of four stages: preprocessing, segmentation, feature extraction and selection, and classification. In this paper, the approaches used in these stages are summarized and their advantages and disadvantages are discussed. The performance evaluation of CAD system is investigated as well.

© 2009 Elsevier Ltd. All rights reserved.
and to increase the diagnostic sensitivity and specificity [13,16]. As much as 65–90% of the biopsies turned out to be benign, therefore, a crucial goal of breast cancer CAD systems is to distinguish benign and malignant lesions to reduce FPs. Many techniques such as linear discriminant analysis (LDA), support vector machine (SVM) and artificial neural network (ANN) [5,10,17,18,20] have been studied for mass detection and classification. Most of the CAD systems need a large number of samples to construct the models or rules, but [22] proposed a novel diagnosis system requiring very few samples.

This survey focuses on summarizing the approaches for breast cancer detection and classification utilizing BUS images. Generally, the ultrasound CAD systems for breast cancer detection involve four stages as shown in Fig. 1.

1. **Image preprocessing:** The major limitations of BUS imaging are the low contrast and interference with speckle [3]. The task of image preprocessing is to enhance the image and to reduce speckle without destroying the important features of BUS images for diagnosis.

2. **Image segmentation:** Image segmentation divides the image into non-overlapping regions, and it will separate the objects from the background. The regions of interest (ROIs) will be allocated for feature extraction.

3. **Feature extraction and selection:** This step is to find a feature set of breast cancer lesions that can accurately distinguish lesion/non-lesion or benign/malignant. The feature space could be very large and complex, so extracting and selecting the most effective features is very important. Most of the reported effective features are listed in Table 4.

4. **Classification:** Based on the selected features, the suspicious regions will be classified as lesion/non-lesion or benign/malignant by various classification methods. The commonly used classifiers are discussed in Section 5.

Some CAD systems do not have image preprocessing and image segmentation components. In such a framework, only some texture features obtained directly from the images or ROIs are used as inputs of classifiers [13,16,20,22]. The advantage of such CAD system is its simple structure and fast processing speed, and disadvantage is that the features extracted directly from ROIs may not provide robust and accurate performance.

At last, we need to measure the performance of CAD systems. There is no a benchmark database of US images for comparing the performance of the algorithms/CAD systems, and it makes the evaluation of different CAD systems very difficult or even impossible. This indicates the necessity to build a benchmark BUS image base accessible to the public.

2. Preprocessing

The preprocessing of BUS images consists of speckle reduction and image enhancement. Speckle is a form of multiplicative noise generated by a number of scatterers with random phase within the resolution cell of ultrasound beam [33,34]. Ref. [29] has demonstrated that the k-distribution is a good model for the amplitude distribution of the received signal. A more generalized statistical model, the homodyned k-distribution, has been analyzed in [30]. It combined the features of the k-distribution and Rice distribution to better account for the statistics of the signal. To detect speckles, the parameters for the speckles should be estimated first. The speckle parameters of the k-distribution model can be estimated based on the moments [31]. An iterative method using the statistics of ultrasound signal is proposed to find the parameters of the homodyned k-distribution model [32]. Speckle makes the visual observation and interpretation difficult. Therefore, removing speckle without destroying important features for diagnosis is critical. Some speckle reduction techniques only work well on additive noise, and logarithmic compression is often employed to convert multiplicative noise into additive noise [33]. Image enhancement is used to improve the quality of low contrast images. We will review speckle reduction and image enhancement separately, however, many techniques can achieve both goals at the same time.

2.1. Speckle reduction

Speckle reduction techniques are classified into three groups: (1) filtering techniques [34–59]; (2) wavelet domain techniques [60–79]; and (3) compounding approaches [80–83].

2.1.1. Filtering techniques

Most filters are traditional techniques in spatial domain and can be categorized as linear and nonlinear filters.

2.1.1.1. Linear filters.

2.1.1.1.1. Mean filter. The mean filter [41,42] replaces each pixel by the average value of the intensities in its neighborhood. It can locally reduce the variance and is easy to implement. It has the effect of smoothing and blurring the image, and is optimal for additive Gaussian noise in the sense of mean square error. Speckled image is a multiplicative model with non-Gaussian noise, and therefore, the simple mean filter is not effective in this case.

2.1.1.1.2. Adaptive mean filter. In order to alleviate the blurring effect, the adaptive mean filters [35–40] have been proposed to achieve a balance between straightforward averaging (in homogeneous regions) and all-pass filtering (where edges exist). They adapt to the properties of the image locally and selectively remove speckles from different parts of the image. They use local image statistics such as mean, variance and spatial correlation to effectively detect and preserve edges and features. The speckle noise is removed by replacing it with a local mean value. The adaptive mean filters
outperform mean filters, and generally reduce speckles while preserving the edges.

The Lee [36], Kuan [35] and Frost [37] filters are well-known examples of adaptive mean filters. They are based on the multiplicative speckle model which can be represented as

\[ I(t) = R(t)u(t) \]

where \( t = (x, y) \) is the coordinates of the current pixel, \( R(t) \) denotes the intensity of the ideal image without speckle, \( \bar{I}(t) \) is the observed image intensity and \( u(t) \) is the speckle with mean \( \bar{u} \) and variance \( \sigma_u^2 \).

The Lee and Kuan filters have the general form:

\[ \tilde{R}(t) = \bar{I}(t) + [I(t) - \bar{I}(t)] \times W(t) \]

where \( \tilde{R}(t) \) is the output of the filter, \( \bar{I}(t) \) and \( \sigma_I^2(t) \) are the local mean and variance of the intensities within the filter window, respectively. \( W(t) \) denotes the coefficient function of the adaptive filter. The Lee filter has a coefficient function defined as \[38\]

\[ W(t) = 1 - \frac{\sigma_I^2(t)}{1 + \sigma_I^2(t)} \quad \text{where} \quad \sigma_I^2(t) = \frac{\sigma_I^2}{\bar{z}^2}, \quad \bar{z}^2(t) = \frac{\sigma_I^2(t)}{\bar{I}(t)} \]

\( \sigma_I^2 \) and \( \bar{z} \) are the intensity variance and mean over a homogeneous area of the image, respectively. Thus \( C_d \) can be considered as a constant for a given image.

The coefficient function of Kuan filter is defined as \[38\]

\[ W(t) = 1 - \frac{C_d^2}{1 + C_d^2} \]

From the formulations of the Lee and Kuan filters, the difference between them is only a constant factor \( 1 + C_d^2 \). So we could discuss them together. In the homogeneous regions, \( C_d^2 \to \sigma_I^2 \). Thus the value of \( W(t) \) approaches 0, which makes the filters act like a mean filter. On the other hand, in the areas where edges exist, \( C_d^2 \to \infty \), and the value of \( W(t) \) approaches 1, which tends to preserve the originally observed image and makes the filters behave like an all-pass filter.

The Frost filter can be represented as

\[ \tilde{R}(x, y) = \sum_i \sum_j m(x + i, y + j) \times I(x + i, y + j) \]

where \( i \) and \( j \) are the indices of the filter window and \( m \) is the weighting function \[38\]:

\[ m(x + i, y + j) = K_0 \exp[-K C_d^2(t) |i^2 + j^2|] \quad \text{where} \quad t = (x, y) \]

where \( K_0 \) is a normalizing constant, and \( K \) is a damping factor. Frost filter has similar attributes as Lee and Kuan filters. The damping factor \( K \) is chosen such that when in homogeneous regions, \( K C_d^2(t) \) approaches 0. Thus the value of \( m(x+i, y+j) \) approaches 1, which makes the filter act like a mean filter; in the areas where edges exist, \( K C_d^2(t) \) becomes so large that the value of \( m(x+i, y+j) \) approaches 0 for the pixels surrounding \( (x, y) \), and remains 1 for the pixel \( (x, y) \). This makes the filter behave like an all-pass filter preserving the originally observed image.

The classical Lee, Kuan and Frost filters are only reliable in a bounded field. Ref. \[38\] enhanced the Lee and Frost filters by dividing the image into three classes according to the local coefficient of variation \( C(t) \). When \( C(t) \) is below a lower threshold, pure averaging is used. All-pass filter is performed when \( C(t) \) is above a higher threshold. When \( C(t) \) exists in between the two thresholds, standard Lee and Frost filters are applied. The enhanced filters adequately average the homogeneous areas and preserve the edges better than the standard filters. Ref. \[45\] proposed a directional adaptive mean filter based on 2D texture homogeneity histogram to suppress speckles in ultrasound images.

2.1.1.2. Nonlinear filters.

2.1.1.2.1. Order-statistic filter. Order-statistic filters are particularly effective in reducing noise whose probability density function has significant tail. The median filter \[34,41,42\] is a special case of order-statistic filters. It preserves the edge sharpness and produces less blurring than mean filter. Specially, it is effective when image is affected by impulsive noise. Several researchers have experimented with adaptive median filters which outperform the median filters \[43,46\]. An adaptive weighted median filter (AWMF) was developed to achieve maximum speckle reduction in uniform areas and to preserve the edges and features \[47\]. The weighted median of a sequence \( X_i \) is defined as the pure median of the extended sequence which is generated by repeating each term \( X_i \) by \( w_i \) times. Here, \( w_i \) are the corresponding weight coefficients. The weight coefficients are adjusted according to the local statistics as

\[ w(i,j) = [w(K + 1, K + 1) - bX_i]/\bar{u} \]

where \( b \) is a scaling constant, \( \bar{u} \) and \( \sigma_u^2 \) are the local mean and variance of the \((2K+1)^2\) window, \( d \) is the distance of the point \((i, j)\) from the center of the window at \((K+1, K+1)\), and \( |X| \) denotes the nearest integer to \( x \) if \( x \) is positive, or zero if \( x \) is negative. However, this algorithm uses an operator which can cause difficulties in enhancing image features such as line segments. To overcome this drawback, \[48\] applied a bank of oriented one-dimensional median filters and retained at each point the largest value among all the filter bank outputs. The directional median filter suppresses speckle noise while retaining the structure of the image, particularly, the thin bright streaks.

2.1.1.2.2. MAP filter. Maximum a posteriori (MAP) filter \[41, 49,50\] estimates an unobserved signal \( x \) by maximizing Bayes theorem:

\[ f(x|z) = f(z|x)f(x)/f(z) \]

where \( f(x|z) \) is the a posteriori probability density function, \( f(x) \) is the denoised original signal model, \( f(z|x) \) is the maximum likelihood term and \( f(z) \) is the model of the observed data. To utilize it, the priori knowledge of the probability density function (PDF) of the image is needed. The PDF is assumed to be Gaussian distributed in \[49\]. Ref. \[51\] modified the MAP filter in \[49\] by assuming a gamma and symmetric beta distribution.

Comparisons of standard de-speckle filters (Fig. 2) with the adaptive MAP filter for ultrasound images are presented in \[41\]. MAP Gauss denotes the MAP filter with Gaussian distribution assigned to the original image, and MAP Pearlman Gauss denotes the MAP Gauss filter using the adaptive windowing proposed in \[154\]. MDb1 is the filter using Daubechies wavelets with Db1 basis. Contrast to speckle ratio (CRS) is used to evaluate the performance and the results show that the MAP Pearlman Gauss is the best among the filters being compared.

2.1.1.2.3. Nonlinear diffusion. Nonlinear diffusion is actually an adaptive filter, where the direction and strength of the diffusion are controlled by an edge detection function. It can remove speckles and enhance edges at the same time. It removes speckles by modifying the image via solving a partial differential equation (PDE).

Ref. \[52\] proposed the nonlinear PDE for smoothing image in a continuous domain:

\[
\begin{cases}
\frac{\partial I}{\partial t} = div[c(|\nabla I|) \cdot \nabla I] \\
I(t = 0) = I_0
\end{cases}
\]

where \( \nabla \) is the gradient operator, \( div \) is the divergence operator, \( |I| \) denotes the magnitude, \( c(x) \) is the diffusion coefficient and \( I_0 \) is the
original image. The diffusion coefficient function $c(x)$ should monotonically decrease, the diffusion decreases as the gradient strength increases, and the diffusion is stopped across edges.

Anisotropic diffusion (AD) performs well with additive Gaussian noise. However, edge estimation using gradient operator makes it difficult to handle multiplicative noisy image. In order to eliminate such disadvantage, Speckle Reducing Anisotropic Diffusion (SRAD) is proposed particularly for envelope US images without logarithmic compression [53]. In SRAD, the instantaneous coefficient of variation servers as the edge detector in speckled images. The function exhibits high values at edges and produces low values in homogeneous regions. Thus, it ensures the mean-preserving behavior in the homogeneous regions, and edge-preserving and edge-enhancing at the edges. Ref. [44] extended 2D SRAD to 3D SRAD to process 3D ultrasound images more efficiently. Nonlinear coherence enhancement diffusion (NCD) is another method for handling speckle noise [33]. Unlike SRAD, NCD works with the US images after logarithmic compression. It combines three different models. According to speckle extent and image anisotropy, the NCD model changes progressively from isotropic diffusion through anisotropic coherent diffusion to, finally, mean curvature motion.

Diffusion stick method for speckle suppression is proposed in [54]. It divides the traditional rectangular filter kernel into a set of asymmetric sticks of variable orientations. The weighted sum of averages along each stick is used to produce the filtered images. Ref. [55] developed a multi-resolution median-anisotropic diffusion interactive method. It used a two resolution level process to convert speckle to quasi-impulsive noise. Then the low-resolution images are processed by the median-anisotropic diffusion interactive algorithm. The weighted sum of averages along each stick is used to produce the filtered images. Ref. [56] developed a multi-resolution median-anisotropic diffusion interactive method. It used a two resolution level process to convert speckle to quasi-impulsive noise. Then the low-resolution images are processed by the median-anisotropic diffusion interactive algorithm. The computational cost is lower than that of the conventional AD schemes. In [56] a hybrid method is designed based on median filtering, improved AD filtering and isotropic diffusion filtering. The gradient matrix is analyzed, and the thresholds are chosen by experiments. The hybrid method combines the three filtering methods for three different grayscale gradient ranges, respectively.

The advantage of nonlinear diffusion is that the speckle reduction is carried out directionally by the edge function and the edges are enhanced. The disadvantage is that it relies on the diffusion flux to iteratively eliminate the small variations caused by noise, and to preserve the large variations caused by edges. For the multiplicative noisy image, however, the general signal/noise relationship no longer exists, since the variations caused by noise may be larger than those caused by signal. Hence, it is not suitable for this case.

2.1.2. Wavelet domain techniques

The discrete wavelet transform (DWT) translates the image into an approximation sub-band consisting of the scale coefficients and a set of detail sub-bands at different resolutions and resolution scales composed of the wavelet coefficients [72]. DWT provides an appropriate basis for separating the noise from an image. As the wavelet transform is good at energy compaction, the small coefficients more likely represent noise, and large coefficients represent important image features. The coefficients representing features tend to persist across the scales and form spatially connected clusters within each sub-band. These properties make DWT attractive for denoising. A number of wavelet-based despeckling techniques have been developed. The general procedure is: (1) calculate the discrete wavelet transform; (2) remove noise by changing the wavelet coefficients; and (3) apply the inverse wavelet transform (IDWT) to construct the despeckled image. The techniques are grouped as: (1) wavelet shrinkage; (2) wavelet despeckling under Bayesian framework; and (3) wavelet filtering and diffusion.

2.1.2.1. Wavelet shrinkage.

The wavelet shrinkage is based on thresholding the wavelet coefficients. It suppresses the coefficients representing noise while retains the coefficients that more likely representing image features. It is usually performed using one of the two dominant thresholding schemes: hard thresholding and soft thresholding.

Suppose the image in wavelet domain is represented as

$$o = s + n$$

where $o$ is the observed wavelet coefficients, $s$ is the noise-free component and $n$ is the additive noise. The wavelet shrinkage estimator can be represented as

$$\hat{s} = Ho$$

**Fig. 2.** CRS values of phantom image using various filters [41].
where \( H \) denotes shrinkage factor. For the classical wavelet thresholding rules, a threshold value \( T \) is defined and \( H \) is specified as follows. For hard thresholding

\[
H = \begin{cases} 
0 & \text{if } |o| < T \\
1 & \text{if } |o| \geq T 
\end{cases}
\]

For soft thresholding

\[
H = \begin{cases} 
0 & \text{if } |o| < T \\
1 - \frac{|o|}{T} & \text{if } |o| \geq T 
\end{cases}
\]

where \( |o| \) denotes the absolute value of \( o \). Soft thresholding provides smoothness when applied to an image while hard thresholding preserves the features of an image. Applications of hard and soft thresholding can be found in [61, 62, 73–75]. Most of them have focused on developing the best uniform threshold. Adaptive thresholding which makes threshold values adaptive to the spatially changing statistics of the images has attracted more attention [76–77]. Adaptive thresholding improves the performance by incorporating additional local information such as the identification of edge into the despeckling algorithm.

The drawback of thresholding methods is that the choice of the threshold is usually done in an ad hoc manner.

2.1.2.2. Wavelet despeckling under Bayesian framework. An alternate approach to the standard thresholding technique is employing Bayesian rules [60, 63–71]. It relies on the knowledge of the wavelet coefficient statistics. This approach assumes that \( p \) is a random variable with a given prior probability density function. Given a set of wavelet coefficients \( q \), the goal is to find the Bayes risk estimator \( \hat{p} \) minimizing the conditional risk, which is the cost averaged over the conditional distribution of \( p \) (denoted as \( P_{p|q}(p|q) \)):

\[
\hat{p}(q) = \operatorname{arg\,min}_p \int_\mathbb{L} \int L[p, \hat{p}(q)]P_{p|q}(p|q) \, dp 
\]

where \( L \) is a cost function to be specified.

In [67], the two-sided generalized Nakagami distribution (GND) is used to model the speckle wavelet coefficients, and the wavelet coefficients are modeled by the generalized Gaussian distribution (GGD). Combining these statistics priors with the Bayesian MAP (maximum a posteriori) criterion, the algorithm can deal with either envelope speckle image or log-compressed image. Ref. [69] designed both the minimum absolute error (MAE) and the MAP estimators for alpha-stable signal mixed in Gaussian noise. Ref. [64] extended the approach [69] in two aspects: the use of bivariate alpha-stable distributions to model the signal wavelet coefficients and the use of oriented 2D dual-tree complex wavelet transform in the multi-scale decomposition step. Ref. [70] employed a preliminary detection of the wavelet coefficients representing the features of interest to empirically estimate the conditional PDFs of the useful feature coefficients and background noise. It has to be applied to the original speckled image before log-compression. A speckle reduction algorithm is developed by integrating the wavelet Bayesian despeckling technique with Markov random field based image regularization [71]. The wavelet coefficients are modeled by a two-state Gaussian mixture model and their spatial dependence is characterized by a Markov random field imposed on the hidden state of Gaussian mixtures.

Most of the thresholding methods do not take into account the specific properties of the image. Wavelet despeckling under Bayesian Framework outperforms the thresholding methods by exploiting the statistics of wavelet coefficients. The disadvantage of wavelet despeckling under Bayesian Framework is that it relies on prior distributions of the noise-free image, however, in the real world, there is no speckle-free US images since speckle is inherent in US images.

Table 1

| Edge map FOM of various filters [155]. |
|---|---|---|---|---|---|---|---|
| Noisy | Lee | Kuan | Gamma | Frost | Geometric | Oddy | Wavelet |
| FOM(%)\(\alpha\) |
| \(L = 1.9\) | 0.9 | 6.1 | 6.4 | 6.2 | 28.2 | 18.3 | 10.9 | 38.1 |
| \(L = 9.4\) | 3.7 | 49.5 | 49.6 | 55.0 | 58.6 | 58.5 | 44.4 | 64.1 |

2.1.2.3. Wavelet filtering and diffusion. Besides thresholding, we can use filtering or diffusion method in wavelet domain to reduce speckle [78, 79]. Wiener filtering is applied in the wavelet domain [78]. The experimental results show that the approach performs better than wavelet thresholding visually and quantitatively. Normalized modulus-based nonlinear multi-scale wavelet diffusion (NMWDD) is proposed for speckle suppression and edge enhancement [79]. The approach has more favorable despeckling properties than that of nonlinear diffusion because the multi-scale representation gives more efficient signal/noise separation. It also outperforms wavelet-based despeckling methods by taking the advantage of edge enhancement inherited from nonlinear diffusion. Both the envelope speckle image and log-compressed image can be directly processed using this technique.

A study that compares different speckle filters in the image domain and wavelet domain is presented in [155]. It compared wavelet coefficient shrinkage (WCS) filter and several standard speckle filters (Lee, Kuan, Frost, Geometric, Kalman, Gamma, etc.) It calculates the figure of merit (FOM) of the edge map to get a quantitative evaluation of edge preservation and the results show that wavelet-domain filters preserve image details better (Table 1).

The disadvantage of wavelet-based despeckling methods is that the time complexity is increased due to the DWT and IDWT operations.

2.1.3. Compounding approaches

In compounding approaches, the image acquisition procedure has been modified to produce several images of the same region that are partially correlated or non-correlated, and averages them to form a single image. There are two general methods for de-correlation among the individual images. Spatial compounding is obtained by generating each original image while the transducer is located at different spatial locations [80, 81]. 3D spatial compounding is adopted to reduce speckle in 3D ultrasound images [83]. Frequency compounding is generated when the transducer operates at different frequencies [82]. The compounding technique reduces speckle at the expense of increasing the complexity of image registration and reconstruction.

Some speckle reduction methods are listed in Table 2.

2.2. Image enhancement

As stated at the beginning of the preprocessing section, many methods enhance the image and remove speckle at the same time. Nonlinear diffusion is such an example. It not only preserves edges but also enhances edges by inhibiting diffusion across edges and allowing diffusion on either side of the edges. Since we already reviewed those techniques in the previous section, now we will focus on the algorithms merely for image enhancement.

Histogram equalization is used to enhance the contrast [13]. The multi-peak generalized histogram equalization was proposed in [85]. It combined multi-peak histogram equalization with local information to enhance the contrast. Ref. [86] proposed stick technique for image enhancement. Sticks (line segments) in different orientations are used as the templates and the orientation which is most likely to represent a line is selected to improve edge information. This
algorithm only enhances edges and the non-line features are not affected. A contrast-enhancement algorithm based on fuzzy logic and the characteristics of breast ultrasound images was proposed in [19]. It used the maximum fuzzy entropy principle to map the original image into fuzzy domain and then the edge and textural information were extracted to describe the lesion features. The contrast ratio measuring the degree of enhancement is calculated and modified. The defuzzification process is finally applied to obtain the enhanced image. Experimental results show that the method could effectively enhance the image details without over- or under-enhancement.

3. Segmentation

Image segmentation is a critical and essential component and is one of the most difficult tasks in image processing and pattern recognition, and determines the quality of the final analysis. Segmentation [87] is a partition of the image I into non-overlapping regions

\[ I_i = I \quad \text{and} \quad I_i \cap I_j = \emptyset \quad i \neq j \]

Computer-aided diagnosis system will help radiologists in reading and interpreting sonography. The goal for the segmentation is to locate the suspicious areas to assist radiologists in diagnoses.

3.1. Histogram thresholding

Histogram thresholding is one of the widely used techniques for monochrome image segmentation [87,88]. Histogram thresholding was proposed for segmenting breast ultrasound images [89–92]. The algorithms [90,91] proposed for segmenting masses in US images involved the following steps: (1) preprocessing using cropping and median filtering, (2) multiplying the preprocessed image with a Gaussian constrain function, (3) determining the potential lesion margins through gray-value thresholding, and (4) maximizing a utility function for potential lesion margins. However, the center, width and height of the lesions needed to be selected manually or semi-manually.

Another thresholding algorithm [89,92] had four steps: First, the regions of interest (ROIs) were preprocessed with a 4×4 median filter to reduce the speckle noise and to enhance the features. Second, a 3×3 unsharp filter was constructed using the negative of a two-dimensional Laplacian filter to emphasize the elements with meaningful signal level and to enhance the contrast between object and background. Third, the ROIs were converted to a binary image by thresholding. The threshold was determined by the histogram of ROIs. If a valley of histogram between 33% and 66% of the pixel population could be found, this intensity value was selected as the threshold. If there was no such valley in that range, the intensity of 50% pixel population was selected as the threshold value. Finally, the selected nodule’s boundary pixels were obtained using morphologic operations.

Ref. [93] adopted an automatic threshold method [94] to obtain the initial image for the gradient vector flow (GVF) snake to locate the tumor contour. The thresholding method is too simple and primitive, and does not perform well for the images with histograms that are unimodal.

3.2. Active contour model

The active contour model, more widely known as snake [95], is a framework for delineating an object outline from a possibly noisy 2D image, and has been massively used as an edge-based segmentation method. This approach attempts to minimize the energy associated with the current contour as the sum of the internal and external energies. The snake model modifies its shape actively and approximates the desired contour. During the deformation process, the force is calculated from the internal energy and external energy. The external energy derived from image feature energy is used to extract the contour of the desired object boundary. The internal energy derived from the contour model is used to control the shape and regularity of the contour [95].

The snake model has been extensively used for US images [93,96–100]. The active contour model was applied to a 3D ultrasonic data file for segmenting breast tumor [93,96], and a snake technique was used to obtain the tumor contour for pre- and post-operative malignant breast excision [93].

Combining intensity and texture with empirical domain-specific knowledge and directional gradient, a deformable shape-based model [97] was studied to find lesion margins automatically. A formulation of the empirical rules used by radiologists in detecting ultrasonic breast lesions was employed to automatically determine a seed point in the image. Followed by region growing to obtain an initial segmentation of the lesion, image pixels were classified according to the intensity and texture. Boundary points were found using the directional gradient of the image. These boundary points were supplied as the initial estimate to a deformable model. No manual initialization of the contour was required. The directional gradient of the image was used as the stopping criterion.

Level set method is employed to improve the active contour segmentation for ultrasound images. Ref. [98] discussed a level set maximum likelihood method to achieve a maximum likelihood segmentation of the target. The Rayleigh probability distribution was utilized to model gray level behavior of ultrasound images. A partial differential equation-based flow was derived as the steepest descent of an energy function taking into account the density probability distribution of the gray levels as well as smoothness constraints. A level set formulation for the associated flow was derived to search the minimal value of the model. Finally, the image was segmented according to the minimum energy.

The methods based on snake-deformation model were used to handle only the ROIs, not the entire image. Automatically generating

Table 2

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filtering techniques [34–59]</td>
<td>Use moving window to convolve the filter with the image to reduce speckles</td>
<td>Simple and fast</td>
<td>1. Single scale representation is difficult</td>
</tr>
<tr>
<td>Wavelet approaches [60–79]</td>
<td>Transform image to wavelet domain and remove noise by modifying wavelet coefficients</td>
<td>1. In wavelet domain, the statistics of the signals are simplified</td>
<td>2. Sensitive to the size and shape of the filter window</td>
</tr>
<tr>
<td>Compounding approaches [80–83]</td>
<td>Average images obtained by varying scanning frequency or view angle</td>
<td>Simple</td>
<td>Requires hardware support. Increases time complexity</td>
</tr>
</tbody>
</table>

speckle reduction methods.
a suitable initial contour is very difficult, and the snake-deformation procedure is very time-consuming.

3.3. Markov random field

Ultrasound image segmentation can be considered as a labeling problem where the solution is to assign a set of labels to pixels, which is a natural representation for Markov random fields. Markov random field model has been used for US image segmentation [101–107]. The algorithm alternatively approximates the maximization of the posterior (MPM) estimation of the class labels, and estimates the class parameters. Markov random field model deals with the spatial relations between the labels obtained in an iterative segmentation process. The process assigning pixel labels iteratively can be achieved by maximizing either a posteriori estimation or posterior marginal estimation.

The algorithms based on Markov random field (MRF)/Gibbs random field (GRF) [101] was adopted to segment US images. The Metropolis sampler approach was used, and a new local energy was defined.

The new local neighboring energy is

\[ U_{N-4}(x_i) = U_{\text{local}}(x_i | y_{N_i}) + U_{\text{local}}(x_{i-1} | y_{N_{i-1}}) + U_{\text{local}}(x_{i+1} | y_{N_{i+1}}) + U_{\text{local}}(x_{i-1} | y_{N_{i-1}}) \]

where \( U_{\text{local}}(x_i | y_{N_i}) \), \( U_{\text{local}}(x_{i-1} | y_{N_{i-1}}) \), \( U_{\text{local}}(x_{i+1} | y_{N_{i+1}}) \), \( U_{\text{local}}(x_{i-1} | y_{N_{i-1}}) \) are the local energies of site \( x_i \), and its four first order neighbors, \( x_{i-1}, x_{i+1}, x_{i-1}, x_{i+1} \), and \( x_{i-1} \).

Then, \( \Delta U \) can be represented as

\[ \Delta U = U_{N-4}(x_i^j) - U_{N-4}(x_i) \]

where \( x_i^j \) is accepted as a new label of site \( x_i \). \( U(x) \) is the global energy of each configuration.

The newly defined local energy can fit into the Metropolis sampler algorithm \( \Delta U \). The Expectation-Maximization (EM) method is used for parameters estimation of each class.

Ref. [103] used a combination of the maximum a posteriori and Markov random field to estimate the US image distortion field following a multiplicative model while labeling image regions based on the corrected intensity statistics. The MAP was used to estimate the intensity model parameters while the MRF provided a way of incorporating the distributions of tissue classes as a spatial smoothness constraint.

A segmentation algorithm for breast lesion was based on multi-resolution texture adaptive clustering [104], which improved the algorithm in [108] by using a new energy function to measure textural properties of various tissues. The segmentation problem was formulated as a maximum a posteriori estimation problem. The MAP estimation utilized Besag's iterative conditional mode algorithm for minimizing an energy function constraining the region to be close to the data, imposing spatial continuity and considering the texture of various regions. However, the input images for this algorithm were only ROIs.

Ref. [105] used a Markov random field to model the region process and to focus on the adaptive characteristics of the algorithm. It introduced a function to control the adaptive properties of the segmentation process, and took into account both local and global statistics during the segmentation process. A new formulation of the segmentation problem was utilized to control the effective contribution of each statistical component.

Ref. [106] combined EM (expectation maximization) for hyper-parameter estimation and MPM (maximization of posterior marginals), and extended the EM/MPM framework to 3D by including pixels from neighboring frames in the Markov random field clique. However, there were many noisy spots in the segmentation results, and the algorithm was quite time-consuming.

The merit of MRF modeling is that it provides a strong exploitation of the pixel correlations. The segmentation results can be further enhanced via the application of maximum a posteriori segmentation estimation scheme based on the Bayesian learning paradigm [102]. However, its iteration process is complex and time-consuming.

3.4. Neural network

Neural network (NN) based methods [11,109,110] are popular in image segmentation, which transform the segmentation problem into classification decision based on a set of input features. In [109], a NN approach was combined with wavelet analysis for US image segmentation. A multi-layered perceptron (MLP) neural network having one hidden layer was designed with variance contrast and auto-correlation contrast as input features, and trained by error back propagation.

A study [110] integrated neural network classification and morphological watershed segmentation to extract the contours of breast tumors. Textural analysis was employed to find the inputs for the NN. Watershed transformation automatically determined the contour of the tumor. However, how to select the training set was problematic, and training a NN was time-consuming.

A Bayesian neural network (BNN) with five hidden units and an output node were employed for segmentation and detection [11], where input features were the depth-to-width ratio, the radial gradient index (RGI) value, texture, and posterior acoustic behavior of the suspected lesion. At first, a radial gradient index filtering technique was used to locate the ROIs and their centers were documented as the points of interest, and a region growing algorithm was used to determine candidate lesion margins. The lesion candidates were segmented and detected by the BNN. However, the algorithm would fail if the lesion was not compact and round-like. In addition, the appropriate number of hidden units for the neural network was determined empirically.

In order to compare different segmentation methods clearly, descriptions, advantages and disadvantages of different methods are discussed briefly in Table 3.

4. Feature extraction and selection

Feature extraction and selection are important steps in breast cancer detection and classification. An optimum feature set should have effective and discriminating features, while mostly reduce the redundancy of feature space to avoid “curse of dimensionality” problem. The “curse of dimensionality” suggests that the sampling density of the training data is too low to promise a meaningful estimation of a high dimensional classification function with the available finite number of training data [111]. For some advanced classification methods, such as artificial neural network and support vector machine, the dimension of feature vectors not only highly affects the performance of the classification, but also determines the training time of the algorithm. Thus, how to extract useful features and make a good selection of the features is a crucial task for CAD systems.

The features of breast US images can be divided into four categories: texture, morphologic, model-based and descriptor features. We summarize and list the typical and effectiveness-proved features in Table 4. Certainly, one cannot use all of them at the same time. Extraction and selection of effective features is a necessary step. The general guidelines for selecting significant features mainly include four considerations: discrimination, reliability, independence and optimality [112]. However, simply combining the best performed features will not definitely make the systems work well and
effectively. The goal of feature extraction and selection is to maximize the discriminating performance of the feature group.

4.1. Texture features

Most of the texture features are calculated from the entire image or ROIs using the gray level values. FT1 (auto-covariance coefficient) is a basic and traditional texture feature which can reflect the inner-pixel correlation within an image. FT2 (BDIP)–FT3 (BVLC) measure the variation of intensities and texture smoothness, respectively. The higher the variance, the larger the variance of intensities and texture smoothness, respectively. The larger the variance is, the larger the variance of intensities in a block is, and the larger BVLC value indicates that the variance in the block is rough[20]. Both the first and second order of FT2 and FT3 can be used as the features too. FT4 is defined as the ratio of the variance, auto-correlation coefficients or intensity average pixel correlation within an image. FT2 (BDIP)–FT3 (BVLC) measure the variation of intensities and texture smoothness, respectively. The higher the possibility of the tumor being malignant is. FT5 is defined as the difference between the gray scale histograms of the ROI containing the lesion: $\text{COR} = \sum_{n=0}^{N_{Ri} - 1} \frac{C_{y}(n)}{C_{y}(0)}$

where $M_{k} = \sum_{m=0}^{M_{k} - 1} C_{y}(m, n)$, $C_{y}(m, n) = \sum_{p=0}^{N_{x} - 1} I_{x}(m, n + p)I_{y}(m, p)$

$M_{k}$ is the number of pixels in the lateral direction of the ROI and $N_{Ri}$ is the number of pixels in the depth direction of the ROI, and $I_{x}$ and $I_{y}$ is the gray level value matrix of the ROI. Because COR is a sum, it includes not only the texture information but also the size information.

Based on understanding of the posterior acoustic behavior or posterior shadow, different numeric expressions are proposed to calculate FT12. In [90], three ROIs were defined whose width and depth were the same as the ROI containing the lesion itself. As Fig. 3 shows, the post ROI represents the posterior region of the lesion and the right ROI and left ROI are adjacent tissues at the same depth of the posterior ROI. The narrow blank boundaries are used to avoid the edge shadows. Finally, the minimum side difference (MSD) is defined as: $\text{MSD} = \text{min}(\Delta_{left}, \Delta_{right})$, where $A_{left}$ and $A_{right}$ are the average gray-level values of the corresponding ROIs. In [131], another method to calculate posterior shadow was proposed. First a skewness image is built by $\text{Skew}(x, y) = \frac{1}{N} \sum_{(x', y') \in A} (l(x', y') - l(x, y))^{3}$

where $A$ is a specified region centered at point $(x, y)$, $l(x', y')$ is the gray value in the original image, $N$ is the total number of data points in region $A$ and $\sigma_{A}$ is the standard deviation of the gray values in area $A$. The skewness image is filtered with a threshold to get the detection points, i.e., the shadow. In [25], the posterior shadow was defined as the difference between the gray scale histograms of the regions inside the lesion and posterior to the lesion. For the same characteristic of breast lesions, we can use different ways to define the numeric expressions. To find more accurate and efficient expressions should be one of the future works.

FT13 is the Boltzmann/Gibbs entropy over the gray scale histogram relative to the maximum entropy. The higher the entropy is, the more homogeneous the lesion is. FT15–FT16 are well-known texture features which have already been well defined. However, they are not frequently used in recent US image characterization. This may be due to their high computation cost. The definition of the fractal dimension (FT17) is similar to the Hausdorff dimension [137]. Informally, the dimension $d$ can be calculated by $N = s^{d}$, where $N$ is...
<table>
<thead>
<tr>
<th>Feature category</th>
<th>Feature description</th>
</tr>
</thead>
</table>
| **Texture features** | FT1: Auto-covariance coefficients [3,13,16,20,22,125–127,132-134]  
FT2: Block difference of inverse probabilities (BDIP) [20]  
FT3: Block variation of local correlation coefficients (BVLC) [20]  
FT4: Variance, auto-correlation, or average contrast [18,109,120]  
FT5: Distribution distortion of wavelet coefficients [109]  
FT6: Mean and variance of the order statistics after wavelet decomposition [109]  
FT7: Contrast of grey level values [22,132,134,139]  
FT8: Correlation of the co-occurrence matrix [139]  
FT9: Dissimilarity [22,132,139]  
FT10: Relative frequency of the edge elements [139]  
FT11: Auto-correlation in depth of \( R \) (COR) [11,90,113,129,130,140,141]  
FT12: Posterior acoustic behavior, minimum side difference (MSD) or posterior acoustic shadow [9,11,25,90,113,117,129,130,140,141]  
FT13: Homogeneity of the lesion [25]  
FT14: Standard deviation of gray value and its gradient of the lesion [129]  
FT15: SGLD matrix based features: correlation, energy, entropy, sum entropy, difference entropy, inertia and local homogeneity [16,21,24]  
FT16: GLD matrix based features: contrast, mean, entropy, inverse difference moment and angular second moment [16]  
FT17: Fractal dimension and related features [21,150] |
| **Morphologic features** | FM1: Spiculation [18,119,141]  
FM2: Depth to width ratio (or width to depth ratio) [9,10,11,18,90,113,114,119,121,136,140-143]  
FM3: Branch pattern [18,119]  
FM4: Number of lobulations [18,25,119,141,152]  
FM5: Margin sharpness [17,123,129,130]  
FM6: Margin echogenicity [17,123]  
FM7: Angular variance in margin [17,123]  
FM8: Number of substantial protuberances and depressions (NSPD) [10]  
FM9: Lobulation index (LI) [10]  
FM10: Elliptic-normalized circumference (ENC) [10]  
FM11: Elliptic-normalized skeleton (ENS) [10]  
FM12: Long axis to short axis ratio (\( L:S \)) [10]  
FM13: Area of lesion [10,25,143]  
FM14: Normalized radial gradient (NRG) along the margin [11,90,113,114,129,130,140,141]  
FM15: Margin circularity [25]  
FM16: Degree of abrupt interface across lesion boundary [152]  
FM17: Angular characteristic [152] |
| **Model-based features** | FB1: \( \beta_0, \beta_\nu, \text{ and } \beta_{\nu_0} \) of PLSN model [122,144]  
FB2: \( m \) and \( \varphi \) of Nakagami model based features [120,122]  
FB3: Single and combined parameters of GS model [122]  
FB4: \( b_{\nu_{00}} \) and \( M \) of \( K \) distribution model based features [120]  
FB5: Normalized skewness \( K \) [124]  
FB6: Signal to noise ratio of the envelope [124]  
FB7: Normalized spectral power \( z \) [124]  
FB8: Margin strength [124]  
FB9: Quality of margin \( \beta \) [120,124]  
FB10: Speckle factor [84,118] |
| **Descriptor features** | FD1: Non-circumscribed or spiculated margins [9,14,115,119,121,135,136,138,142,143]  
FD2: Shape (round, oval or irregular) [9,14,115,119,121,135,136,138,142]  
FD4: Posterior shadow or posterior echo [119,121,135,136,142]  
FD5: Decreased sound transmission or acoustic transmission [14]  
FD6: Echogenicity [14,121,135,136,142]  
FD7: Heterogeneous echo texture [115,121,135,136,143]  
FD8: Duct extension [119]  
FD9: Thickened cooper ligaments [143]  
FD10: Antiparallel orientation [14,135,138]  
FD11: Distortion, echogenic halo or rim of surrounding tissue [14,136]  
FD12: Bilateral refraction sign [121]  
FD13: Microllobution [119,136] |

In the polar coordinates \((r,\theta)\), each boundary pixel is represented as \( r(\theta) \) and FM1 (spiculation) is the ratio of low-frequency component (area under the graph \(|R(w)|\) from 0 to \(\pi/4\)) to high-frequency component (area under the graph \(|R(w)|\) from \(\pi/4\) to \(\pi\)), where \(|R(w)|\) is the Fourier transform of \(r(\theta)\) and the cutoff frequency \(\pi/4\) was experimentally chosen [18]. The larger the value is, the lower possibility of the tumor being malignant is. FM2 is one of the most effective distinguishing features mentioned in many papers. Malignant lesions tend to have the ratio bigger than 1 while benign lesions usually have the ratio smaller than 1. FM3–FM4 are the

---

4.2. Morphologic features

Unlike texture features extracted from the rough ROIs, the morphologic features focus on some local characteristics of the lesion, such as the shape and margin.
numbers of local extremes in the low-pass-filtered radial distance graph and the curve-fitted radial distance graph, respectively. Malignant lesions tend to have higher value of FM3 or FM4. The ROC curves of classification using FM1–FM4 are compared in Fig. 4. For FM5–FM7, the lesion is divided into \( N \) sectors, and in each sector, the mean gray levels of the pixels in the inner and outer shells are compared. By using a user-defined threshold, some of the sectors are chosen distinctly. The margin sharpness is calculated as \( \frac{\text{number of distinct sectors}}{N} \times 100 \). Margin echogenicity is the mean gray level difference of the inside and outside of each sector. Angular variance in margin is the ratio SD/mean for the difference in the mean gray level of the inside and outside of each sector. All of the above three features are proved to be significantly different by Student \( t \) test when they are used to distinguish benign and malignant lesions \[123\]. Table 5 summarizes the mean ± SD and \( p \) values of the three features.

FM8–FM12 are newly proposed morphologic features \[10\]. As shown in Figs. 5 and 6, a breast lesion is delineated by a convex hull and concave polygon. Given a threshold angle of \( \theta \in \{20^\circ, 30^\circ, 40^\circ, 50^\circ, 60^\circ\} \), let \( A = \{\omega_1, \omega_2, \ldots, \omega_p\} \) and \( \Omega = \{\omega_1, \omega_2, \ldots, \omega_p\} \) denote the set of representative convex and concave points of a lesion boundary, respectively, where \( p \) and \( d \) are the numbers of points in each set. Thus, the NSPD is defined as \( p+d \). Ideally, a malignant breast lesion has a larger NSPD. FM9 (LI) is defined as \( (A_{\text{max}} - A_{\text{min}}) \times N/\sum A_i \) where \( A_{\text{max}} \) and \( A_{\text{min}} \) are the sizes of maximum and minimum lobes as illustrated in Fig. 6 and \( N \) is the total number of the lobes. LI is an effective complement of NSPD, and can correctly characterize the benign lesions with multiple large lobes of similar sizes which are easily to be misclassified by NSPD \[10\]. FM10 (ENC) is defined as the circumference ratio of the lesion to its equivalent ellipse (Fig. 7), and it represents the anfractuosity of a lesion which is a characteristic of malignant lesions. FM11 (ENS) is defined as the number of skeleton points normalized by the circumference of the equivalent ellipse of the lesion. The calculation cost of this feature is relatively high. Same as FM10, a malignant lesion tends to have a higher value of FM11. These four features capture mainly the contour and shape characteristics of the lesion. FM12 (L:S) is the ratio of the long- to short-axis, where the long and short axes are determined by the major and minor axes of the equivalent ellipse. Therefore, L:S ratio is different from the traditional depth/width ratio (FM2) because it is independent of the scanning angle. For both FM12 and FM13 (lesion area), clinically, the larger the value is, the lesion is more likely malignant. Among the five newly proposed features (FM8–12), NSPD is proved to be the most important feature, and NSPD, LI, ENS and ENC are better than lesion size, L:S ratio and depth/width ratio. FM14 is used to measure the average orientation of the gray level...
gradients along the margin. The formula for FM14 is

\[
NRG = \frac{\sum_{j=0}^{J-1} \| \nabla I(x_j, y_j) \|}{\sum_{j=0}^{J-1} \| [\nabla I(x_j, y_j)] \|}
\]

where \( \nabla I \) is the gradient image computed using Sobel filter, \( J \) is the number of points on the margin and \( \hat{r}(x_j, y_j) \) is the unit vector in the radial direction from the geometric center to the point \((x_j, y_j)\). FM15 is defined as the standard deviation of the normalized distances of every boundary point to the ROI's geometric centers [25]. A high value of FM15 is a sign of malignancy. FM16 and FM17 are proved to be two important features in [152]. To define FM16 and FM17, a distance map should be first calculated. For each pixel in the image, its value in the distance map is defined as the shortest distance from the pixel to the lesion boundary. FM16 is designed to estimate the degree of abrupt interface across the lesion boundary. The formula is \( LB_D = \text{avg}_{Tissue} - \text{avg}_{Mass} \).

\[
\text{avg}_{Tissue} = \frac{\sum_{k=0}^{k-1} l(n)}{N_{Tissue}}, \quad \text{avg}_{Mass} = \frac{\sum_{k=0}^{k-1} l(n)}{N_{Mass}}
\]

where \( l(n) \) is the gray level value of pixel \( n \), dis(n) is the value of pixel \( n \) in the distance map, \( N_{Tissue} \) is the number of pixels in the surrounding tissue and \( N_{Mass} \) is the number of pixels in the outer mass. Both the surrounding tissue and outer mass areas are composed of pixels whose distances to the lesion boundary are no more than \( k \) in the distance map. Width \( k \) was set to 3 in [152]. The likelihood of malignancy was decreased as the increasing of FM16. FM17 (angular characteristics) is defined as the sum of numbers of local maxima in each lobulate area [152]. Fig. 8 shows an example of lobulate areas. By the maximum inside circle, some lobulate areas are partitioned from the mass. Some small lobulate areas with maximum distance less than 4 are discarded. For the remained lobulate areas, the local maxima in each lobulate area are grouped as follows: when a new local maximum is discovered, the Euclidean distance from it to the center of the grouped local maxima is calculated. If the distance is larger than a predefined threshold, this local maximum is regarded as a new grouped local maximum. At last, the total number of local maxima in the lobulate areas represents the angular characteristics. In [152], the predefined threshold was set to 10. The larger the value is, the lesion is more likely malignant.

4.3. Model-based features

Model-based feature is a special type of US features which focuses on the backscattered echo from the breast tissue. Different models are developed to simulate the echo of the backscattered envelope. Once a model is chosen and the echo is modeled, the parameters of the model can be used as the features to distinguish malignant and benign lesions. The models have been used for breast cancer diagnosis including power-law shot noise (PLSN) model, Nakagami model, \( K \) distribution model and generalized spectrum (GS) model. Comparing with texture features and morphologic features, the advantages of model-based features are that they are not influenced by the experience of the radiologists, and not influenced by the ways in which the images are collected. They are operator- and machine-independent [120]. The disadvantage of model-based features is that the background of the models is quite complex and the estimation of the parameters is very complicated as well.

4.4. Descriptor features

Descriptor features are easier to understand because they are actually the empirical classification criteria of the radiologists. Most of them are descriptive and have no numeric expressions. The reason we list this type of features is that some of the useful features for current CAD systems are transferred from the descriptor features. There are still useful descriptor features that have not been transferred to numeric expressions, and cannot be used in CAD systems. In this subsection, we have chosen the most frequently used descriptor features which are proved to be effective in distinguishing benign and malignant lesions. Except FD2 and FD12, all the listed descriptor features are malignant characteristics.

FD1 and FD2 are the most powerful features to characterize malignant lesions. For FD2, oval or round shape is a sign for benign and “taller than wide” or other irregular shape is a sign for malignancy. FD3 describes calcifications or microcalcifications in the lesion. FD4 is called posterior shadow or posterior echo, and it focuses on the region posterior to the lesion ROI which has darker gray value than that of the surrounding. FD5 is defined as the shadow effect of surrounding tissues. FD6 works well for differentiating large tumors but not small tumors. FD7 is an argumentative feature. This might be caused by the subjective nature of this feature so that an accurate numeric expression of FD7 is needed. FD8 is a projection from the nodule that extends radially within or around a duct and toward the nipple [119]. FD9 represents thickened suspensory ligaments of the breast which tend to stretch over time. FD11 describes the echogenicity of the surrounding tissue of the tumor. FD12 is an acoustic phenomenon that mostly occurs in benign tumors. FD13 is recognized by the presence of many small lobulations on the surface of the solid lesion. Most of these descriptor features are included in the Breast Imaging Reporting and Data System (BI-RADS) [146].
4.5. Other features

Sometimes, other information can be integrated to help the diagnosis. Patient’s age is proved to be an effective feature to diagnose malignancy [14,17,122,123]. Also, family disease history is another useful feature for the diagnosis.

4.6. Reduce feature space dimension

With so many features available, the crucial task is to find an optimal set of features with relative low dimension. The feature extraction transforms the coordinate system to improve a determined goal; whereas feature selection only reduces the dimensionality, i.e., does not change the coordinate system of the data [156].

4.6.1. Feature extraction

Feature extraction, linearly or nonlinearly, transforms the coordinate system of the original variables [156]. The most well-known feature extraction technique is principal component analysis (PCA). PCA performs on the symmetric covariance matrix or symmetric correlation matrix, and solves the eigenvalues and eigenvectors of the matrix. PCA is good at reducing the high dimensional correlated features into low dimensional features. The feature vector of the auto-covariance coefficients can be optimized by PCA effectively [13,16]. Other feature extraction techniques such as factor analysis (FA) [162], independent component analysis (ICA) [158], and discriminant analysis (DA) [159] can also be used to reduce the feature dimension.

4.6.2. Feature selection

Generally, algorithms for feature selection can be categorized into two classes: wrapper and filter. Filter approach (such as FOCUS and Relief algorithms [157]) selects features using a preprocessing step and does not take into account the bias of induction algorithms. On the contrary, to search for a good subset of the features, wrapper approach uses the induction algorithm as a part of the evaluation function. Ref. [157] provided detailed explanations and summaries of these two classes of feature selection algorithms. As the wrapper approach has obvious advantages over filter approach, especially for complex feature data set, wrapper approach has more applications in breast cancer detection [9,21,113,114]. For example, [114] applied a wrapper approach (linear stepwise feature selection) to a feature set composed of 15 sonographic features of breast cancer and found that the two most significant features were the average orientation of gray level gradients along the margin and depth-to-width ratio.

5. Classifiers

After the features have been extracted and selected, they are input into a classifier to categorize the images into lesion/non-lesion or benign/malignant classes. Majority of the publications focuses on classifying malignant and benign lesions (usually called lesion classification), and some of the articles focus on classifying lesions and non-lesions (usually called lesion detection), and only a few of them focus on both. Lesion detection is necessary before lesion classification. We summarize the different classifiers commonly used in breast cancer detection and classification in Table 6.

5.1. Linear classifiers

Frequently used linear classifiers for breast cancer detection and classification are linear discriminant analysis [160] and logistic regression (LOGREG) [161]. The main idea of LDA is to find the linear combination of the features which best separate two or more classes of the data. If there are n classes, and LDA classifies the data by the following linear functions:

\[ f_i = \mu_i X^T - \frac{1}{2} \mu_i X_i^T + \ln(p_i), \quad 1 \leq i \leq n \]

where

\[ C = \frac{1}{n} \sum_{i=1}^{n} n_i C_i, \quad P_i = \frac{n_i}{N} \]

\[ n_i \text{ is the number of samples in the } i \text{th class}, \quad N \text{ is the number of total samples}, \quad \mu_i \text{ is the mean of class } i, \quad C_i \text{ is the covariance matrix of class } i. \]

The above parameters can be obtained from the training data. When a new data \( x_k \) is in, it is assigned to class \( i \) with the highest \( f_i \).

Logistic regression is a model for predicting the probability of an event occurring as a function of other factors. The probability of \( X = x_1, x_2, \ldots, x_n \) is formulated as

\[ \logit(P) = \log \frac{P}{1 - P} = b_0 + \sum_{i=1}^{n} b_i x_i \]

where \( b_0, \ldots, b_n \) are model parameters which could be estimated from the training data. When LOGREG is used to classify two-class problem, for each feature vector \( x_i \), threshold \( 0.5 \) is used to decide which class \( X \) belongs to.

In [90], LDA was applied to the data set of 400 cases with four automatically extracted features. The average \( A_2 \) under ROC curve was 0.87 over eleven independent trials. In [123], LOGREG was used to determine the probability of malignancy in a database of 58 cases. Three margin-based features were evaluated and the area under the ROC curve with the best feature combination of age, margin echogenicity and angular variation was 0.87 ± 0.05. Here, we can see that the performances of LDA and LOGREG are not high because the classifiers are linear, and for nonlinear separable data, the methods have intrinsic limits.

5.2. Artificial neural networks

Artificial neural networks are the collection of mathematical models that imitate the properties of biological nervous system and the functions of adaptive biological learning [2]. It is a self-learning system that changes its parameters based on external or internal information that flows through the network during the learning phase. ANN is composed of an input layer, an output layer and one or more hidden layers. Layer is composed of neurons. In the field of breast cancer detection and classification, three types of artificial neural networks are frequently used: Back-propagation neural network, self-organizing map (SOM) and hierarchical ANN [10,17,18,109,125–127].

5.2.1. Back-propagation neural network

Back-propagation (BP) neural network is a feed-forward ANN with supervised learning process. Frequently used back-propagation neural networks have one or two hidden layers and 2–10 neurons in each layer. There is no universal rule to decide the number of layers or number of neurons in each layer. In [10], a BP neural network was used in the proposed CAD and the result was compared with that in [125]. The CAD in [10] achieved \( A_2 = 0.959 \pm 0.005 \) with the selected morphologic features and outperformed the one in [125]. Ref. [150] combined K-means classification with BP neural network. K-means classification was used to select training samples for BP neural network and only those samples within a distance to the cluster center would be used for training. The performance of back-propagation neural network is better than that of linear classifiers. However, the training process is stochastic and unrepeatable even with the same data and same initial conditions.
SOM can automatically classify input data into different classes (number of classes could be more than 2) and it is a totally unsupervised method. The disadvantages are that the number of parameters grows exponentially with the dimension of input space, and user cannot decide the number of classes. In [126], SOM was used to classify US images [14,115,135,136,138,142].

5.2.2. Hierarchical ANNs
In a hierarchical ANN structure, individual ANNs are combined into a tree structure, and each node is associated with an ANN. In [127], a 2-layer hierarchical ANN was developed with 4 ANNs in the first layer and one ANN in the second layer. The performance of the proposed hierarchical ANN system was $A_r = 0.9840 \pm 0.0072$ on a data set of 1020 images. Although the performance is high, the extraction of the 4 ROIs is troublesome and the whole training process is time-consuming.

5.3. Bayesian neural networks
Bayesian neural network is a kind of neural network using Bayesian method to regularize the training process [145]. The idea behind BNN is to cast the task of training a network as a problem of inference, which is solved using Bayes' theorem [128]. Bayesian neural network is more optimal and robust than conventional neural networks, especially when the training data set is small.

A BNN with one hidden layer and five neurons in the hidden layer was chosen to detect lesions [11]. This work focused on distinguishing true lesions from non-lesions. The performance was $A_r = 0.84$ on the database of 757 images. Using the same database, in [129], two BNNs were trained and tested separately with different tasks. One was used to classify true lesions from non-lesions, and the other was used to classify malignant lesions from other detections. The performance of these two BNNs were $A_r = 0.91$ and 0.81, respectively. In [130], a 3-way BNN was used to classify the data into three classes (malignant, benign and non-lesion). To evaluate the performance, the output can be projected to 2-way classifiers. In this way, on database of 858 cases (1832 images), the performance of classifying lesions from non-lesions was $A_r = 0.92$, and the performance of classifying malignant from other detections was $A_r = 0.83$. The BNN model is easy to incorporate priori information, but to estimate those statistical parameters requires a relatively huge database.

5.4. Decision tree
A decision tree is a simple tree structure where non-terminal nodes represent tests on one or more attributes and terminal nodes reflect decision outcomes. Each non-terminal node has a threshold associating with one or more features to divide the data into its descendents, and the process stops when each terminal node only contains one class. Thus decision tree can be used as a classification tool after the thresholds are set in the training process. Comparing with neural networks, the decision tree approach is much simpler and faster [2]. However, it highly relies on the design of classification rules on each non-terminal node and the set of threshold values.

A well known algorithm for constructing decision trees is C4.5 [131]. This algorithm has been incorporated into the free classifier package WEKA (it is called J48 in WEKA) and widely used in artificial intelligence. An updated version C5.0 provides a number of improvements on C4.5. In [132], algorithm C5.0 was used to build the decision tree with 153 training samples and 90 testing samples. Covariance coefficients of the ROIs were features inputting to the decision tree, and the performance on the testing data set was accuracy $= 96\%$ (86/90), sensitivity $= 93.33\%$ (28/30) and specificity $= 96.67\%$ (58/60), respectively. The performance was compared with that of an experienced physician on the same testing data set and experiment result showed that the proposed CAD did a better job. Ref. [22] used bootstrap technique to train the decision tree with small size training sets which were parts of the database in [132]. Bootstrap technique was proved to be effective and useful, especially, for the case that a huge database was not available.

### Table 6: Classifiers

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Features used</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linear classifiers:</strong> Construct decision boundaries by optimizing certain criteria: LDA and LOGREG [9,90,116,120,121,122,123,124,139,143,152]</td>
<td>Text features (FT6–FT8, FT10, FT12), morphologic features (FM2, FM5–FM7, FM14), descriptor features (FD1–FD4, FD6–FD7, FD9, FD12)</td>
<td>Simple and effective for linearly separable data</td>
<td>Poor performance for nonlinearity separable data. Poor adaptability for complex problem</td>
</tr>
<tr>
<td><strong>ANNs:</strong> Construct nonlinear mapping functions: Back-propagation, SOM and a hierarchical ANN [10,16,17,18,109,125,126,127,150]</td>
<td>Text features (FT1, FT4, FT5), morphologic features (FM1–FM4, FM8–FM13)</td>
<td>Robustness, no rule or explicit expression is needed, and widely applicable</td>
<td>Long training time, initial value dependent, unrepeatable, over-parameterization and over-training</td>
</tr>
<tr>
<td><strong>SOM:</strong> A probabilistic approach to estimate the class conditional probability density functions [11,129,130,141]</td>
<td>Texture features (FT11, FT12, FT14), morphologic features (FM2, FM5, M14)</td>
<td>Priori information can incorporate in models, useful when there is finite training data</td>
<td>Need to construct model and estimate the associated parameters</td>
</tr>
<tr>
<td><strong>Decision tree:</strong> A tree structure with classification rules on each node [22,132]</td>
<td>Text features (FT1, FT7, FT9)</td>
<td>Training process is faster than NNs, Repeatable training process, good performance</td>
<td>Accuracy depends fully on the design of the tree and the features</td>
</tr>
<tr>
<td><strong>Support vector machines:</strong> Map the input data into a higher dimension space and seek an optimal hyperplane to separate samples [3,20,21,25,133]</td>
<td>Text features (FT1–FT3, FT12, FT13), morphologic features (FM4, FM13, FM15)</td>
<td>No training process needed, new data can be directly added to the system</td>
<td>Supervised learning (training data should be labeled), parameter-dependent</td>
</tr>
<tr>
<td><strong>Template matching:</strong> Uses retrieval technique to find the most alike image in the database and assign the query image to the class of the most alike image [13,16,134]</td>
<td>Text features (FT1, FT7, FT9, FT15, FT16)</td>
<td>Incorporate human knowledge and use the features that cannot be used by computers</td>
<td>Requiring large size database, images should come from the same platform to archive better performance</td>
</tr>
<tr>
<td><strong>Human classifiers:</strong> Physicians/radiologists, use empirical criteria to classify US images [14,115,135,136,138,142]</td>
<td>Descriptor features (FD1–FD14), morphologic features (FM2)</td>
<td>No training process needed in the system</td>
<td>Interobserver variability, unstable and inaccurate, human error, and subjectiveness</td>
</tr>
</tbody>
</table>
5.5. Support vector machine

Support vector machine, is a supervised learning technique that seeks an optimal hyperplane to separate two classes of samples. Kernel functions are used to map the input data into a higher dimension space where the data are supposed to have a better distribution, and then an optimal separating hyperplane in the high-dimensional feature space is chosen.

In [3,20,21,133], SVM was applied to classify malignant and benign lesions. Both the performance and time cost of SVM were compared with ANN on the same data set [133]. Experiment result showed that SVM ($A_z = 0.970$) not only outperformed the ANN ($A_z = 0.956$), but also was almost 700 times faster than ANN. Ref. [21] proposed a fuzzy support vector machine (FSVM) based on a regression model. The performance of the FSVM outperformed both the ANN and SVM with classification accuracy = 94.25%. The drawback of SVM is that the parameters such as the cost parameter C which controls the trade-off between allowing training errors and forcing rigid margins, and the kernel type needs to be tuned. Also, the mapping to higher dimension is complex and training time increases exponentially with the input data dimension.

5.6. Template matching

Image retrieval technique can be used to classify malignant and benign lesions. The methods use feature vectors to represent the query image and the images in the database. Based on the similarity formula, the distance between the query image and each image in the database was calculated. The final class of the query image was decided by combining the first $K$ retrieved images with the $K$ highest similarity scores.

In [134], texture features were used directly as the feature vector to calculate the similarity score and the disadvantage of the method is that it requires the images in the database come from the same platform. In [13], the principle component analysis (PCA) was applied to the entire database to form a basic set of the images and each image was represented by a weighted linear combination of the images in the basic set. The weight vector was the new feature vector used to calculate the similarity score. This method was robust with the images from different sources.

The advantage of using image retrieval technique to classify breast lesions is that no training is needed and new images can be incorporated into the system easily. The disadvantages are that for some systems the images in the database have to come from the same platform to guarantee that the similarity measure is fair, and the running time of the algorithm increases if the size of database increases. However, to get a better performance, the method requires that the database is big enough to include various lesion types. There is a trade-off between database’s size and algorithm’s efficiency.

5.7. Human classifier

Human classifiers imply the radiologists who classify the lesions using empirical criteria. They are not a component of CAD systems.
Other frequently used criteria are [10,14,17,18,21,115]:

- **Sensitivity** = \( \frac{TP}{TP + FN} \)
- **Specificity** = \( \frac{TN}{TN + FP} \)
- **Positive predictive value (PPV)** = \( \frac{TP}{TP + FP} \)
- **Negative predictive value (NPV)** = \( \frac{TN}{TN + FN} \)
- **Accuracy** = \( \frac{TP + TN}{TP + TN + FP + FN} \)
- **Matthew’s correlation coefficient (MCC)** = \( \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \)

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives.

The last formula is Matthew’s correlation coefficient (MCC), which has seldom been used for breast cancer CAD performance evaluation. However, MCC is a powerful accuracy evaluation criterion of machine learning methods. Especially, when the number of negative samples and positive samples are obviously unbalanced, MCC gives a better evaluation than overall accuracy. As more and more breast cancer CAD systems employed machine learning methods, such as SVM, ANN and BNN, MCC should be used as an additional evaluation criterion.

The performance of some CAD systems and the databases used are listed in Table 8.

<table>
<thead>
<tr>
<th>Database</th>
<th>Description</th>
<th>Source</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB1 [129]</td>
<td>The database consists of two parts: 1. 757 images, including 229 complex cysts, 334 benign solid lesions and 194 malignant lesions 2. 1740 images, including 258 complex cysts, 520 simple cysts, 210 benign solid lesions, 87 other benign breast disease, 87 malignant lesions and 578 normal images</td>
<td>The first subset of DB1 is obtained from Northwestern University, Chicago, IL. The second subset of DB1 is obtained from University of Chicago, Chicago, IL</td>
<td>( A_c = 0.94 ) and 0.91 on training and testing data sets, respectively</td>
</tr>
<tr>
<td>DB2 [116]</td>
<td>The database has 204 US images. 204 ROIs are labeled with lesions and 816 ROIs are labeled without lesions</td>
<td>Images are acquired at Thomas Jefferson University Hospital, Philadelphia, Taiwan, China</td>
<td>( A_c = 0.91 )</td>
</tr>
<tr>
<td>DB3 [133]</td>
<td>The database consists of two subsets obtained from different periods: 1. 88 benign lesions and 52 malignant lesions 2. 215 benign lesions and 35 malignant lesions</td>
<td>Images are acquired at a database of a medical center in Taiwan, China</td>
<td>( A_c = 0.95 )</td>
</tr>
<tr>
<td>DB4 [10]</td>
<td>This database consists of two subsets: 1. 91 benign lesions and 39 malignant lesions 2. 40 benign lesions and 71 malignant lesions</td>
<td>Images are collected from a database of a medical center in Taiwan, China</td>
<td>( A_c = 0.95 )</td>
</tr>
<tr>
<td>DB5 [18]</td>
<td>The database consists of two subsets: 1. 300 benign lesions and 284 malignant lesions 2. 167 benign lesions and 99 malignant lesions</td>
<td>Images are provided by the Seoul National University Hospital, Seoul, Korea</td>
<td>( A_c = 0.95 )</td>
</tr>
<tr>
<td>DB6 [21]</td>
<td>51 Benign lesions and 36 malignant lesions</td>
<td>Images are acquired from 2nd Affiliated Hospital of Harbin Medical University, Harbin, China</td>
<td>Accuracy = 94.25%, sensitivity = 91.67%, specificity = 96.08%, PPV = 94.29%, NPV = 94.23%</td>
</tr>
</tbody>
</table>

As mentioned before, there is no a benchmark accessible to the public yet, therefore, the results listed just have some reference value.

### 7. Future directions

Masses and microcalcifications are both important signs of breast cancer [1,2]. Currently, in the field of breast cancer CAD systems using US images, most of the works focus on mass detection and classification since the ordinary US images can hardly show microcalcifications. One of the future directions is high-resolution US imaging devices which can support microcalcification detection [27,28]. Some successful experiments proved that high-resolution US can show microcalcifications within breast cancer with a sensitivity of 95%. With the development of US equipments and refinement of image techniques, the rate of microcalcification detection and characterization will be higher. Besides, such advancement will also improve the detection of blood flow, an indicator of malignancy [147].

Three-dimensional ultrasound imaging is another future direction which has been paid more and more attention. Three-dimensional ultrasound imaging can provide more comprehensive information of the breast lesion than 2D imaging and incorporate all 2D characteristics. The advantages of 3D US are especially obvious in a CAD system because CAD system is good for processing a large amount of data in a short time, which can greatly reduced the variability of the observations and the work load of radiologists. Most of the 2D techniques can be directly applied to 3D US images with some preprocessing or post-processing methods [9,16,23–25,136]. Some newly developed methods, especially for 3D US images, can be found in [26,82,83,93,96].

In order to handle the fuzzy and uncertainty nature of the US images, some new techniques and approaches based on fuzzy logic, rough set and neutrosophic logic have been developed. Neutrosophic logic, a new powerful theory which handles indeterminate and uncertain characteristics in different sets, could be applied to medical image processing [148]. Some research works using fuzzy logic and fuzzy entropy have obtained good results [19,21]. Quantitative ultrasound (QUS) technique is recently used for breast cancer detection and diagnosis. A research group implemented a multi-parameter approach using QUS technique and the experimental result showed...
that this approach improved the diagnostic potential of ultrasound for breast cancer detection [149]. Besides, whole breast US images which provide more breast information are available to detect breast cancer with the support of advanced scanner [151]. Also developing more accurate numeric expressions for image features is another future work to improve the performance of CADs.

Comparative studies of different preprocessing methods, segmentation methods, features and classifiers should be carried out fairly, deeply and accurately. As we stressed above, the most urgent task is to build a benchmark database of US images accessible to the public to support the comparison and evaluation of different algorithms and CAD systems. Besides, more clinical trials of the currently developed CAD systems for breast cancer should be conducted, which can provide important second opinion to physicians and can provide the feedback from the physicians to the CAD system designers.

8. Conclusions

In this paper, we reviewed CAD systems for breast cancer detection and classification using ultrasound images in the literature. The techniques developed in the four stages (preprocessing, segmentation, feature extraction and selection, classification) are summarized, and their advantages and disadvantages are discussed. Different performance evaluation metrics are studied, and the future developments and trends are also investigated. The paper will be useful for the researches in BUS imaging, computer vision, image processing and radiology.

References

Sixth International Conference on Computer Vision, Pattern Recognition and Computer Imaging, Salt Lake City, USA, 2005.


About the Author—HENG-DA CHENG received Ph.D. degree in Electrical Engineering from Purdue University (Supervisor: Prof. K.S. Fu), West Lafayette, IN, 1985. Now, he is a Full Professor, Department of Computer Science, and an Adjunct Full Professor, Department of Electrical Engineering, Utah State University, Logan, Utah. Dr. Cheng is an Adjunct Professor and Doctoral Supervisor of Harbin Institute of Technology. He is also a Guest Professor of the Institute of Remote Sensing Application, Chinese Academy of Sciences, a Guest Professor of Wuhan University, a Guest Professor of Shantou University, and a Visiting Professor of Northern Jiaotong University. Dr. Cheng has published more than 250 technical papers and is the Co-editor of the book, Pattern Recognition: Algorithms, Architectures and Applications (World Scientific Publishing Co., 1991). His research interests include image processing, pattern recognition, computer vision, artificial intelligence, medical information processing, fuzzy logic, genetic algorithms, neural networks, parallel processing, parallel algorithms, and VLSI architectures. Dr. Cheng is the General Chair of the 11th Joint Conference on Information Sciences (JCIS 2008), was the General Chair of the 10th Joint Conference on Information Sciences (JCIS 2007), was the General Chair of the Ninth Joint Conference on Information Sciences (JCIS 2006), was the General Chair of the Eighth Joint Conference on Information Sciences (JCIS 2005), General Chair and Program Chair of the Sixth International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2007), General Chair and Program Chair of the Sixth International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2006), General Chair and Program Chair of the Sixth International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2005), was the General Chair and Program Chair of the Fifth International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2003), and was the General Chair and Program Chair of the Fourth International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2002), the General Chair and Program Chair of the Third International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2000), the General Chair and Program Chair of the First International Workshop on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 1998), and the Program Co-Chairman of Vision Interface 1990. He served as program committee member and Session Chair for many conferences, and as reviewer for many scientific journals and conferences. Dr. Cheng has been listed in Who’s Who in the World, Who’s Who in America, Who’s Who in Communications and Media, Who’s Who in Finance and Industry, Who’s who in Science and Engineering, Men of Achievement, 2000 Notable American Men, International Leaders in Achievement, 500 Leaders of Influence, International Dictionary of Distinguished Leadership, etc. Dr. Cheng is also an Associate Editor of Pattern Recognition, an Associate Editor of Information Sciences and Associate Editor of New Mathematics and Natural Computation.

About the Author—JUAN SHAN received her B.S. degree from Computer Science School of Harbin Institute of Technology (HIT), Harbin, China, in 2004. She is a Ph.D. student in the Department of Computer Science, Utah State University from 2004 to now. Her research interests include pattern recognition, computer vision and medical image processing.

About the Author—WEN JU received her B.S. degree in Computer Science from University of Science and Technology of China in 2002. Currently, she is a Ph.D. student in the Department of Computer Science, Utah State University. Her research interests include: pattern recognition, computer vision and medical image processing.

About the Author—YANHUI GUO was born in 1979, PR China. He received B.S. degree in Automatic Control from Zhengzhou University, PR China, in 1999, and the M.S. degree in Pattern Recognition and Intelligence System from Harbin Institute of Technology, Harbin, Heilongjiang Province, PR China, in 2002. He is currently working toward the Ph.D. degree in the Department of Computer Science, Utah State University. His research interests include image processing, pattern recognition, medical imaging processing, fuzzy logic, and neural networks.

About the Author—LING ZHANG received the Bachelor of Engineering degree from the Department of Computer Science, Shandong University of Technology, Jinan, China, in 1993 and the Master of Science degree from the Department of Computer Science, Shandong University of Technology, Jinan, China, in 1997. From 1993 to now, she worked as a teacher in Shandong University. Now she is a Ph.D. candidate in School of Computer Science and Technology in Shandong University. Her research interests include artificial intelligence, nature language processing, and database.