國立臺灣大學工學院工業工程學研究所

博士論文

Institute of Industrial Engineering

College of Engineering

National Taiwan University

Doctoral Dissertation

基於超音波影像特徵之腫瘤偵測及邊緣擷取-以 2D 甲狀腺

及 3D 乳房影像為例

Tumor Detection and Segmentation based on Sonographic

Features with Examples of 2D Thyroid Images and 3D Breast

Images

邱齡瑩

Ling-Ying Chiu

指導教授：陳正剛 博士

Advisor: Argon Chen, Ph.D.

中華民國 109 年 07 月

July, 2020

To my sweet husband, beloved parents, and dear friends.

中文摘要

欲在超音波影像上執行腫瘤的電腦輔助診斷，需要明確定義腫瘤的位置和邊界。然而，腫瘤本身的生物學特性、超音波成像的物理性質和品質、操作者的主觀認知與操作條件等種種因素，都使得辨識腫瘤的位置及邊界更加困難。

本研究的主題主要聚焦於超音波影像當中腫瘤的自動偵測以及邊緣分割。我們應用了兩種超音波影像：甲狀腺和乳房，分別應用於探討自動邊界分割和位置偵測的問題。我們將提出之方法應用於實際臨床案例的 2D 和 3D 超音波影像上，以評估所提方法之效能。

在邊界分割問題上，我們提出了一種新穎的半自動分割方法，使用 Variance-Reduction 的統計方法來對甲狀腺腫瘤的邊界進行分割，且不需對影像進行預處理。而在位置偵測問題上，我們提出了一種全自動電腦輔助偵測演算法，將 two-phase merge-filter 演算法應用於自動三維乳房超音波影像。

研究結果顯示，我們提出的分割方法對於超音波影像上甲狀腺腫瘤的邊界分割是可靠且有效的。另外，我們提出的電腦輔助偵測系統具備極高的靈敏度，並伴隨相當低的偽陽性值，因而具有高度潛力成為自動三維乳房超音波影像的良好輔助工具。

關鍵詞

自動偵測，自動分割，自動三維乳房超音波，乳房腫瘤，電腦輔助偵測，two-phase merge-filter，腫瘤位置，腫瘤邊界，甲狀腺腫瘤，超音波影像，Variance-Reduction statistic.

ABSTRACT
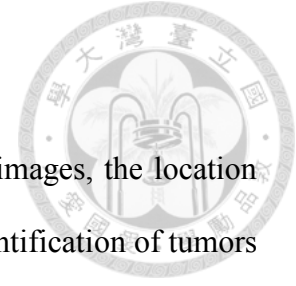
To perform computer-aided diagnosis of tumors on ultrasound images, the location and boundary of tumors should be clearly defined. However, the identification of tumors location and boundary are difficult issues due to the biological characteristics of the tumors, the physics and quality of ultrasound imaging, and the subjective factors and operating conditions of the operator.

The main focus of this research is on the automatic detection and segmentation of tumors in ultrasound images. Two types of ultrasound images, thyroid and breast, were used to explore the issues of automatic boundary segmentation and location detection, respectively. The performances of the proposed methods were then applied on 2D and 3D ultrasound images of actual clinical cases.

In boundary segmentation, a novel and semi-automatic method was proposed for segmenting the boundary of thyroid nodule based on the Variance-Reduction (V-R) statistics without image preprocessing. In location detection, we proposed a fully automatic computer-aided detection (CADe) algorithm applying on automated three-dimensional breast ultrasound (ABUS) images with a two-phase merge-filter algorithm.

It was shown that the segmentation method was reliable and effective in segmenting thyroid nodule boundary on ultrasound images. Moreover, the proposed CADe system had a great potential of becoming a good companion tool of the ABUS imaging by ensuring high sensitivity with a relatively small number of false-positives.
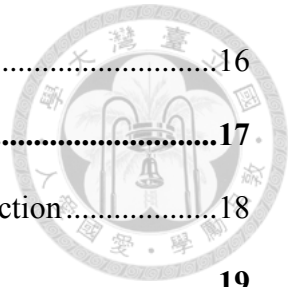
Keywords

Automatic detection, automatic segmentation, automated three-dimensional breast ultrasound, breast tumor, computer-aided detection, two-phase merge-filter, tumor location, tumor boundary, thyroid nodule, ultrasound image, Variance-Reduction statistic.
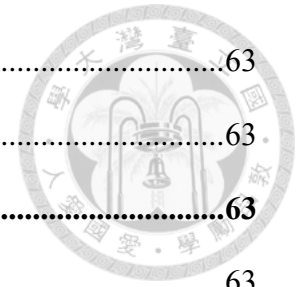
CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1   Introduction

Ultrasound imaging is one of the most effective screening tools for the discovery of tumors. It possesses the advantages of non-invasiveness and irradiation-free, making it a suitable choice in regular screening. By interpreting the ultrasound images, doctors make decisions if the patients requires further examinations or interventions. However, the interpretation of ultrasound images are mostly subjective and highly depend on the experience and judgment of the operator, and the inter-observer variation often results in substantially different decisions. Due to these reasons, how to objectively and quantitatively describe a tumor in ultrasound imaging becomes a pressing issue facing the medical staffs.
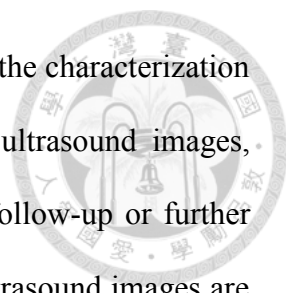
To effectively quantify the sonographic features, an objective accurate tumor location and boundary have to be first determined. Therefore, computer-aided systems in our research are respectively proposed to automatically segment boundary of tumor and detect tumor location on ultrasound images.

1.1 Boundary segmentation for ultrasonic thyroid nodules

1.1.1   Background

Thyroid cancer has increased significantly over the past four decades in United States [1]. The advance in diagnostic tools allows the detection of small thyroid nodules [1]-[4]. Although early treatment is the key to the cure of cancer and further reduces the mortality rate [3], it may also lead to over-diagnosis with unnecessary biopsies and treatments [1]-[4]. Therefore, an appropriate screening tool with effective diagnosis ability has become an increasingly important issue [5]-[7].

Ultrasound imaging is one of the most effective non-invasive screening tools for the detection of thyroid nodules [6]-[14]. Beside conventional B-mode sonography,

1

elastography which reflects tissue stiffness plays a promising role in the characterization of thyroid nodules [17]. Based on impressions of observing the ultrasound images, clinicians make suggestions for patients to be subject to periodic follow-up or further cytological tests [8]. However, the acquisition and observation of ultrasound images are mostly subjective and highly depend on the experience and judgment of the operator, and the inter-observer variation often results in significantly different decisions [6], [11]-[13]. According to aforementioned reasons, objective quantification of the sonographic thyroid nodule features has become a pressing issue facing the medical staffs [6]. To effectively quantify the sonographic features, an objective accurate nodule boundary has to be first determined. Therefore, computer-aided system was proposed to automatically detect and identify the boundary of thyroid nodules on ultrasound images [9]-[14].

1.1.2 Objective

For computer-aided systems, speckle is one of the most challenging problems [9]-[14], [15], [16]. To reduce the impact of speckle, previous publications used image preprocessing methods or altered the detailed information provided during the execution steps [9]-[11], [14]. However, speckle also carries information about tissue characteristics and may be exploited in several applications. Approaches in attempt to reduce speckle may result in the loss of such information within original ultrasound images, and may even affect the capturing of nodule features and the correctness of the computer-aided system operation.
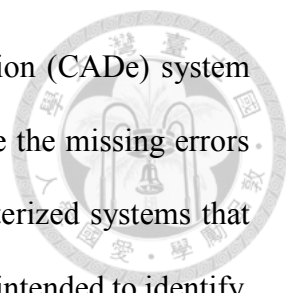
To the best of our knowledge, this is the first work to apply Variance-Reduction (V-R) statistics-based approach to detect nodule boundaries on thyroid ultrasonography. In this study, we propose a novel and semi-automatic method without image preprocessing.

2

1.2 Breast tumor detection in 3D ultrasound imaging

1.2.1 Background

Breast cancer has the highest incidence and causes the second leading mortality of cancer among women in 2017 in the United States [18]. Since early treatment is the key to the cure of breast cancer, effective diagnosis system and the adjunctive detection system have become an important issue to further reduce the mortality rate [19].

Mammography [20] and breast ultrasound [21] are two of the most commonly used screening tools for the diagnosis and detection of breast tumor. Each of them has its own advantage and disadvantage in clinical examination [22]. Mammography is adopted as a major screening tool for its high sensitivity in detecting micro-calcification and can be used to further detect early changes in abnormal tissues, but it also suffers from high false positive rate especially in women with dense breast tissue [23]-[26]. Breast ultrasound is the most important alternative to mammography [23]-[25], with the advantages of radiation-free and less pain, but ultrasound imaging highly depends on the medical staff's experience and judgment, resulting in inter-observer variations that often lead to significantly different clinical decisions, and the results are poorly reproducible [26]. Based on these problems, automated breast volume scanner (ABVS) as an innovative high-end ultrasound scanner has been developed by combining the advantages of mammography and breast ultrasound to automatically scan the entire breast [9]-[11]. Additionally, ABVS allows the scan results to have less operator dependency and higher reproducibility, and provides coronal images to further assist the planning of surgical intervention [26], [27]. However, the reviewing process to discover suspicious abnormalities from hundreds of image slices produced by ABVS is often time-consuming [29], [30]. Besides, tumor which is tiny or isoechogenic to the surrounding normal tissue

3

may be missed by reviewer [30]. Therefore, computer-aided detection (CADe) system has been proposed to accelerate the reviewing process and to reduce the missing errors [33]-[39]. As defined by the U.S. FDA, "CADe devices are computerized systems that incorporate pattern recognition and data analysis capabilities and are intended to identify, mark, highlight, or in any other manner direct attention to portions of an image, or aspects of radiology device data, that may reveal abnormalities during interpretation of patient radiology images or patient radiology device data by the intended user" [40]. CADe system is needed to solve a practical clinical problem of ABVS where thousands of images were acquired by automatically scanning the patient breasts and will be time-consuming and tiresome for physicians to review the slice one by one to discover any suspicious lesions.

1.2.2 Objective

The aim of this study is to develop an intuitive, simple, efficient CADe solution to quickly screen all the images generated by the automated three-dimensional breast ultrasound (ABUS), in particular Siemens ACUSON2000 ABVS in this study, to highlight all possible lesions for further review and diagnosis by physicians.

Chapter 2 Literature Review
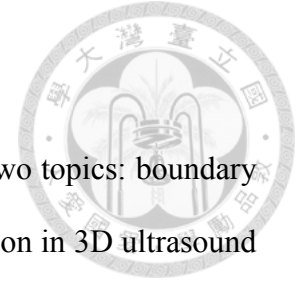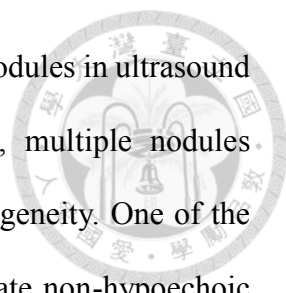
In this chapter, the discussion of our research is separated into two topics: boundary segmentation for ultrasonic thyroid nodules and breast tumor detection in 3D ultrasound imaging. For each topic, we will review related works and summarize the advantages and disadvantages. Then, we will introduce some of the methods related to our proposed methods.

2.1 Boundary segmentation for ultrasonic thyroid nodules

Various methods have been employed for computer-aided systems to automatically detect and identify the boundary of thyroid nodules on ultrasound images.

Tsantis *et al*. [9] proposed a hybrid multi-scale model which integrated the wavelet edge detection procedure and the Hough transform to extract the final contour of thyroid nodules in ultrasound images. This method was advantageous by combining the ability of wavelet transform to detect sharp variation and the efficiency of Hough transform to discriminate target from noisy background. Nevertheless, it also had several shortcomings that the delineation outcome was highly affected by human error, and the step of Hough transform was time-consuming. Besides, it required prior knowledge about the shape of nodule to be detected, which limited the value in practical use. Iakovidis *et al.* [10] presented a segmentation framework applied on thyroid ultrasonography by incorporating a level set approach named Variable Background Active Contour model and a parameter tuning mechanism based on Genetic Algorithm to search for optimal parameters automatically. This GA-VBAC framework was relatively unaffected by intensity inhomogeneity in the thyroid ultrasound images. However, this method required huge amount of time in its training phase, which posed a major limitation. Maroulis *et al.* [11] developed an algorithm of Variable Background Active Contour model based on the

Active Contour Without Edges model for the delineation of thyroid nodules in ultrasound images. This algorithm had the advantages of noise robustness, multiple nodules delineation capability, and the ability to cope with intensity inhomogeneity. One of the drawbacks of their proposed algorithm was that it could not delineate non-hypoechoic thyroid nodules. Savelonas et al. [12] presented a joint echogenicity–texture model based on a modified Mumford–Shah function that incorporated regional image intensity and statistical texture information. Their approach were capable of segmenting hypoechoic, isoechoic, and hyperechoic nodules. It required no prior assumption about the shape of nodule to be detected, and was noise-tolerant. Nevertheless, it also had some limitation that it might get confused by anatomical structures within nodules such as blood vessels. Koundal et al. [14] proposed an automated delineation method that integrated spatial information with the Neutrosophic L-Means clustering and level-sets method for the segmentation of thyroid nodules in ultrasound images. This method first segmented possible target via Spatial Neutrosophic L-Means clustering, then applied Distance Regularizer Level Set Evolution method to delineate the nodule contour. It had benefits of noise robustness and multiple nodules delineation capability, and could handle intensity inhomogeneity as well. However, it had weakness in its ability to detect isoechoic thyroid nodules. Ma et al. [31] employed a deep convolutional neural network for thyroid nodule segmentation based on 2D ultrasound images. The convolutional neural network was fed with image patches from thyroid nodules and normal thyroid as input, then output the segmentation probability. This method could delineate thyroid nodules accurately and was capable of multiple nodules delineation. It was also noise-tolerant. Despite of these, this deep-learning based method showed unsatisfying performance while dealing with heterogeneous nodules or complicated background. Chang et al. [32] developed a method to retrieve thyroid nodule contour on the basis of a

6

manual input contour. The manual input contour was then processed with Active Contour Method into a smooth thyroid nodule contour, and the sonographic features of the region enclosed by the contour were calculated.
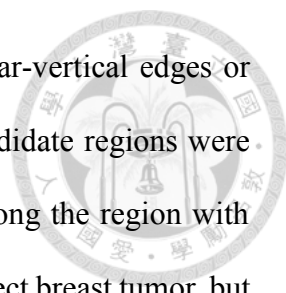
In recent years, deep learning has gradually been discussed and valued in various field of applications. Among them convolutional neural network (CNN), such as Ma et al. [31] used in their work, has gained the most attention. However, unlike traditional knowledge-based methods, the "black-box" nature of these CNN-based system may lead to severe consequences and arouses incremental concerns recently. CNN requires huge amount of ground truth data for training, and we even do not know what it learns, how it learns, and where it errs.

The region of interest (ROI) delineation usually plays an important role in ultrasound image analysis. With ROI being delineated, the computation cost of ultrasound image analysis is greatly reduced, and the accuracy is improved. However, the delineation of ROI for ultrasound images is not always readily available in clinical settings. It adds an extra step beyond the standard operations of ultrasound imaging, and the ROI delineation step itself is error-prone if not performed by experienced operator. To take advantage of ROI delineation and simultaneously avoid its drawback, it is essential to incorporate the step of ROI delineation with the standard operation procedures of ultrasound examinations.
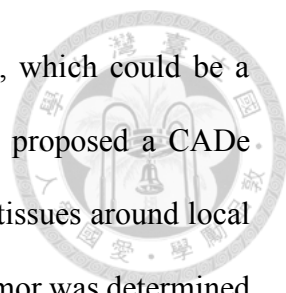
2.2 Breast tumor detection in 3D ultrasound imaging

Computer-aided detection system has been proposed to accelerate the reviewing process and to reduce the missing errors. Recently, the issue has received special attention and discussion in the 3D ultrasound images.

Ikedo *et al*. [33] proposed a fully automatic scheme based on the edge information

7
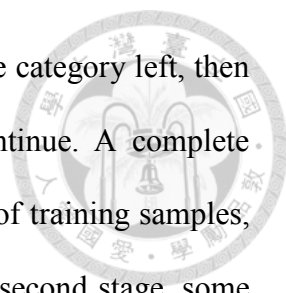
detected by Canny edge detector and further classified edges as near-vertical edges or near-horizontal edges by a morphological method. Initial tumor candidate regions were then generated by watershed algorithm and density information among the region with near-vertical edges. This method achieved sensitivity of 80.6% to detect breast tumor, but has difficulty in detecting flat-shaped tumors due to their short vertical dimensions. Besides, it frequently misinterpreted fat and rib regions. Chang *et al*. [34] used the gray-level slicing method to segment images into numerous regions, and then seven quantitative features were extracted to further determine each region whether a part of tumor or not. Their proposed system achieved decent sensitivity of 92.3%, but had a long processing time. Besides, the case number in this study was small, which hampered the reliability of their conclusion. Moon *et al*. [35] developed a CADe system based on algorithm of Hessian analysis with multi-scale blob detection. Then three categories of quantitative features were extracted to estimate the tumor likelihood with a logistic regression model. Their proposed system could detect every tumor and achieved the sensitivity of 100%, but accompanied with a lot of erroneous detections. Besides, their study did not involve early-stage breast cancers, so the ability of detecting small malignant tumors was unknown. Tan *et al*. [36] presented a two-stage CADe system. After landmarks such as nipple and chest wall were provided, tumor candidates were generated by using voxel features including coronal spiculation patterns, blobness, and contrast. The likelihood of malignancy was evaluated with an ensemble of five neural networks. A major shortcoming of their proposed system was that the maximal achievable sensitivity was suboptimal. Moon *et al*. [37] proposed a CADe system based on quantitative tissue clustering algorithm. The fuzzy c-means clustering was used to generate tumor candidates among the regions segmented by fast 3D mean shift method. Seven quantitative features were used to estimate the likelihood of the candidates being

8

tumors. A limitation of their study was the high false positive rate, which could be a problem for clinical applications of CADe systems. Lo *et al*. [38] proposed a CADe system based on topographic watershed transform gathering similar tissues around local minima into homogeneous regions. The likelihood of region being tumor was determined with quantitative features of morphology, echogenicity, and texture. The proposed CADe system achieved the sensitivity of 100% with tolerable false positive rates. Besides, the processing time was satisfactory. There were some limitations in their study though. Their system might not perform well for all kinds of breast carcinomas. Furthermore, the data used in their study all came from a single medical center and might hinder the generalization of their system.

Utilizing deep learning, Chiang *et al*. [39] proposed a method based on 3D convolutional neural networks and prioritized candidate aggregation. Their proposed system first implemented a sliding window detector to extract volume of interest, which were then determined their tumor probability by 3D convolutional neural networks. Their proposed system was advantageous for its fast processing time. However, a major drawback of their proposed system was that it did not perform well at high sensitivities and produced a lot of false positives.

Decision tree learning is one of the predictive modelling approaches used in statistics, data mining, and machine learning. Among them, classification and regression tree (CART) [41] is one of the most well-known models. CART is constructed through a sequential binary classification, and its goal is to make the child nodes with a higher homogeneity. The first stage is to build a complete classification tree. At this stage, subjects are binary-classified over and over until classification is no longer possible. The method will search for the best feature and cut point in the parent node, and then use this feature and cut point to divide the parent node into two child nodes. If the number of

9

samples of the child node is too small or the child node has only one category left, then the classification will stop; otherwise, the classification will continue. A complete classification tree is established through the complete classification of training samples, but it usually leads to the problem of over-fitting. Therefore in the second stage, some branches of the complete classification tree need to be pruned to avoid over-fitting.

We construct our detection algorithm inspired by CART, to implement the free-of-parameters advantage thus ideal for our application to analyze huge amount of data contained within 3D ultrasound images.

Chapter 3　Variance-reduction methods for boundary segmentation
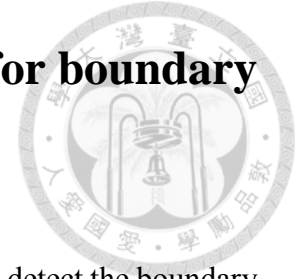
In this study, we developed a novel and semi-automatic system to detect the boundary from thyroid nodule on ultrasound images. The architecture of our proposed method in this chapter was shown as flow chart and schematic diagram in Fig. 3-1 and Fig. 3-2, respectively. Firstly, region of interest (ROI) was automatically generated according to the initial inputs of the nodule's major and minor axes. Subsequently, the boundary candidate points were extracted using the V-R statistics. Three filtering methods including the direction searching method, the outlier elimination method, and the inner product method were used in sequence to further filter out the outliers among the remaining boundary candidate points of the previous step. The remaining boundary candidate points were then smoothened and linked together to form the final boundary.
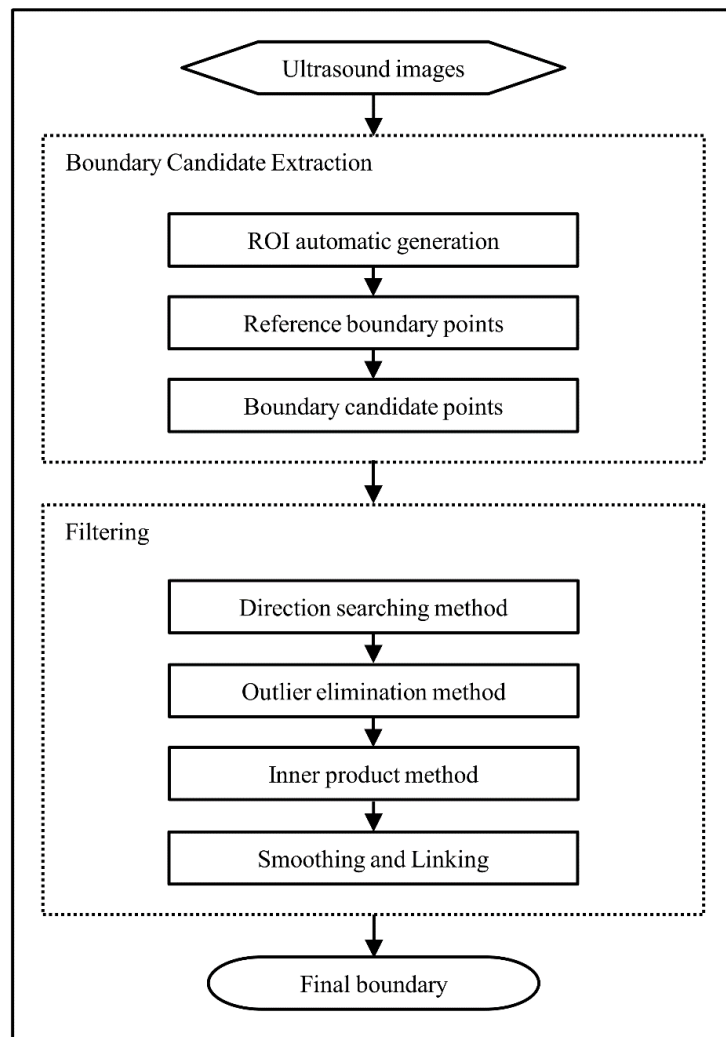
Fig. 3-1. The flow chart of our proposed method.
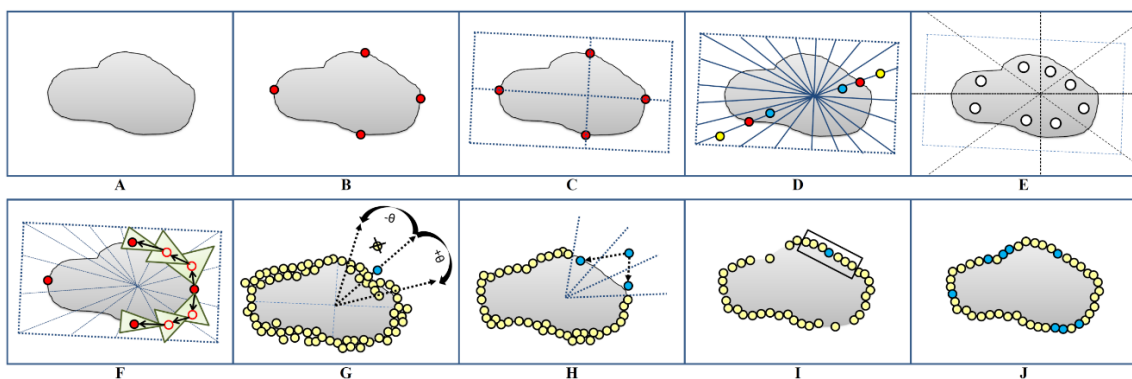
Fig. 3-2. Schematic diagram of semi-automatic segmentation of nodule boundary. (A) thyroid nodule. (B) manually input four extreme nodes on approximate location of the nodule. (C) the region of interest (ROI) was automatically generated based on the major

axis and minor axis. (D) reference boundary points. (E) new cut points and new radial lines defined by the centers of the eight 45-degrees sectors in the ROI. (F) direction searching method. (G) outlier elimination method. (H) inner product method. (I) smoothing. (J) linking.
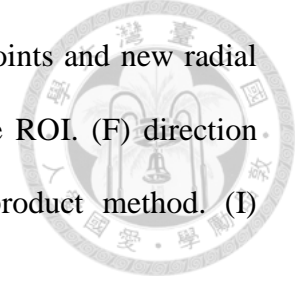
3.1 Boundary Candidate Extraction

The nodule boundary on the ultrasound image was generally characterized by significant difference in grayscale values between the inside and outside of the nodule boundary. Therefore, we proposed a method that automatically detects the thyroid nodule boundary with segmentation and identification methods by combining the radial gradient algorithm and the V-R statistics.

3.1.1   ROI automatic generation

Before the proposed method was implemented, the ROI was automatically generated based on the approximate major axis and minor axis of the nodule manually input by medical staff, as shown in Fig. 3-2 (A) to (C).

3.1.2   Reference boundary points

The radial lines were then grown from the center of the ROI and three original cut points on each radial line were searched by the V-R statistics so that each cut point resulted in minimum sum of two group total sum of square, as in (3-1), Fig. 3-2 (D) and Fig. 3-3.

$$\min_{a}\left\{\sum_{i=1}^{a-1}\left(x_i-\mu_1\right)^2+\sum_{i=a+1}^{N}\left(x_i-\mu_2\right)^2\right\} \tag{3-1}$$

In (3-1) and corresponding Fig. 3-3, a radial line has N pixel points and their grayscale values are represented as a set $\{x_1, x_2, \ldots, x_a, \ldots, x_{N-1}, x_N\}$ where x_i is the grayscale value of the pixel point i in the order from the ROI center to the ROI boundary, and μ_1 and μ_2

13

are the averages of grayscale values of the inner pixels and outer pixels which are separated by pixel point a, respectively. After searching pixel point a throughout N pixel points, the first cut point is defined as the one resulting in minimal sum of two total sum of squares and further marked with red. Moreover, two additional cut points are defined by repeating the above process from inner and outer segments of the radial line separated by the first cut point. The two newly generated cut points are then marked with blue and yellow, respectively, as shown in Fig. 3-2 (D) and Fig. 3-3. Three original cut points from all radial lines are regarded as reference boundary points.
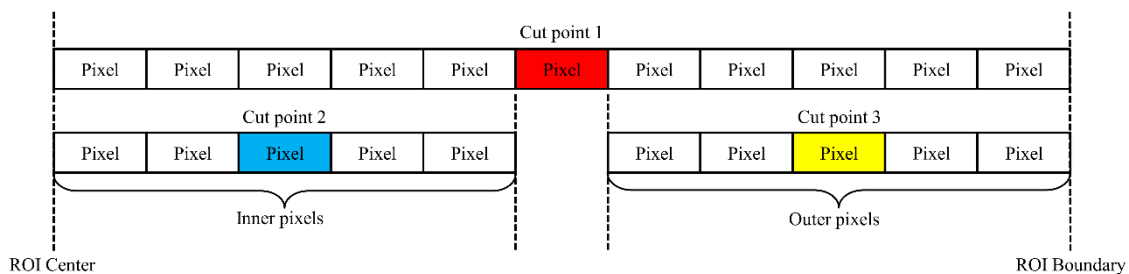
Fig. 3-3. The radial line was grown from the center of the ROI and then three cut points were searched on each radial line by Variance-Reduction statistic.

3.1.3 Boundary candidate points

In order to prevent the influence from the variation of the nodule's inner tissues, we further searched more useful cut points based on (3-1) and the reference boundary points. Firstly, new radial lines were defined by the new centers of the reference boundary points within eight 45- degrees sectors in the ROI, as shown in Fig. 3-2 (E). Secondly, three cut points were generated by applying (3-1) on each new radial lines and were then regarded as boundary candidate points.

14
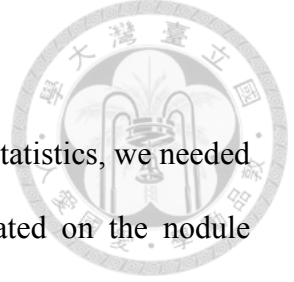
3.2 Filtering

With the plenty boundary candidate points extracted by the V-R statistics, we needed to determine which candidate points were more likely to be located on the nodule boundary. To further obtain suitable boundary candidate points, we developed three filtering methods including the direction searching method, the outlier elimination method, and the inner product method.

The filtering methods were executed based on the information of the four nodes, which provided approximate locations of the nodule boundary points. The nodule boundary ought to go through these four nodes, so our filtering method worked in a manner that boundary candidate points should be on the path going through these four nodes.

3.2.1   Direction searching method

The direction searching method started from one of the four nodes and defined the feasible region according to the path direction from one node served as initial starting point toward its two neighboring nodes, respectively. The path was further divided into many parts to increase the number of searching steps in order to ensure accuracy. The searching condition was set as that if a boundary candidate point was located in the feasible region and satisfied the criterion of the distance from the starting point, then it was preserved, otherwise it would be filtered as in (3-2).

$$|d_{ij}| > w \quad or \quad |\theta_{nij}| > \theta_{Threshold} \qquad (3\text{-}2)$$

In (3-2), d_{ij} represents the distance between starting point i and the boundary candidate point j, w is a parameter tuning the threshold. θ_{nij} represents the angle formed by neighboring node n, starting point i and boundary candidate point j. $\theta_{Threshold}$ represents the feasible region.

15

The process was separately performed on both sides of the initial starting point, iteratively replacing the starting point by the center of the points satisfying the condition in previous step, and gradually decreasing the feasible region in each iteration to ensure convergence onto the neighboring nodes. The iteration stopped at the arrival of the two neighboring nodes, and then the above process restarted with another node as a new initial starting point. Finally, the whole process ceased after all four nodes completed the searching process, as shown in Fig. 3-2 (F).

3.2.2 Outlier elimination method

Since an appropriate nodule boundary should not have candidate points extremely outlying from other points, the outlier elimination method and the inner product method were used after the direction searching step to eliminate these outlier points.

The outlier elimination method was implemented to filter out the boundary candidate points which were relatively distant from other boundary candidate points in a sector area, as shown in (3-3) and Fig. 3-2 (G).

$$\left| d_{ic} \right| > \left(\mu_K + a\sigma_K \right) \tag{3-3}$$

In (3-3), there are k boundary candidate points in the sector area K, and the distance between the ith boundary candidate point and the center of ROI are represented as the set of $\{d_{1c}, d_{2c}, …, d_{ic}, …, d_{(k-1)c}, d_{kc}\}$. μ_K and σ_K are the mean and standard deviation of all distance values in the sector area K, respectively, and a is a parameter tuning the threshold.

3.2.3 Inner product method

The inner product method filtered out the boundary candidate points resulting in a sharp angle between two neighboring boundary candidate points, as shown in (3-4) and Fig. 3-2 (H).

$$\left(\vec{V}_{i(i+1)} \quad \bullet \quad \vec{V}_{i(i-1)} \right) > b \tag{3-4}$$

16

In (3-4), we sort all boundary candidate points in order, where $\vec{V}_{i(i-1)}$ and $\vec{V}_{i(i+1)}$ are vectors from boundary candidate point i to its two neighboring boundary candidate points $(i-1)$ and $(i+1)$, respectively, and b is a parameter tuning the threshold.
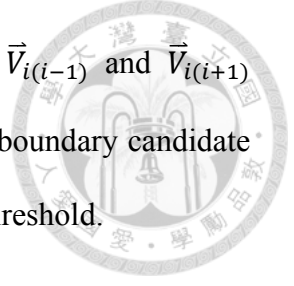
3.3 Smoothing and Linking

After these three filtering steps, the remaining boundary candidate points were smoothened (Fig. 3-2 (I) and (3-5)), and extra points were filled in to link the distant neighboring boundary candidate points to form the final boundary (Fig. 3-2 (J)).

Smoothing process applied a pulling effect on each remaining boundary candidate point from its surrounding neighborhoods, in order to smoothen the extreme convexity or concavity of the remaining boundary candidate points, as in Fig. 3-2 (I) and (3-5).

$$(X_i', Y_i') = (\frac{\sum_{w=i-s}^{i+s} X_w}{2s+1}, \frac{\sum_{w=i-s}^{i+s} Y_w}{2s+1}) \tag{3-5}$$

Equation (3-5) shows the new coordinate (X_i', Y_i') of the boundary candidate point i which is smoothened by taking the average of coordinates from the $(i\text{-}s)^{\text{th}}$ candidate point to the $(i\text{+}s)^{\text{th}}$ candidate point.

The linking process took place when a boundary candidate point i and its neighbor point $i+1$ were not adjacent in 8-connected manner. Then the extra points were added to the line connecting point i and point $i+1$ to fill the gap. The linking process repeated until there was no gap in the boundary. The extra points filled in were shown with blue color in Fig. 3-2 (J).

Chapter 4　Two-phase merge-filter methods for 3D tumor detection
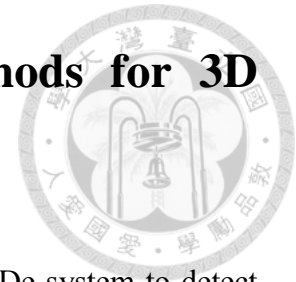
We developed an intuitive, simple, efficient, and automatic CADe system to detect breast tumor on ABUS images. The detection method proposed in this chapter was shown in Fig. 4-1 and described as follows.

The detection method was performed in a two-phase manner. First, in two-dimensional (2D) phase, after image preprocessing, square cells with hypoechogenicity and homogeneous texture were extracted and merged into initial tumor candidates [50]. Intensity-related features were then generated to characterize the 2D candidates to subsequently distinguish tumors from non-tumors. A classifier was then constructed to stage-wise classify 2D candidates and to filter out non-tumor ones efficiently. Secondly, in three-dimensional (3D) phase, remaining 2D candidates were merged longitudinally into 3D masses, and the merged 3D masses were then characterized by four types of features. Finally, the classifier was used again based on 3D features to further filter out non-tumor masses to obtain the final detected masses.
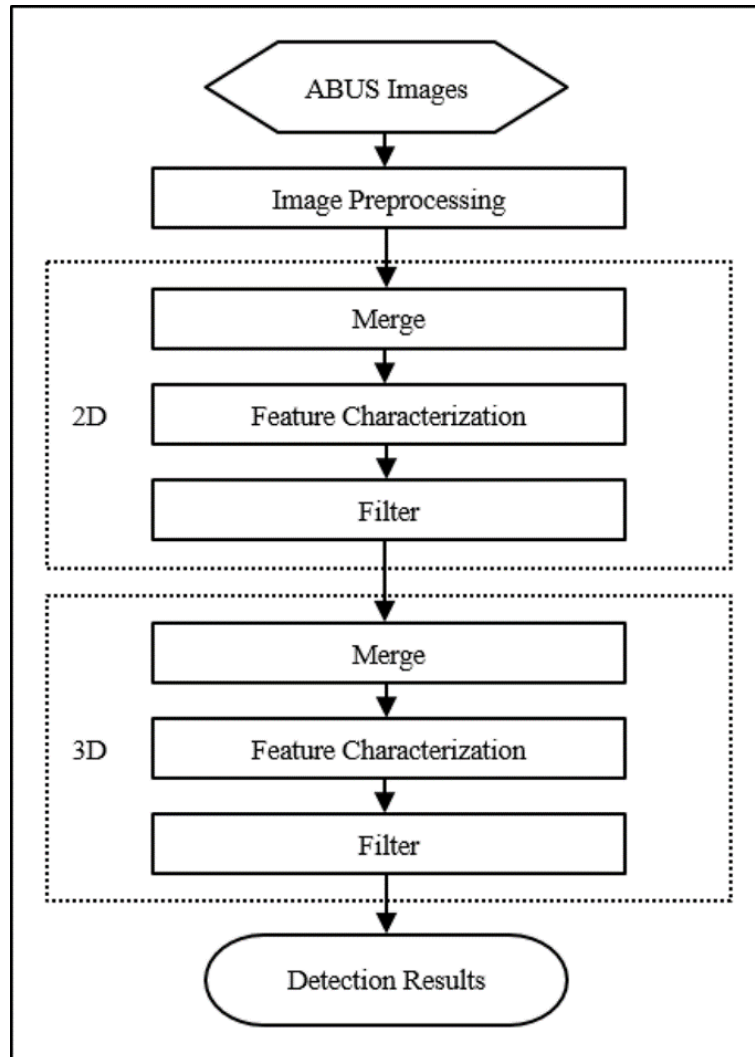
Fig. 4-1. The flow chart of our proposed CADe system.

4.1 Image Preprocessing

The ABUS image pass consists a series of 318 slices with plenty pixel points providing lots of image data information. Because of this, the analysis of the ABUS images is usually a time-consuming reviewing process [29], [30]. Therefore, according to the subtle variation between the image slices [38], we extracted one image slice from every five slices for better efficiency. Furthermore, to acquire an appropriate image region for subsequent analysis, we excluded two lateral shadow regions in each image which were caused by the gap between the transducer and the breast skin during scanning. The

result of the appropriate region of interest for analysis was shown in Fig. 4-2 (B).

Ultrasound images generally contain plenty of speckle and low contrast due to the complex imaging environment and imaging principle. Accordingly, the quality of the ultrasound image may be diminished, resulting in ineffective diagnosis or interpretation. Besides, ultrasound images may appear globally dark or bright between different image slices, and this leads to an unfavorable phenomenon in dealing with an image pass. Histogram equalization (HE) [50], [51] is a famous methods for image processing to enhance contrast [52] and uniform the probability distribution of the grayscale values without destructing the important features. However, HE may enhance meaningless detail and hide important but small high-contrast features [53]. To avoid these effect, we applied a modified HE algorithm contributed by Richard Kirk [53] combining a transformation function by using the square root of the histogram values. After preprocessing, images in each pass had uniform grayscale distribution and showed more detail in the suspicious regions. The result was shown in Fig. 4-2 (C).
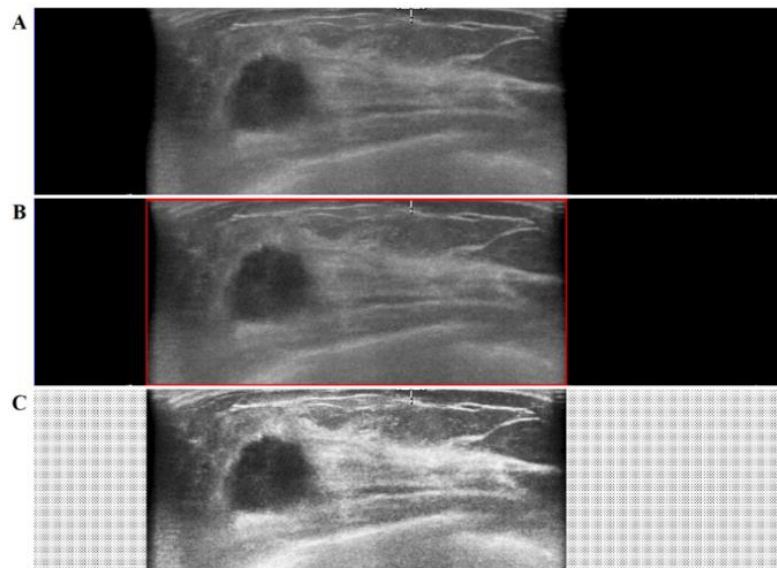
Fig. 4-2. Steps of Image Preprocessing. (A) The original ABUS image. (B) The red rectangle showed the appropriate region for analysis excluding two lateral shadows. (C) The result of applying similar HE algorithm.

20
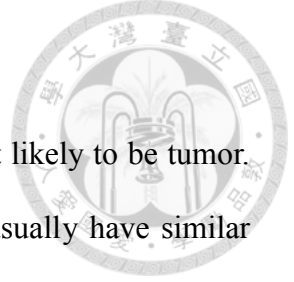
4.2 2D merge

In this step, we intended to extract candidates for the region that likely to be tumor. According to clinical experience, tissues within the tumor region usually have similar echotextures [37]. The procedure was taken in three steps, not only extracted the initial candidates, but also reduced the processing time for the subsequent steps.

Firstly, all pixel points on an image slice were cut into non-overlapping square cells, as shown in Fig. 4-3 (A). Secondly, owing to the hypoechogenicity of breast tumors [37], we collected the square cells with average grayscale value lower than twenty-fifth percentile among all grayscale values within an image slice. The collection of these hypoechogenic square cells were regarded as a set of appropriate analysis regions, as shown in Fig. 4-3 (B).

Thirdly, if two adjacent appropriate analysis regions were similar in grayscale values, they were merged together. Modified t-statistic and F-statistic were adopted to judge similarity in grayscale values due to clinical observation that tumor region usually has similar echotextures [37]. Two adjacent appropriate analysis regions satisfying the following two criteria were merged together: similarity in average grayscale values judged by modified t-statistic and similarity in standard deviations of grayscale values judged by F-statistic [57]. Furthermore, the denominator used in modified t-statistic was changed to the square root of overall average variance from all appropriate analysis regions to avoid the effect of outliers on an image slice. Modified t-statistic and F-statistic were formulated as (4-1) and (4-2), respectively.

$$T_{jk} = \left| (\mu_j - \mu_k) \middle/ \sqrt{\frac{\sum_{m=1}^{N_i} \sigma_m^2}{N_i}} \right| \le t \tag{4-1}$$

$$F = \begin{cases} \dfrac{\sigma_j^2}{\sigma_k^2} \leq f & , \quad if \quad \sigma_j^2 \geq \sigma_k^2 \\[3mm] \dfrac{\sigma_k^2}{\sigma_j^2} \leq f & , \quad if \quad \sigma_j^2 < \sigma_k^2 \end{cases} \qquad (4\text{-}2)$$

If an image slice i contains N_i appropriate analysis region, and region j contains texture information about mean grayscale value μ_j and standard deviation value σ_j. T_{jk} and F represent the modified t-value and F-value between two adjacent regions j and k, respectively, where $j < k$. t and f are parameters tuning the threshold for T_{jk} and F, respectively.

The merging procedure was iteratively performed until there were no adjacent region composing statistically similar tissue texture. The products of merging procedure were referred as initial tumor candidates, as shown in Fig. 4-3 (C). The scheme of whole mergence procedure was demonstrated in Fig. 4-3.
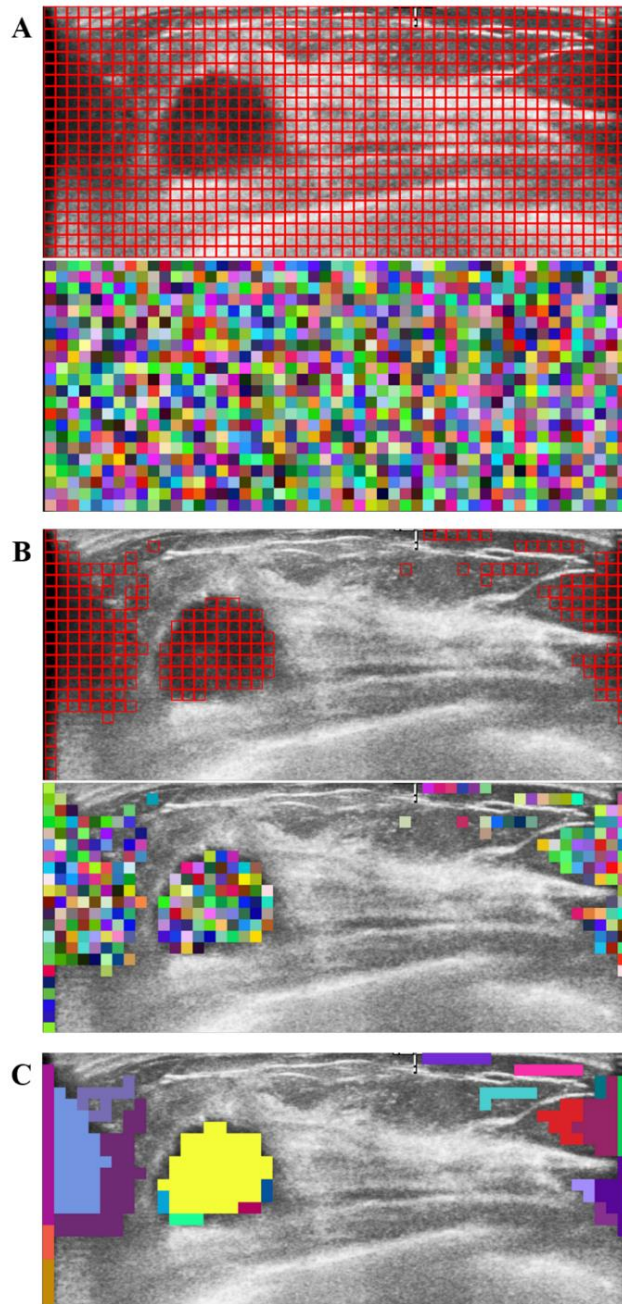
Fig. 4-3. Step-by-step visualization of 2D merge. (A) All pixels on image slice were cut into non-overlapping square cells. The upper one showed the schematic diagram of the square cells, and the lower one visualized each individual square cell with random color. (B) The appropriate analysis region set was a collection of the hypoechogenic square cells with average grayscale value lower than twenty-fifth percentile. (C) Adjacent appropriate analysis regions with similar grayscale values were merged together to form initial tumor candidates.

23
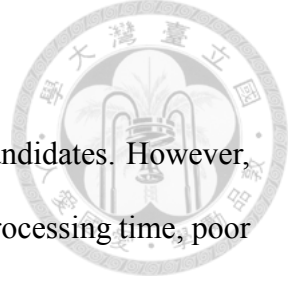
4.3 2D Features Characterization

After 2D merge, these regions were regarded as initial tumor candidates. However, too many candidates among non-tumor areas may cause prolonged processing time, poor performance, and frequent false positives (FPs). To further distinguish tumor from non-tumor candidates, we extracted intensity-related features in 2D phase to characterize candidates in ABUS images [54]-[56].

The intensity of candidate within tumor region is generally darker than that of the surrounding tissues. Based on this characteristics, two intensity-related features were generated to distinguish tumor and non-tumor candidates, namely intensity mean value ($I(c)_{mean}$, defined in (4-3)) and the percentile of $I(c)_{mean}$ within the image slice (*Percentile_$I(c)mean$*).

$$I(c)_{mean} = \frac{1}{n_c} \sum_{p \in c} I_p \qquad (4\text{-}3)$$

Candidate c has a set of pixel points $\{p\}$, and the amount of pixel points in c is n_c. I_p indicates the grayscale value of pixel point p within candidate c.

4.4 3D merge

The 3D phase of our proposed detection method began with an algorithm of 3D merge. 2D candidates in neighboring transverse image slices which met the criterion of continuity were then merged longitudinally by volume growth characteristics. Finally, the merged 3D objects were referred as masses.

The criterion of continuity were based on the fact that tumors generally distribute in a spatially continuous manner. First, we searched a square cell belonging to one 2D candidate and projected it onto corresponding position in neighboring slice. Then we checked if the projected square cell located in a 2D candidate on that slice or contacted

24

with it in 8-connected manner. If so, then the two 2D candidate were merged together because of high likelihood belonging to the same tumor. The procedure iteratively repeated until there were no 2D candidates meeting the criterion of continuity. The schematic diagram was shown in Fig. 4-4.
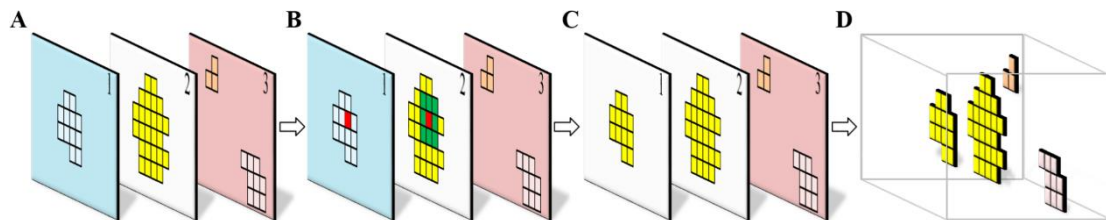
Fig. 4-4. 3D merge. (A) Tumor candidates from 2D phase were marked in different colors. (B) Criterion of continuity: we searched a square cell belonging to one 2D candidate and projected it onto corresponding position in neighboring slice. Then we checked if the projected square cell located in a 2D candidate on that slice or contacted with it in 8-connected manner. (C) 2D Candidates in slice 1 and slice 2 satisfied the criterion of continuity and thus merged together. The two 2D candidates on slice 3 were left untouched due to non-continuity. (D) The schematic view of image (C) in 3D space. The merged candidates (yellow) were referred as masses.

4.5 3D Features Characterization

To further improve the performance after 3D merge, we extracted four types of features to characterize 3D masses in ABUS images for further classification between tumor and non-tumor [54]-[56]. The features extracted are solely for lesion detection and are different from those features specified in BIRADS for breast cancer diagnosis.

4.5.1   Morphology

Morphology features were mainly quantified from the relevant information of the 3D mass boundary and provided the most useful and direct information for tumor detection.

We implemented eight morphological features in three major categories: size-related features, shape-related feature, and aggregation-related feature.

Four size-related features, the size of mass (*Size*), *Length*, *Width*, and *Height*, were implemented to exclude speckles and shadows in ABUS images. The speckles appearing in ABUS images are small, whereas the shadows shown in the bottom, near-nipple region, and two lateral sides usually occupy a large area. Consequently, *Size* could be used to filter masses with extremely small or extremely large size. The *Length*, *Width*, and *Height* were the dimensions of the minimal bounding cube containing the mass. These three features were auxiliary to help distinguish speckles and shadows in ABUS images.

Three shape-related features including the Length-to-Width ratio ($Ratio_{L-W}$), Length-to-Height ratio ($Ratio_{L-H}$), and Width-to-Height ratio ($Ratio_{W-H}$) were introduced to filter outliers such as fat, ribs, and shadows in ABUS images. Fat tissues generally have flat and thin shape in horizontal plane, while the ribs are as dark as tumor in ABUS images but have $Ratio_{W-H}$ around 1. Shadows described above generally have narrow *Width* and span across several image slices leading to a large $Ratio_{L-W}$. According to these properties, the ratios were introduced to filter the fat-like masses (large $Ratio_{L-H}$ or large $Ratio_{W-H}$), rib-like ones ($Ratio_{W-H}$ around 1), and shadow-like ones (large $Ratio_{L-W}$). The definitions of three ratios were shown in (4-4), (4-5), and (4-6), respectively.

$$Ratio_{L-W} = \frac{Length}{Width} \tag{4-4}$$

$$Ratio_{L-H} = \frac{Length}{Height} \tag{4-5}$$

$$Ratio_{W-H} = \frac{Width}{Height} \tag{4-6}$$

26

The aggregation-related feature was the ratio of size-to-bounding cube ($Ratio_{S-B}$). This feature was defined to describe the aggregation of the mass. A mass with higher degree of aggregation was more likely to be a tumor. The $Ratio_{S-B}$ was defined as (4-7).

$$Ratio_{S-B} = \frac{Size}{Length \times Width \times Height} \tag{4-7}$$

4.5.2 Texture

We generated five texture-related features in the 3D phase. The main three of the five features, $I(m)_{mean}$, $I(m)_{stdev}$, and the coefficient of variation ($CV(m)$), were used to distinguish tumor from non-tumor masses. Their definitions were shown in (4-8), (4-9), and (4-10), respectively.

$$I(m)_{mean} = \frac{1}{n_m} \sum_{p \in m} I_p \tag{4-8}$$

$$I(m)_{stdev} = \sqrt{\frac{1}{n_m} \sum_{p \in m} \left(I_p - I(m)_{mean} \right)^2} \tag{4-9}$$

$$CV(m) = \frac{I(m)_{stdev}}{I(m)_{mean}} \tag{4-10}$$

Mass m has a set of voxel points $\{p\}$, and the amount of voxel points in m is n_m. I_p indicates the grayscale value of voxel point p within mass m.

Furthermore, to get rid of the influence from the different distribution of the intensity between slices and to consider the relative echogenic characteristics within single image slice, we implemented another two texture features, the rank ratio of intensity mean value ($IR(m)_{mean}$) and the rank ratio of intensity standard deviation ($IR(m)_{stdev}$) among all masses M in a pass, as an auxiliary consideration. The definitions of these two texture features were shown in (4-11) and (4-12), respectively. $|M|$ represents the number of all masses in a pass.

27

$$IR(m)_{mean} = \frac{\underset{m \in M}{Rank}\{I(m)_{mean}\}}{|M|} \qquad (4\text{-}11)$$

$$IR(m)_{stdev} = \frac{\underset{m \in M}{Rank}\{I(m)_{stdev}\}}{|M|} \qquad (4\text{-}12)$$

4.5.3 Location

Four location features were generated based on the anatomy of the breast and the interpretation of ABUS images. The breast tissues generally distribute obliquely from the superior-lateral to inferior-medial side. Consequently, the location of the 3D mass in foot-to-head direction ($Location_{F\text{-}H}$) was implemented as one of the features.

In ABUS images, ribs sometimes look similar to breast tumors. To exclude the ribs, we introduced a feature in addition to aforementioned morphology features, *Depth*, according to the fact that the distance between skin and rib is greater than that between skin and breast tumor. In other words, the masses in the rib region usually have a large *Depth*. Furthermore, the rank ratio of Depth (*DR*) with respect to all masses in a pass was also considered as location feature to further filter out the outliers.

Besides, nipple often results in shadows on ABUS images. To exclude these shadows, we introduced a feature, normalized distance between mass and nipple marker ($Distance_{M\text{-}N}$), to improve the performance of filtering. This feature also helped us ruling out lateral shadows extremely far from the nipple. To normalize the effect of breast size between people, $Distance_{M\text{-}N}$ was derived by the ratio of measured mass-to-nipple distance relative to the width of breast. $Distance_{M\text{-}N}$ was based on operator-input nipple marker, which is an essential step required by ABVS Workplace system to finalize the examination [58]. Due to the existence and correctness of the nipple marker, we used this as a-priori knowledge to generate the feature.
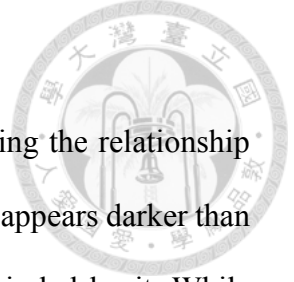
4.5.4   Rim Effect

We considered the tissue characteristic on ABUS images regarding the relationship between tumor and surrounding breast tissue. Breast tumor generally appears darker than surrounding tissue and grows from the mammary gland and is encircled by it. While tumor grows, tissue surrounding the tumor is compressed. Accordingly, tumor is generally accompanied with a thin rim of hyperechogenic tissue on the ABUS images and has obvious intensity difference from the tumor interior, as shown in Fig. 4-5. A total six features were constructed to describe this fact, including the maximum upper variance (Var_U, defined in (4-13)), the maximum lower variance (Var_L, defined in (4-14)), the sum of these two variances (Var_{UL}, defined in (4-15)), and the respective rank ratio of these three features within a pass.

$$Var_{u,j} = \max{}_{x=x_{\min}}^{x_{\max}} \left\{ \frac{1}{y_c - y_{\min} + 1} \sum_{y=y_{\min}}^{y_c} \left(I(x_i, y) - \mu_{x_i, y} \right)^2 \right\} \qquad (4\text{-}13)$$

$$Var_U = Var_{u,J_a} + Var_{u,(J_a+1)} + ... + Var_{u,(J_b-1)} + Var_{u,J_b}$$

$$Var_{l,j} = \max{}_{x=x_{\min}}^{x_{\max}} \left\{ \frac{1}{y_{\max} - y_c + 1} \sum_{y=y_c}^{y_{\max}} \left(I(x_i, y) - \mu_{x_i, y} \right)^2 \right\} \qquad (4\text{-}14)$$

$$Var_L = Var_{l,J_a} + Var_{l,(J_a+1)} + ... + Var_{l,(J_b-1)} + Var_{l,J_b}$$

$$Var_{UL} = Var_U + Var_L \qquad (4\text{-}15)$$

In (4-13), $Var_{u,j}$ is the maximum variances of the vertical line segments between the horizontal axis of this area's center and the uppermost of ROI, calculating over a range from the minimal x-coordinate to the maximal x-coordinate of the partial area of the mass on image slice j. Var_U sums $Var_{u,j}$ over the range from image slice J_a to image slice J_b in which the corresponding mass exists. In (4-14), Var_L is difined similarily regarding the opposite half of ROI.

29
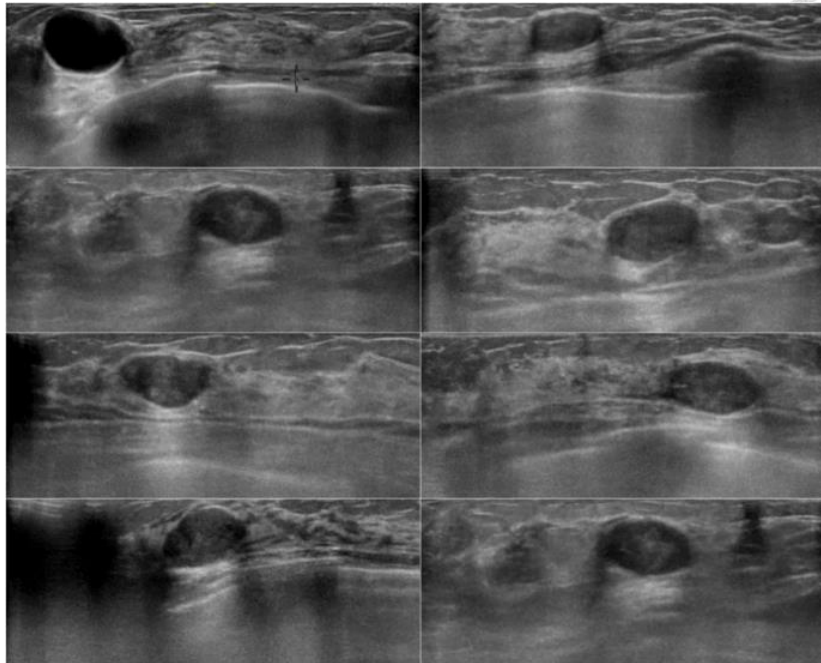
Fig. 4-5. Rim effect. While tumor grows, the tissues surrounding the tumor are compressed. Accordingly, tumor is generally accompanied with a thin rim of hyperechogenic tissues on ABUS images.

4.6 2D filter/ 3D filter

To further reduce the FPs from 2D candidates and 3D masses, a classifier was constructed by using an intuitive and efficient method applying stage-wise classification onto 2D features and 3D features.

The procedure began with all 2D candidates/3D masses involved to distinguish tumors from non-tumors through the classifier. The optimal cut-point for each feature resulting in best reduction ability was first automatically determined from all 2D candidates/3D masses. For 2D filter, the best reduction ability was defined by least false positives at 100% sensitivity. For 3D filter, the best reduction ability was defined by two criteria, the least false positives at 100% sensitivity, and the closest-to-(0, 1) criterion as in (4-16) [59]. If no cut-point with best reduction ability could be determined, then the corresponding feature was removed. After the optimal cut-points were chosen, the feature

with best performance was automatically determined by the least false positives among all optimal cut-points. Then the feature with best performance was used to classify 2D candidates/3D masses into tumors and non-tumors. To reduce the processing time, 2D candidates/3D masses which were classified as non-tumors were discarded directly at each iteration, and this feature was no longer used thereafter. The next iteration involved the remaining 2D candidates/3D masses from previous iteration, automatically determined the feature with best performance among unused features, and classified remaining 2D candidates/3D masses. The iteration of classification stopped when all features have been used.

The determination of optimal cut-point, the choice of feature with best performance, and classification were all fully automatic, without parameters or human intervention.

Algorithm for classifier

A. Initialization:

Step 1. Collected all feature values from all 2D candidates/3D masses as an input dataset S

Step 2. Statistical normalization of S

B. Iteration t:

Step 1. For each feature i, optimal cut-points C_i was determined by searching throughout all 2D candidate/3D mass j satisfying the best reduction ability at 100% sensitivity:

    i. In 2D filter: $X_i = \min_{j} \{FP_j\}$

    ii. In 3D filter:

$$Y_{i,j} = (1 - \frac{TP_j}{TP_j + FN_j})^2 + (1 - \frac{TN_j}{TN_j + FP_j})^2, \quad Z_{i,j} = FP_j \quad\quad (4\text{-}16)$$

31

The best reduction ability for each feature was recursively determined. We compared the $Y_{i,j}$ and $Z_{i,j}$ values of mass j to those of the current best. If both $Y_{i,j}$ and $Z_{i,j}$ values were smaller than those of the current best, then mass j substituted the current best; otherwise the current best retained. The procedure ran through all masses and determined the one with the best reduction ability.

       iii. If no cut-point with best reduction ability could be determined, then the corresponding feature i was removed.

False positive (FP) showed the number of 2D candidate/3D mass which were classified as tumors but were not actually located at tumor position. True positive (TP) showed the number of 2D candidate/3D mass which were classified as tumor and were correctly located at tumor position. True negative (TN) showed the number of 2D candidate/3D mass which were classified as non-tumor and were correctly located at non-tumor position.

Step 2. Optimal feature f with the best performance from all of the C_i:

    i. In 2D filter/3D filter: $\min_i \{FP(C_i)\}$

Step 3. Classification by f with value $V_{f,C}$ from C_f

μ_1 : The average value of feature f from 2D candidates/3D masses labeled as tumor

μ_0 : The average value of feature f from 2D candidates/3D masses labeled as non-tumor

$V_{f,j}$: The value of feature f from 2D candidates/3D masses j

$V_{f,C}$: The value of feature f from cut-point C_f

    i. In case of $\mu_1 < \mu_0$, 2D candidate/3D mass j would be classified as tumor if $V_{f,j} < V_{f,C}$. If $V_{f,j} \geq V_{f,C}$, then 2D candidate/3D mass j would

32

be discarded.

ii. In case of $\mu_1 \geq \mu_0$, 2D candidate/3D mass j would be classified as

tumor if $V_{f,j} \geq V_{f,C}$. If $V_{f,j} < V_{f,C}$, then 2D candidate/3D mass j would

be discarded.

Step 4. f was no longer used thereafter, and then iteration t proceeded to $t+1$

C. Stop criteria: All features have been used.
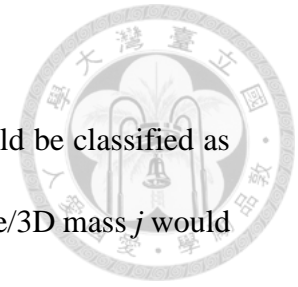
Chapter 5　Applications and Results
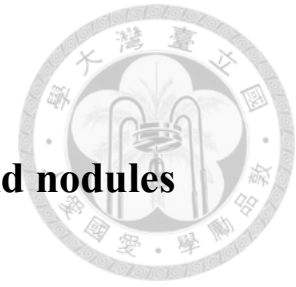
5.1 Boundary segmentation for ultrasonic thyroid nodules

5.1.1　Materials

The ultrasound images of thyroid nodules used in this study were recruited from 28 August 2007 to 14 January 2013 in the National Taiwan University Hospital. Ultrasound examinations were performed by using Philips HDI 5000 scanner (Philips Healthcare, Bothell, WA, USA; manufactured in 2000) with a 5-12 MHz linear multi-frequency probe (L12-5). The process of scanning was performed with the patients in supine position and the neck hyperextended. All informed consents were obtained from all patients recruited for this study and approved by our institutional review board.

The 538 ultrasound images of thyroid nodules including 322 benign and 216 malignant cases, proven by cytology or histology examinations, were used for performance validation and comparison. The details were shown in Table 5-1. The manual delineation of the nodule boundary by experienced radiologist was used to serve as a gold-standard.

	Histopathologic diagnosis			
	Benign (n=322, 60%)		Malignant (n=216, 40%)	
Characteristics	Number			
Gender				
Female	256		177	
Male	66		39	
Age				
Female	48.95		46.98	
Male	52.31		47.01	
Location				
Right	174		108	
Intermediate	9		5	
Left	139		103	
Subtypes	Nodular hyperplasia	268	Papillary carcinoma	193
	Follicular adenoma	54	Follicular carcinoma	15
			Medullary carcinoma	6
			Anaplastic carcinoma	2

Table 5-1. Patient clinical data and nodule characteristics.

5.1.2 Performance analysis and evaluation

To evaluate the performance of our proposed method, we divided our image database into two datasets with approximately equal size by the order of inspection date. One of them was used in training of the optimal parameters, and the other one was used as testing dataset to evaluate the effectiveness of our proposed method.

We analyzed the difference between the boundary automatically generated by our proposed method (automatic boundary, AB) and the gold-standard boundary delineated by experienced radiologist (manual boundary, MB). Since the evaluation considered two major objectives, similarity and accuracy, the boundary error metrics and the overlapping area metrics were used to evaluate the effectiveness of our proposed method [14].

35

Two boundary error metrics included the Hausdorff distance (HD) and the mean absolute distance (MD) by using the set of minimum distance errors (MDE). MDE was the distance from a pixel point on AB to its nearest pixel point on MB, as shown in (5-1) and Fig. 5-1. HD and MD measured the largest MDE and the average MDE, respectively. Their definitions were shown in (5-2) and (5-3).

$$MDE_i = \min_m \left\| AB_i - MB_m \right\| \tag{5-1}$$

$$HD = \max_i \left(MDE_i \right) \tag{5-2}$$

$$MD = \frac{1}{N} \sum_i^N \left(MDE_i \right) \tag{5-3}$$

In (5-1), (5-2), and (5-3), if AB has N pixel points, and their MDE are represented as a set of $\{MDE_1, MDE_2, ..., MDE_i, ..., MDE_{N-1}, MDE_N\}$, where MDE_i is the MDE of pixel point i on the AB, and the HD and MD are the maximum and the average value from the above set collecting from N pixel points on AB, respectively. To avoid the effect of different nodule size, the corresponding normalized boundary errors metrics are derived by dividing HD and MD by the number of boundary pixel points on MB, namely normalized HD (NHD) and normalized MD (NMD).
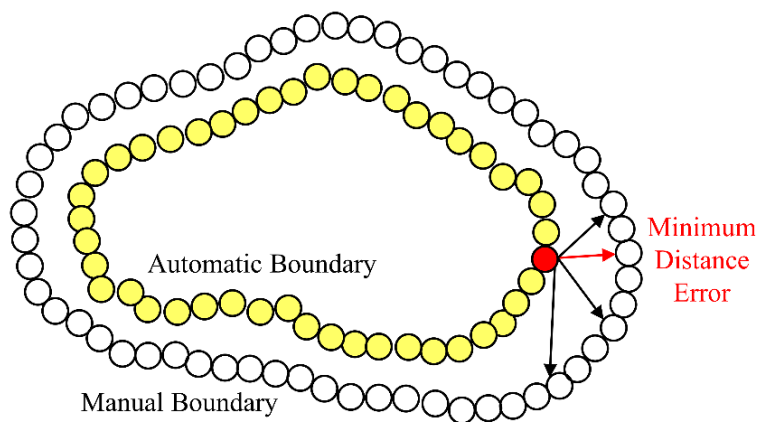
Fig. 5-1. The schematic diagram of the minimum distance error for every automatic boundary point.

Furthermore, five overlapping area metrics were used to evaluate the degree of overlapping between two areas delineated by AB and MB, including the true positive area ratio (TPR), the false positive area ratio (FPR), the false negative area ratio (FNR), the Jaccard similarity index (JSI), and the positive predictive value (PPV), as shown in (5-4), (5-5), (5-6), (5-7), (5-8) and Fig. 5-2.
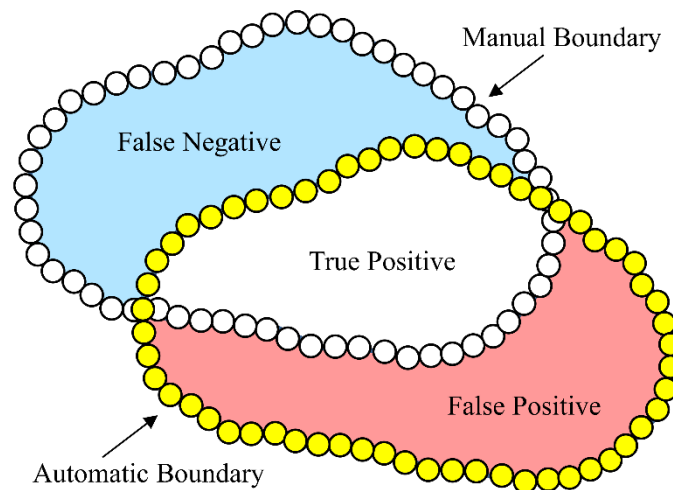
Fig. 5-2. The schematic diagram of the corresponding area of FP, TP, and FN.

The true positive area ratio (TPR) showed the ratio of number of pixel points which were designated as nodule both in AB and MB divided by the number of pixel points in MB. Higher TPR indicated that the segmentation results were more similar to the gold-standard.

$$TPR = \frac{\left| A_m \cap A_a \right|}{\left| A_m \right|} \tag{5-4}$$

A_m and A_a are the sets of pixel points within the delineated areas by MB and AB, respectively.

The false positive area ratio (FPR) showed the ratio of number of pixel points which were designated as nodule in AB but not in MB divided by the number of pixel points in MB. Lower FPR indicated less incorrect segmentations.

37

$$FPR = \frac{\left| A_m \cup A_a - A_m \right|}{\left| A_m \right|}$$ (5-5)

The false negative area ratio (FNR) showed the ratio of number of pixel points in MB which were not detected by AB divided by the number of pixel points in MB.

$$FNR = \frac{\left| A_m \cup A_a - A_a \right|}{\left| A_m \right|}$$ (5-6)

The Jaccard similarity index (JSI) showed the overlapping percentage. Higher JSI represented higher degree of overlapping between two areas.

$$JSI = \frac{\left| A_m \cap A_a \right|}{\left| A_m \cup A_a \right|}$$ (5-7)

The positive predictive value (PPV) showed the predictive ability of our proposed method to detect the nodule boundary.

$$PPV = \frac{\left| A_m \cap A_a \right|}{\left| A_a \right|}$$ (5-8)

The comparison between our proposed method and previously published works was accomplished to objectively evaluate the performance. Imitating the comparison in work by Chang *et al.* [44], we compared with Active Contour Model (ACM) [42] and Watershed Model (WM) [43], and further add Distance Regularized Level Set Evolution (DRLSE) [45] into comparison. For the sake of standardization, ACM and WM were reproduced with OpenCV library [44], and DRLSE is realized with source code provided by author(s) [46]. During the evaluation process, all methods used identical training dataset to train the optimal parameters and then tested their respective performance with identical testing dataset. Without image preprocessing, each case implemented all comparison methods with identical input principle. In accordance to the requirement for different algorithms, the initial input originated from the identical four nodes was

38

prepared in appropriate form, and was accompanied with identical ROI which was automatically generated based on the identical four nodes.
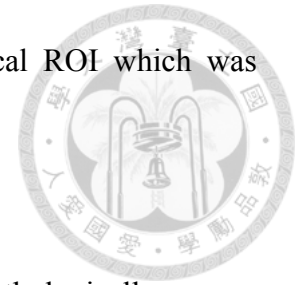
5.1.3　Results

The database of ultrasound images consisting a total of 538 pathologically proven thyroid nodules, including 322 benign and 216 malignant cases were used to evaluate the performance of our proposed method. Among them, the training and testing dataset had 275 and 263 cases, respectively. The effectiveness of the performance results was then reported in qualitative and quantitative terms.

Qualitative analysis

A step-by-step visualization showed the results of our proposed method actually applied to the ultrasound images corresponding to the above schematic diagram (Fig. 3-2). So that, two clinical thyroid nodules, one benign and one malignant, were used to demonstrate in Fig. 5-3 and Fig. 5-4, respectively. To further show comparison between original images, AB, and MB, aforementioned two clinical thyroid nodules were used to demonstrate the final results of AB in the third row of Fig. 5-5, and showed that in both case AB had a very similar results to their corresponding MB (Fig. 5-5, second row).
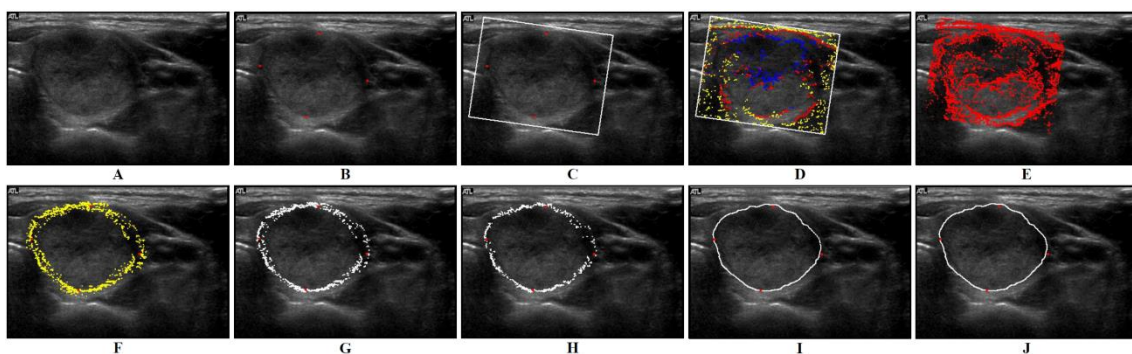
Fig. 5-3. Result of applying our proposed method on a benign case of thyroid nodule corresponding to the steps of Fig. 3-2. (A) original thyroid nodule on ultrasound image. (B) manually input four extreme nodes on approximate location of the nodule. (C) the

region of interest (ROI) was automatically generated based on the major axis and minor axis. (D) reference boundary points. (E) new cut points and new radial lines defined by the centers of the eight 45-degrees sectors in the ROI. (F) direction searching method. (G) outlier elimination method. (H) inner product method. (I) smoothing. (J) linking.
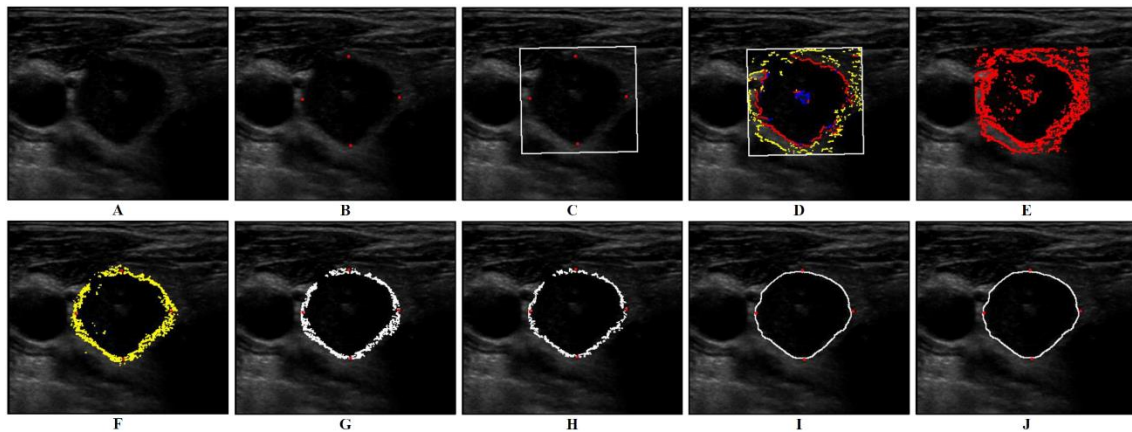
Fig. 5-4. Result of applying our proposed method on a malignant case of thyroid nodule corresponding to the steps of Fig. 3-2. (A) original thyroid nodule on ultrasound image. (B) manually input four extreme nodes on approximate location of the nodule. (C) the region of interest (ROI) was automatically generated based on the major axis and minor axis. (D) reference boundary points. (E) new cut points and new radial lines defined by the centers of the eight 45-degrees sectors in the ROI. (F) direction searching method. (G) outlier elimination method. (H) inner product method. (I) smoothing. (J) linking.
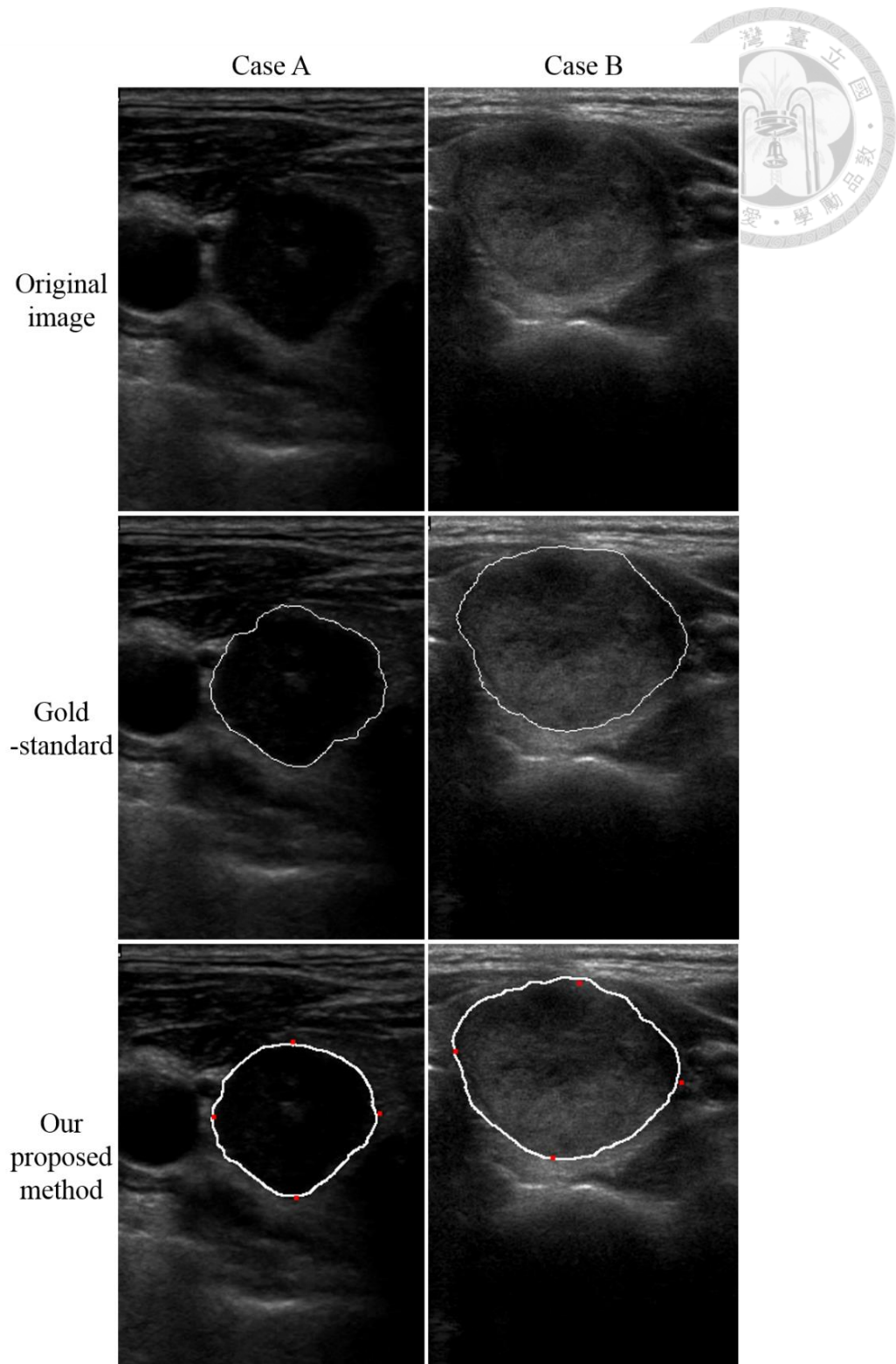
Fig. 5-5. The comparison between the boundary generated by our proposed method and the gold-standard boundary by experienced radiologist. The first row were original images, the second row were the corresponding gold-standard boundaries, and the third row were the corresponding automatic boundaries, respectively.

doi:10.6342/NTU202003153

To show that our proposed method is applicable on complicated cases with variation of the nodule's inner tissues, Fig. 5-6 demonstrated that AB agreed well with corresponding MB in the case of weak boundary (Fig. 5-6, first row), blurring boundary (Fig. 5-6, second row), missing boundary (Fig. 5-6, third row), inhomogeneity (Fig. 5-6, fourth row), and cysts in nodule (Fig. 5-6, fifth row), respectively.

Fig. 5-7 demonstrated the comparison between gold-standard (Fig. 5-7, second row), our proposed method (Fig. 5-7, third row), WM (Fig. 5-7, fourth row), ACM (Fig. 5-7, fifth row), and DRLSE (Fig. 5-7, sixth row) applied to one benign nodule (Case A), one malignant nodule (Case B), and another malignant nodule with inhomogeneity (Case C), respectively.
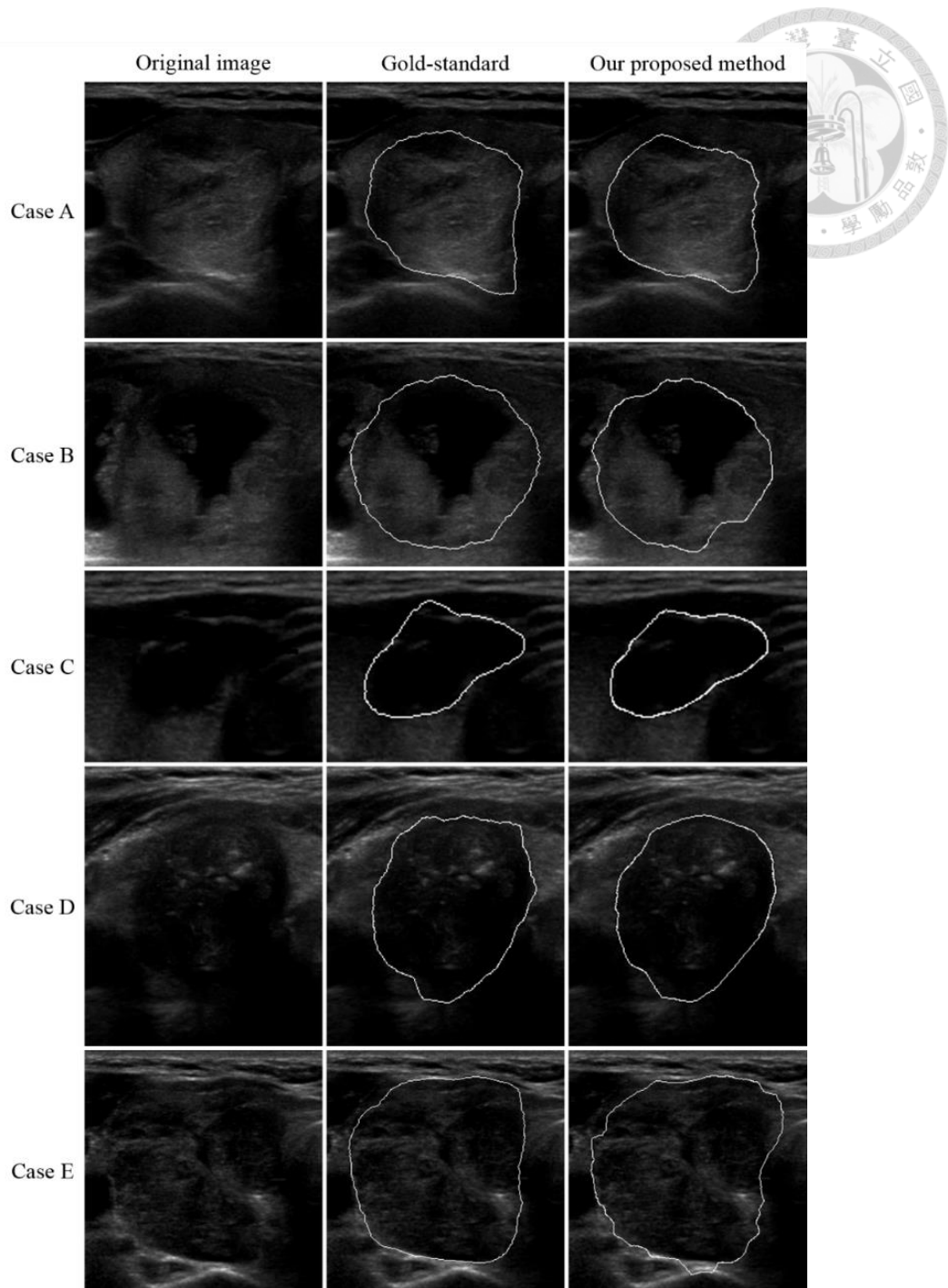
Fig. 5-6. The comparison between the boundary generated by our proposed method and the gold-standard boundary by experienced radiologist on complicated cases. The first row (Case A) was a nodule with weak boundary. The second row (Case B) was a nodule with blurring boundary. The third row (Case C) was a nodule with missing boundary. The fourth row (Case D) showed a nodule with intensity inhomogeneity. The fifth row (Case E) showed a case with cysts in nodule.
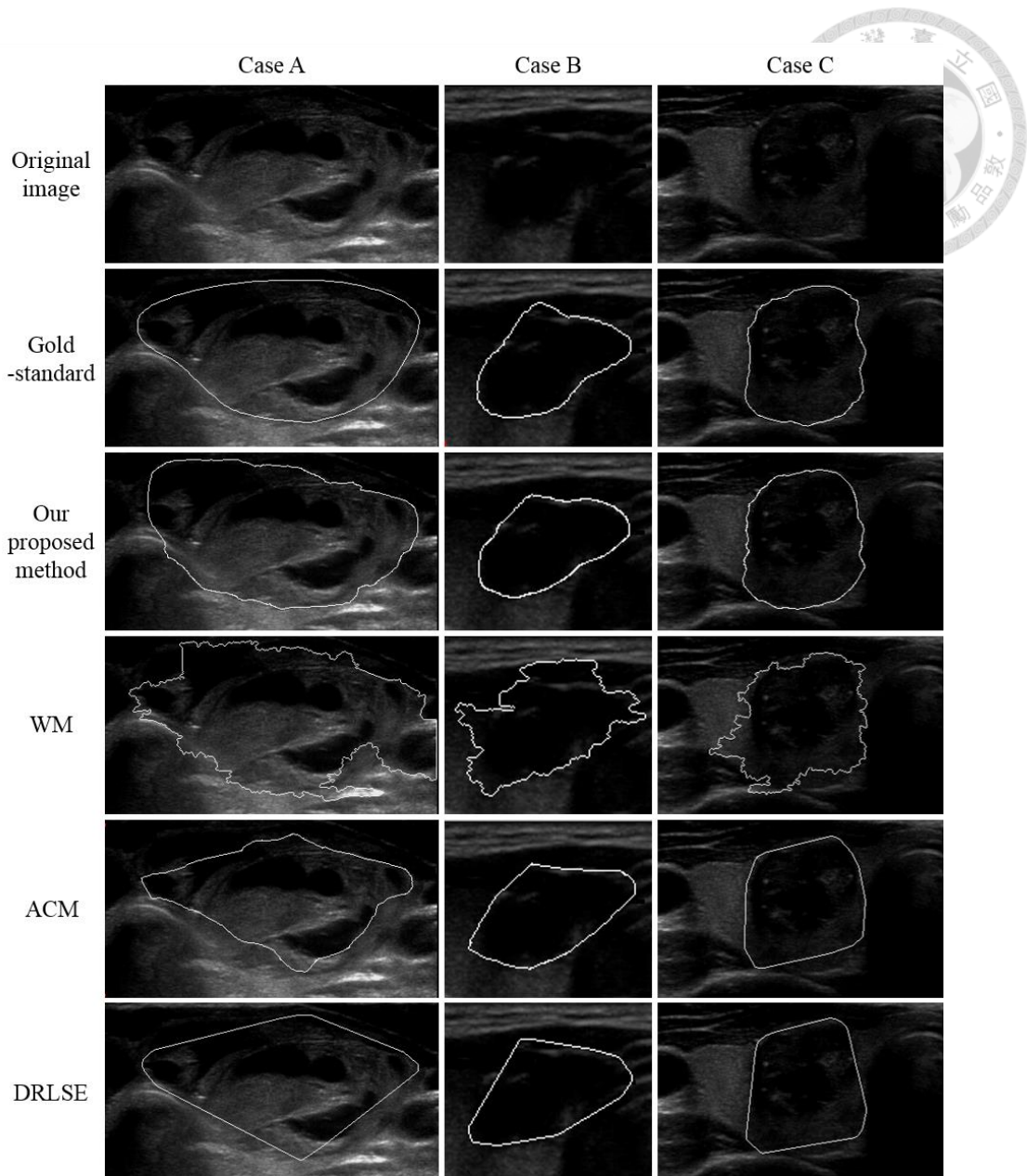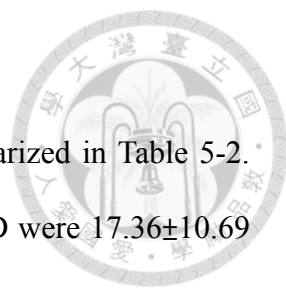
Fig. 5-7. The screenshots of identical view of one benign nodule (Case A), one malignant nodule (Case B), and another malignant nodule with inhomogeneity (Case C), showing the results of comparison between our proposed method and other standardized methods. The first row were the original images, the second row were the gold-standard boundaries delineated by experienced radiologist, the third row were automatic boundaries by our proposed method, the fourth row were boundaries by Watershed Model, the fifth row were boundaries by Active Contour Model, the sixth row were boundaries by Distance Regularized Level Set Evolution, respectively.

Quantitative analysis

The performance results from 538 thyroid nodules were summarized in Table 5-2. Among them, the testing dataset (263/538) showed that HD and MD were 17.36±10.69 pixel points and 4.69±3.00 pixel points, respectively. The NHD and NMD were 3.65%± 1.15% and 1.02%±0.40%, respectively. The above results clearly indicated a high degree of correlation between AB and MB. The five overlapping area metrics showed the similarity and accuracy between two areas delineated by AB and MB. The TPR achieved 93.66%±5.07%, whereas FPR was 7.68%±5.76%. These results demonstrated that our proposed method delineated thyroid nodules with high sensitivity and high specificity.

To investigate the generalizing ability of our proposed method, we analyzed the performance of the proposed method in both the training dataset (275/538) and the testing dataset (263/538) with Mann-Whitney U test. The testing performance was similar to the training performance, as shown by *p*-values in most metrics (Table 5-2). It indicated that our proposed method was not over-fitted with the training dataset and could be applied generally to the testing dataset.

To see if our proposed method biased toward benign or malignant nodules, we analyzed the respective performances in 161 benign modules and 102 malignant nodules from testing dataset (263/538). The results were listed in Table 5-3. In both benign and malignant nodules our proposed method achieved good results, and there was no statistical difference between them shown by Mann-Whitney U test.

We compared our proposed method with WM, ACM, and DRLSE using the same testing dataset. We used the packaged functions directly provided by the original authors for analysis of non-preprocessed images in the comparison study. The respective performances of these method were listed in Table 5-4. Then we analyzed the difference

45

in performance by comparing our proposed method to WM, our proposed method to ACM, our proposed method to DRLSE with Mann-Whitney U test. The *p*-values were summarized in Table 5-5.

In order to evaluate if our proposed method was insensitive to adopted parameters, we analyzed the effect of different parameter values on performance. The results of sensitivity analysis regarding tunable parameters "distance w", "angle $\theta_{Threshold}$", "parameter a", and "inner product b" were summarized in Table 5-6, Table 5-7, Table 5-8, and Table 5-9, respectively. It showed that the performance varies with different values of w and $\theta_{Threshold}$. The larger w we chose, the worse result we got. $\theta_{Threshold}$ affected the performance in opposite direction: the smaller $\theta_{Threshold}$, the worse result. On the contrary, our proposed method was insensitive to a and b. All performance metrics were almost the same regardless of a and b values.

All performance experiments were accomplished on an Intel Core i7-4500U 1.80 GHz processor with 8 GB RAM. The time performance with the same dataset of our proposed method was 3.30 seconds per nodule, whereas WM and ACM were 5.68 and 0.79 seconds per nodule, respectively. DRLSE had the worst time performance with 16.01 seconds per nodule.

Cases	Dataset	Boundary Error Metrics			
		HD	MD	NHD (%)	NMD (%)
538	Training (275/538)	18.07±10.31	4.30±2.62	3.23±1.12	0.78±0.29
	Testing (263/538)	17.36±10.69	4.69±3.00	3.65±1.15	1.02±0.40
	p-value	0.267	0.184	<0.005	<0.005

Cases	Dataset	Overlapping Area Metrics				
		TPR (%)	FPR (%)	FNR (%)	JSI (%)	PPV (%)
538	Training (275/538)	94.22±4.38	7.84±5.39	5.78±4.38	87.47±3.97	92.66±4.49
	Testing (263/538)	93.66±5.07	7.68±5.76	6.34±5.07	87.08±4.44	92.82±4.75
	p-value	0.337	0.542	0.337	0.575	0.555

Table 5-2. The individual performance of our proposed method from training and testing datasets by using boundary error metrics and overlapping area metrics. (HD = Hausdorff distance, MD = mean absolute distance, NHD = normalized HD, NMD = normalized MD, TPR = true positive area ratio, FPR = false positive area ratio, FNR = false negative area ratio, JSI = Jaccard similarity index, PPV = positive predictive value. Unit for HD and MD is pixel. All values are represented as average±standard deviation. p-values are obtained with Mann-Whitney U test.)

Dataset	Data Type	NHD (%)	NMD (%)	TPR (%)	FPR (%)	FNR (%)	JSI (%)	PPV (%)
Testing (263/538)	Benign (161/263)	3.59±1.08	0.99±0.37	93.98±4.93	7.90±5.52	6.02±4.93	87.19±4.14	92.62±4.59
	Malignant (102/263)	3.75±1.24	1.06±0.45	93.15±5.25	7.34±6.11	6.85±5.25	86.91±4.86	93.13±4.98
	p-value	0.390	0.337	0.116	0.250	0.116	0.935	0.254

Table 5-3. The individual performance of our proposed method from benign and malignant cases by using boundary error metrics and overlapping area metrics. (HD = Hausdorff distance, MD = mean absolute distance, NHD = normalized HD, NMD = normalized MD, TPR = true positive area ratio, FPR = false positive area ratio, FNR = false negative area ratio, JSI = Jaccard similarity index, PPV = positive predictive value.

47

All values are represented as average±standard deviation. p-values are obtained with

Mann-Whitney U test.)
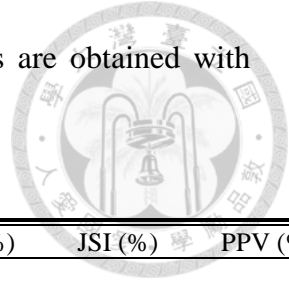
Dataset	Method	NHD (%)	NMD (%)	TPR (%)	FPR (%)	FNR (%)	JSI (%)	PPV (%)
Testing (263/538)	Our proposed method	3.65±1.15	1.02±0.40	93.66±5.07	7.68±5.76	6.34±5.07	87.08±4.44	92.82±4.75
	Watershed model	6.22±1.87	1.73±0.42	94.93±3.00	17.50±7.16	5.07±3.00	81.02±4.24	84.82±5.07
	Active contour model	4.60±1.69	1.60±0.67	86.93±7.35	8.31±11.15	13.07±7.35	80.56±5.84	92.62±8.10
	DRLSE	4.22±1.03	1.47±0.36	87.61±6.31	7.75±7.34	12.39±6.31	81.37±4.02	92.61±5.85

Table 5-4. The comparison of our proposed method with other methods by using the same dataset. (HD = Hausdorff distance, MD = mean absolute distance, NHD = normalized HD, NMD = normalized MD, TPR = true positive area ratio, FPR = false positive area ratio, FNR = false negative area ratio, JSI = Jaccard similarity index, PPV = positive predictive value. All values are represented as average±standard deviation.)

Dataset	Method	NHD	NMD	TPR	FPR	FNR	JSI	PPV
Testing (263/538)	Our proposed method v.s. Watershed model	<0.005	<0.005	0.051	<0.005	0.050	<0.005	<0.005
	Our proposed method v.s. Active contour model	<0.005	<0.005	<0.005	<0.005	<0.005	<0.005	<0.005
	Our proposed method v.s. DRLSE	<0.005	<0.005	<0.005	0.174	<0.005	<0.005	0.529

Table 5-5. Mann-Whitney U test *p*-values for method comparisons. (HD = Hausdorff distance, MD = mean absolute distance, NHD = normalized HD, NMD = normalized MD, TPR = true positive area ratio, FPR = false positive area ratio, FNR = false negative area ratio, JSI = Jaccard similarity index, PPV = positive predictive value. All values are represented as average±standard deviation.)

Dataset	w	NHD (%)	NMD (%)	TPR (%)	FPR (%)	FNR (%)	JSI (%)	PPV (%)
Testing (263/538)	2	3.65±1.15	1.02±0.40	93.66±5.07	7.68±5.76	6.34±5.07	87.08±4.44	92.82±4.75
	3	4.59±1.43	1.34±0.57	88.83±7.01	5.20±5.00	11.17±7.01	84.43±5.45	94.88±4.34
	4	5.53±1.64	1.78±0.75	83.60±8.83	3.86±4.78	16.40±8.83	80.43±7.04	96.08±4.22
	5	6.26±1.75	2.17±0.93	79.10±10.28	2.87±3.89	20.90±10.28	76.78±8.61	96.94±3.59

Table 5-6. Performance of different values of distance w. (HD = Hausdorff distance, MD = mean absolute distance, NHD = normalized HD, NMD = normalized MD, TPR = true positive area ratio, FPR = false positive area ratio, FNR = false negative area ratio, JSI = Jaccard similarity index, PPV = positive predictive value. All values are represented as average±standard deviation.)

Dataset	$\theta_{Threshold}$	NHD (%)	NMD (%)	TPR (%)	FPR (%)	FNR (%)	JSI (%)	PPV (%)
Testing (263/538)	45	5.01±1.08	1.92±0.59	76.51±6.40	0.86±1.62	23.49±6.40	75.83±5.84	98.99±1.74
	90	4.71±1.23	1.59±0.61	81.73±7.06	1.70±2.56	18.27±7.06	80.32±6.07	98.16±2.53
	135	4.04±1.32	1.19±0.51	88.54±6.37	3.75±3.92	11.46±6.37	85.34±5.38	96.21±3.64
	180	3.65±1.15	1.02±0.40	93.66±5.07	7.68±5.76	6.34±5.07	87.08±4.44	92.82±4.75

Table 5-7. Performance of different values of angle $\theta_{Threshold}$. (HD = Hausdorff distance, MD = mean absolute distance, NHD = normalized HD, NMD = normalized MD, TPR = true positive area ratio, FPR = false positive area ratio, FNR = false negative area ratio, JSI = Jaccard similarity index, PPV = positive predictive value. All values are represented as average±standard deviation.)

Dataset	a	NHD (%)	NMD (%)	TPR (%)	FPR (%)	FNR (%)	JSI (%)	PPV (%)
Testing (263/538)	1	3.65±1.15	1.02±0.40	93.66±5.07	7.68±5.76	6.34±5.07	87.08±4.44	92.82±4.75
	2	3.94±1.23	1.05±0.39	94.39±4.76	8.47±5.93	5.61±4.76	87.14±4.24	92.17±4.80
	3	3.96±1.22	1.05±0.39	94.44±4.74	8.50±5.98	5.56±4.74	87.16±4.24	92.15±4.83
	4	3.96±1.22	1.05±0.39	94.44±4.74	8.49±5.99	5.56±4.74	87.17±4.25	92.16±4.84
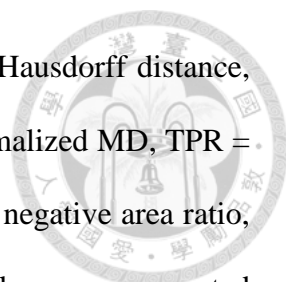
Table 5-8. Performance of different values of parameter *a*. (HD = Hausdorff distance, MD = mean absolute distance, NHD = normalized HD, NMD = normalized MD, TPR = true positive area ratio, FPR = false positive area ratio, FNR = false negative area ratio, JSI = Jaccard similarity index, PPV = positive predictive value. All values are represented as average±standard deviation.)

Dataset	b	NHD (%)	NMD (%)	TPR (%)	FPR (%)	FNR (%)	JSI (%)	PPV (%)
Testing (263/538)	6	3.64±1.14	1.02±0.40	93.50±5.12	7.52±5.70	6.50±5.12	87.06±4.47	92.95±4.72
	8	3.65±1.14	1.02±0.40	93.60±5.08	7.63±5.75	6.40±5.08	87.07±4.45	92.86±4.75
	10	3.65±1.15	1.02±0.40	93.66±5.07	7.68±5.76	6.34±5.07	87.08±4.44	92.82±4.75
	12	3.66±1.14	1.02±0.40	93.69±5.06	7.72±5.79	6.31±5.06	87.08±4.43	92.78±4.77
	14	3.66±1.14	1.02±0.40	93.72±5.05	7.77±5.81	6.28±5.05	87.07±4.43	92.75±4.78
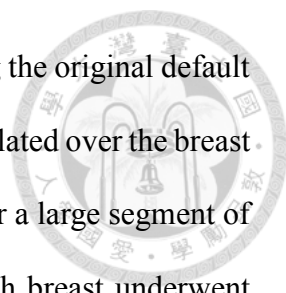
Table 5-9. Performance of different values of inner product *b*. (HD = Hausdorff distance, MD = mean absolute distance, NHD = normalized HD, NMD = normalized MD, TPR = true positive area ratio, FPR = false positive area ratio, FNR = false negative area ratio, JSI = Jaccard similarity index, PPV = positive predictive value. All values are represented as average±standard deviation.)

5.2 Breast tumor detection in 3D ultrasound imaging

5.2.1 Materials

Automated 3D Breast Ultrasound Images

The ABUS images used in this study were acquired by an ACUSON-S2000 Automated Breast Volume Scanner (ABVS) (Siemens Medical Solutions USA, Inc., Malvern, PA) in Breast Center of National Taiwan University Hospital. The scanner consisted of a flexible arm with a broadband linear array transducer (5-14 MHz with the center frequency at 11 MHz) at its end, a touchscreen, and a 3D imaging workstation. The

50
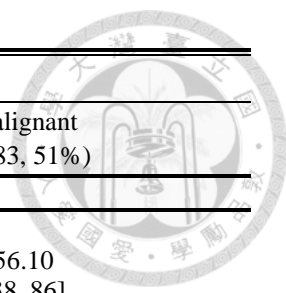
scan process was performed with the patients in supine position using the original default settings of the Siemens ACUSON-S2000 ABVS. The transducer translated over the breast with a constant speed to obtain a 3D volume of imaging data to cover a large segment of the breast. The medial, anterior-posterior, and lateral regions of each breast underwent standardized examinations to form a total six sets of ABUS image passes of entire breasts. Each examination consisted of 318 2D transverse plane images with 0.525mm slice-to-slice interval. The sagittal plane images and coronal plane images were then reconstructed by ABVS Workplace system to form a 3D volume. However, the number of two reconstructed plane image slices varied for different patients.

Patients

Informed consents were obtained from all patients recruited in this study and approved by National Taiwan University Hospital Research Ethics Committee.

For performance validation, consisting of 176 passes with 132 abnormal passes and 44 normal passes from women were used in this study. Among the 132 abnormal passes, there were 162 tumors all proven by the histopathological examination of biopsy specimens, including 79 benign and 83 malignant ones. The details of the tumor characteristics were shown in Table 5-10. In pursuit of reality and standardization, the passes of transverse plane images in which visible tumor existed were used for analysis. All tumors were annotated by experienced radiologist to serve as a gold-standard. For generalization of our proposed system, 44 normal passes were also included in the database. All the normal passes were confirmed by multiple examination modalities.

	Histopathologic diagnosis			
	Benign (n=79, 49%)		Malignant (n=83, 51%)	
Characteristics	Number			
Age (years)				
Average	43.00		56.10	
Range	[20, 77]		[38, 86]	
$\mu \pm \sigma$	43.03±13.35		56.10±10.66	
Size (cm)				
Average	1.56		2.46	
Range	[0.11, 11.48]		[0.40, 17.82]	
$\mu \pm \sigma$	1.56±1.43		2.46±2.04	
BIRADS distribution	Grade 2	2	Grade 2	0
	Grade 3	39	Grade 3	4
	Grade 4A	26	Grade 4A	10
	Grade 4B	9	Grade 4B	13
	Grade 4C	2	Grade 4C	24
	Grade 5	1	Grade 5	32
Subtypes	Fibroadenoma	55	Invasive ductal carcinomas	67
	Fibrocystic disease	20	Ductal carcinomas in situ	9
	Phyllodes tumor	2	Invasive lobular carcinomas	5
	Intraductal papilloma	2	Malignant phyllodes tumor	2

Table 5-10. Patient clinical data and tumor characteristics. (Size was the longest length among the tumors.)

5.2.2 Performance Analysis and Evaluation

The detection results was compared to gold-standard indicated by experienced radiologist. If a mass classified as tumor overlapped with gold-standard tumor location, then it was considered as true positive, otherwise it was considered as false positive [37]. To validate the performance of our proposed CADe system, we randomly divided our database into two sets, set A and set B, with approximately equal size and approximately same proportions of benign, malignant and normal cases. At first, we used set A for training and set B for testing to evaluate the effectiveness of our proposed method. Then we swapped the two sets and repeated the same validation process for obtaining credible validation results. The performance of proposed CADe system was presented in terms of sensitivity and false positive per pass, since the conventional sensitivity-specificity presentation is not suitable for detecting lesions in medical images that may contain

52

multiple abnormalities per image [60]. In addition, this presentation was also used by related works [33]-[39] in the field of detection on ABUS images and therefore allowed cross-literature comparison.
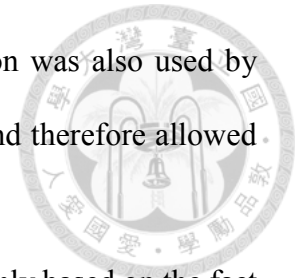
Furthermore, the performance evaluation of this system was mainly based on the fact that all tumors ought to be detected correctly without missing. With 100% sensitivity, the less detected errors (FPs), the better performance.

All procedures in this system were fully automatic without human intervention and accomplished on a Windows-based PC with Intel Core i7-6700 3.4 GHz processor with 32GB RAM.

5.2.3 Results

The ABUS images database consisting a total of 176 passes, including 132 abnormal passes and 44 normal passes was used to evaluate the performance of our CADe system. Among the 132 abnormal passes, there were 162 pathologically proven tumors, in which 79 were benign and 83 were malignant. The effectiveness of performance was shown in two aspects, qualitative analysis and quantitative analysis.

Qualitative analysis

To illustrate the detection result and the reduction of false positives, we visualized our method applied to ABUS images step-by-step, 2D merge, 2D filter, 3D merge, and 3D filter. The left column and right column of Fig. 5-8 demonstrated the results in a benign and a malignant tumor, respectively. The tumors were correctly detected without forming false positives.

Fig. 5-9 showed the examples of false positive in two conditions, false positive coexisting with true positive, and false positive existing alone in the normal pass, as demonstrated in the left and right columns, respectively.
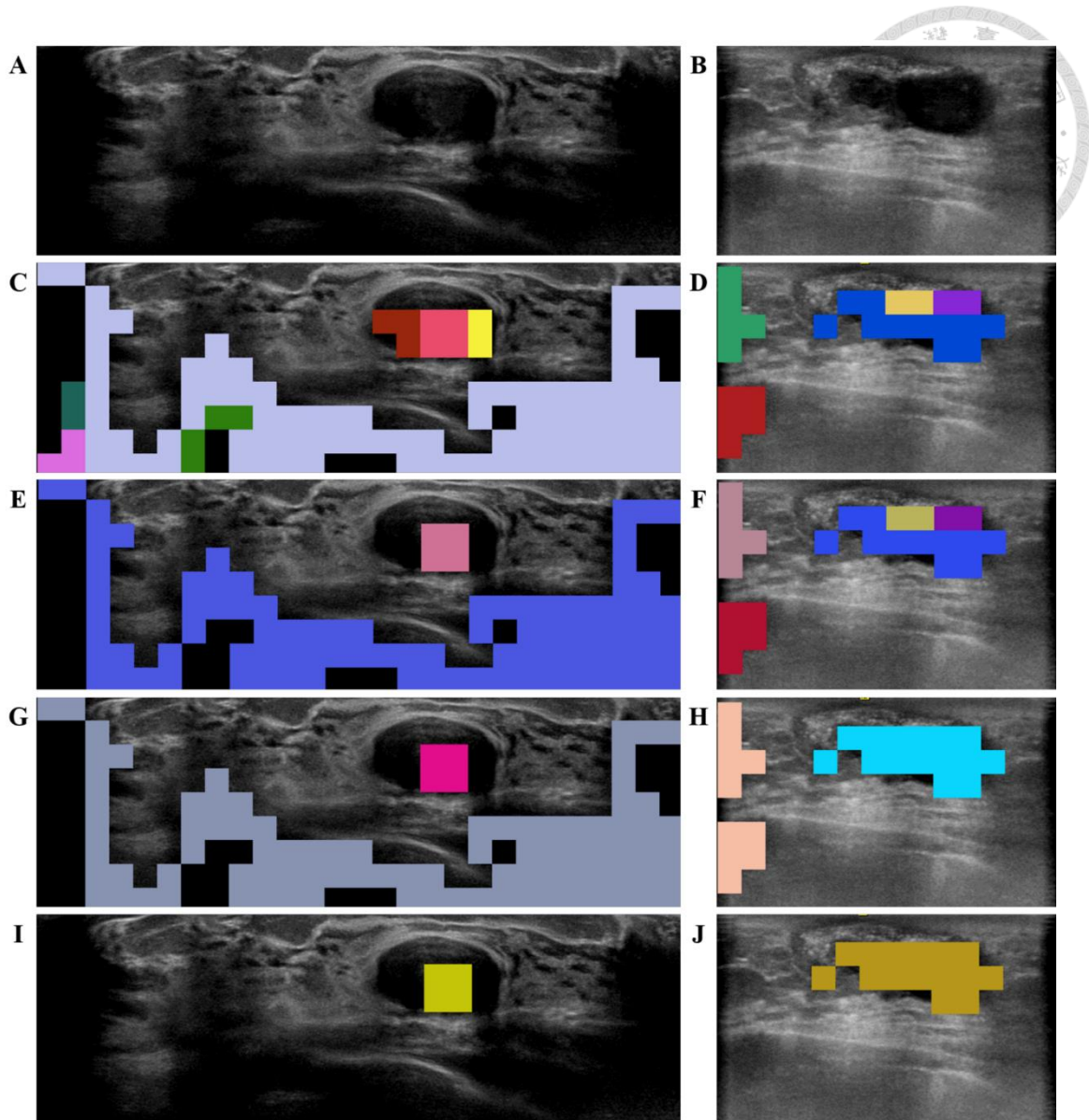
Fig. 5-8. The detection results of applying our method on a benign (left column) and malignant tumor (right column) corresponding to the steps, including 2D merge, 2D filter, 3D merge, and 3D filter. Each individual 2D candidate/3D mass was marked in random color on an image slice. (A) and (B) were original ABUS images. (C) and (D) were the results of 2D merge. (E) and (F) were the results of 2D filter. (G) and (H) were the results of 3D merge. (I) and (J) were the results of 3D filter, which showed excellent results of correct tumor detection without false positive.
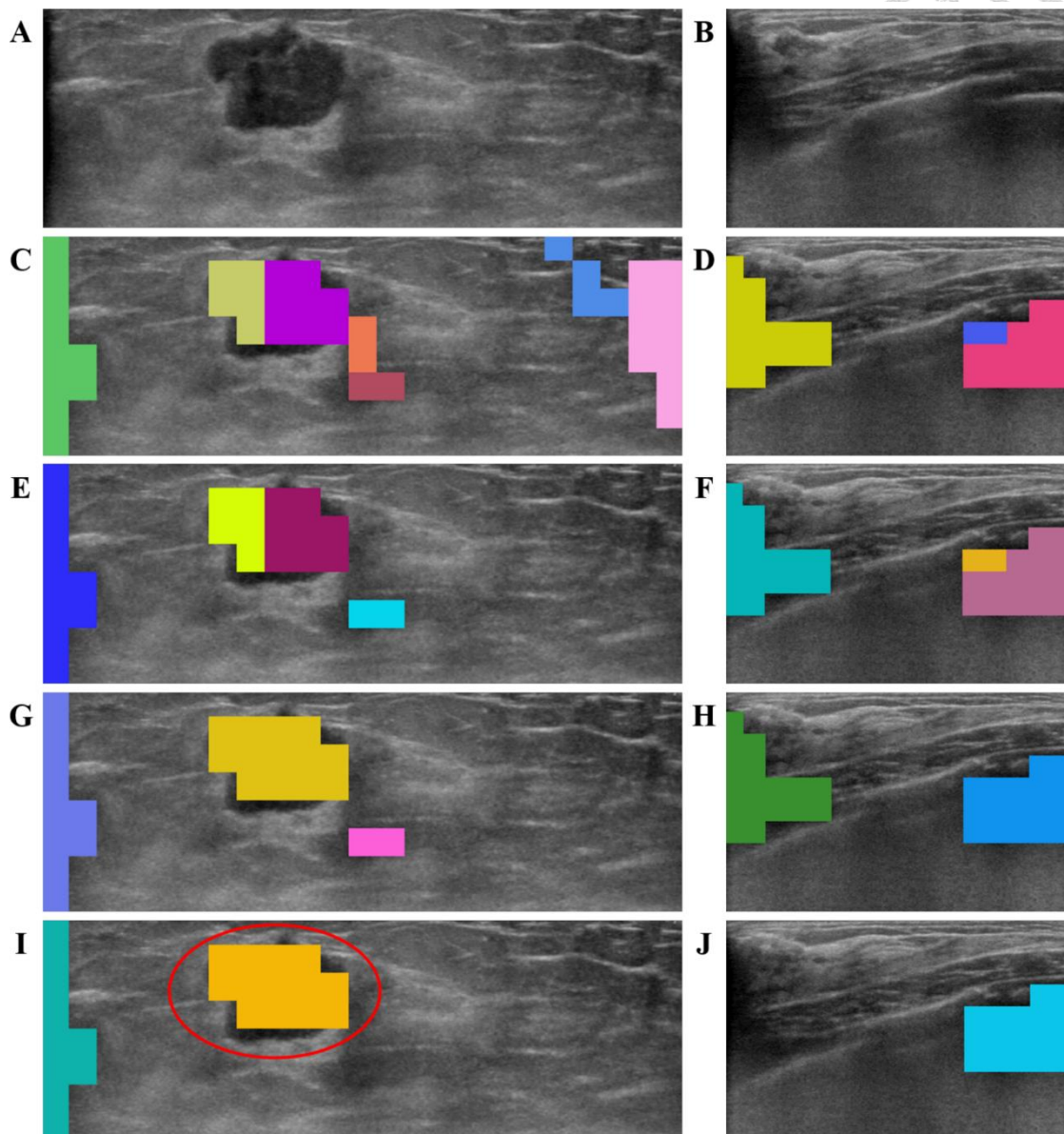
Fig. 5-9. The examples of false positives in two conditions, false positive coexisting with true positive, and false positive existing alone in the normal pass. Each 2D candidate/3D mass was marked in random color on an image slice. (A) and (B) were original ABUS images. (C) and (D) were the results of 2D merge. (E) and (F) were the results of 2D filter. (G) and (H) were the results of 3D merge. (I) and (J) showed the results of 3D filter. In (I) and (J), red circle indicated the mass was true positive and uncirculated masses were false positives.

Quantitative analysis

The overall result from four major parts of our proposed method, including 2D merge, 2D filter, 3D merge, and 3D filter was shown in Table 5-11. These result showed that our proposed method had good detection results and had a particularly strong filtering ability in 3D filter. Our proposed method achieved 100% and 90% sensitivity with 6.71 and 5.14 FPs per pass, respectively.

The performance of our proposed method in different types of cases was summarized in Table 5-11. In all conditions, our proposed method attained excellent detection results. To investigate if the performance of our proposed method was biased toward certain type of cases, we analyzed the performance with Mann-Whitney U test and Student's t-test. The results showed that the differences between normal and abnormal cases and between benign and malignant cases were both statistically insignificant, as shown in Table 5-12.

In order to evaluate if our proposed method was insensitive to adopted parameters, we analyzed the effect of different parameter values on performance. The results of sensitivity analysis regarding all four tunable parameters "square size", "percentile of grayscale value", "threshold for t-value t", and "threshold for F-value f" were summarized in Table 5-13, Table 5-14, Table 5-15, and Table 5-16, respectively. It showed that the performance was insensitive to the choice of parameters, except percentile of grayscale value. The standard deviations of FPs per pass at 100% sensitivity with different square size, t, and f were 1.75, 0.00, 0.58, respectively (first, third and fourth rows in Table 5-17). On the contrary, the choice of percentile grayscale value affected the performance, with standard deviation 5.81 FPs per pass at 100% sensitivity (second row in Table 5-17).

The processing time for four major parts of our proposed method were 98.9907,

0.0014, 0.2104, and 0.0069 seconds per pass, respectively.

For comparison with the previously published methods for detecting tumor on ABUS images, Table 5-18 listed the results reported in literatures [35], [38], [39] presented in FPs per pass at different sensitivities and the method processing time.
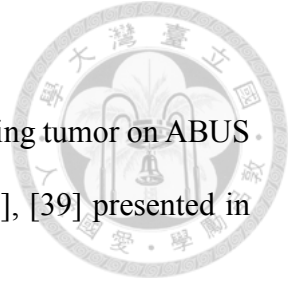
Type	False Positives at 100% sensitivity			
	2D merge	2D filter	3D merge	3D filter
Overall	257.99 ± 73.33	211.08 ± 61.00	57.03 ± 32.39	6.71 ± 4.06
Normal	231.25 ± 35.29	194.59 ± 34.55	61.00 ± 23.37	7.16 ± 2.61
Abnormal	267.00 ± 79.81	217.13 ± 65.23	55.87 ± 34.57	6.56 ± 4.42
Benign	219.20 ± 86.81	176.52 ± 67.05	44.89 ± 31.27	5.13 ± 4.73
Malignant	254.69 ± 75.03	207.70 ± 65.42	53.20 ± 35.22	5.96 ± 4.32
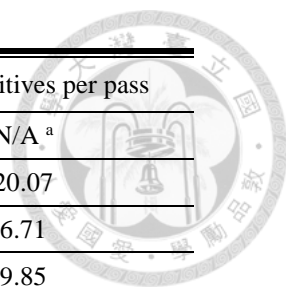
Table 5-11. The performance from four major parts of our proposed method. (Numbers for overall, normal, and abnormal types were false positives per pass, and numbers for benign and malignant types were false positives per tumor.)

Type	Sensitivity	False Positives (per pass)	p-value (Mann-Whitney U test)	p-value (Student's t-test)
Normal	100%	7.16 ± 2.61	>0.05	>0.05
Abnormal	100%	6.56 ± 4.42		
Benign	100%	5.27 ± 4.37	>0.05	>0.05
Malignant	100%	5.42 ± 4.36		

Table 5-12. Mann-Whitney U test p-values and Student's t-test p-values for different types of cases.

Square Size	Sensitivity	False Positives per pass
21	90%	10.32
23	100%	10.75
25	100%	6.71
27	100%	8.42
29	100%	10.96

Table 5-13. Performance with different values of square size.

57

Percentile of Grayscale Value	Sensitivity	False Positives per pass
0.15	97% [a]	N/A [a]
0.20	100%	20.07
0.25	100%	6.71
0.30	95%	9.85
0.35	100%	9.14

Table 5-14. Performance with different values of percentile of grayscale value. ([a]Since the result of extraction of suspect candidates did not reach 100% sensitivity, the process skipped 2D filter and 3D filter which demanded 100% sensitivity.)

t	Sensitivity	False Positives per pass
0.4	100%	6.71
0.6	100%	6.71
0.8	100%	6.71
1.0	100%	6.71
1.2	100%	6.71
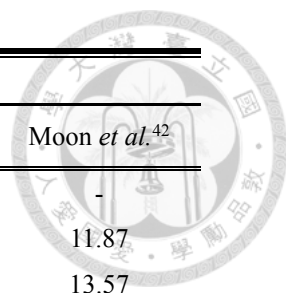
Table 5-15. Performance with different values of Threshold for t-value t.

f	Sensitivity	False Positives per pass
40	100%	8.31
45	100%	7.04
50	100%	6.71
55	100%	7.01
60	100%	6.79

Table 5-16. Performance with different values of Threshold for f-value f.

Parameters	Standard Deviation (False Positives per pass)
Square Size	1.75
Percentile of Grayscale Value	5.81
t	0.00
f	0.58

Table 5-17. Performance variation with respect to parameters at 100% sensitivity.

58

Sensitivity (%)	False Positives per pass			
	Our method	Chiang et al. [46]	Lo et al. [45]	Moon et al. [42]
90.00	5.14	1.15	5.42	-
93.38	-	1.94	-	11.87
95.59	-	2.42	-	13.57
97.01	-	4.06	-	--
99.26	-	-	-	15.42
100.00	6.71	> 20.00	9.44	17.45
Execution time (seconds/per pass)	99.21	20.17	74.00	780

Table 5-18. Comparison of performance in terms of sensitivity and corresponding FPs. (Note: Since Chiang *et al.* reported results in terms of FPs per patient whereas other literatures used FPs per pass, and each patient scanned in six passes for entire breasts from a standardized examination, so the results obtained by Chiang *et al.* were divide by 6 to allow fair comparison.)

59

Chapter 6   Discussions and conclusions
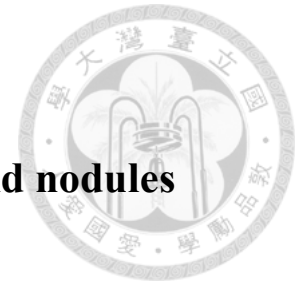
6.1 Boundary segmentation for ultrasonic thyroid nodules

6.1.1   Discussion

The purpose of this study was to propose a novel method for semi-automatic segmentation of the boundary of thyroid nodules on non-preprocessed ultrasound images. Therefore, the correlation and accuracy between AB and MB was an important issue.

The results in Fig. 5-3 and Fig. 5-4 showed that the application of our proposed method on clinical thyroid nodules perfectly conformed to the original design concept, and the boundary drawn by our proposed method agreed well with gold-standard, as shown in Fig. 5-5. Furthermore, our proposed method achieved high effectiveness and successfully prevented the influence from the variation of the nodule's inner tissues without any preprocessing of ultrasound images. It could handle complicated cases with blurred boundaries, inhomogeneity, or cysts in nodule, as shown in Fig. 5-6.

The overall results of a large amount of cases showed that our proposed method achieved a high degree of correlation between AB and MB with low error rates (summarized in Table 5-2), as demonstrated in Fig. 5-5, Fig. 5-6, and the second and third row of Fig. 5-7. Besides, the results showed that the difference of all the metrics between the training and testing dataset was around 0.35%, and this difference was statistically insignificant proved by Mann-Whitney U test. These results clearly indicated the reliability and consistency of our proposed method.

The results of benign and malignant nodules showed that our proposed method performed ideally both on benign or malignant cases (Table 5-3). The performance metrics of benign cases were slightly better than that of malignant cases, but the difference in-between was around only 0.46% and statistically insignificant, expressing the

reliability and high degree of accuracy on both benign and malignant cases.

Three standardized methods used for comparison were chosen in order to avoid additional adjustments, thereby reinforcing the impartiality and credibility of the comparison. The manipulations to be avoided included additional filtering methods, extra parameters, human intervention, or subjective errors. Consequently, two of the standardized methods, ACM and WM, established with OpenCV library were chosen for comparison as in work by Chang *et al.*[44]. The third standardized method, DRLSE [45], was chosen in comparison and realized with the source code provided by author(s) [46].
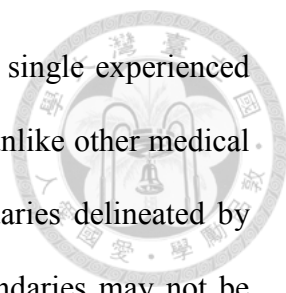
The comparison results with the same dataset were shown in Table 5-4, Table 5-5, and Fig. 5-7. Our proposed method significantly outperformed ACM in all performance metrics and also significantly outperformed DRLSE in most metrics except FPR. As for comparison with WM, the TPR in WM was 1% slightly higher than our proposed method, but was accompanied by the largest FPR (17.50%) due to the boundaries outreaching unnecessary areas, as shown in the fourth row of Fig. 5-7. The FPR in WM was 9.82%, 9.19%, and 9.75% higher than our proposed method, ACM, and DRLSE, respectively. The NHD in WM was much larger than other methods comparative to NMD, and this implied that there were extreme outliers in WM. Besides, the JSI and PPV of our proposed method were 6.06% and 8.00% significantly higher than that of WM, respectively. WM referred to more information from human intervention, but detected boundary with excessive sensitivity leading to a large error rate, therefore reduced the effectiveness of the boundary segmentation.

Sensitivity analysis showed that the performance varies with different values of w and $\theta_{Threshold}$, but not with a and b. It was observed that larger w resulted in worse performance. This might be due to that a larger value of w might cover too many unnecessary boundary candidate points. By the way, smaller $\theta_{Threshold}$ was accompanied with worse performance.

61

Too small value for $\theta_{Threshold}$ limited the path searching and might cause the path to deviate from the real boundary of nodule, therefore resulting in low TPR. On the other hand, the choice of value for *a* and *b* had very limited influence on overall performance. It implied that direction searching method was the principal filtering step of our proposed method, so the two following steps only played an auxiliary role.

Since OpenCV is an optimized cross-platform library [47], the two methods coded with OpenCV, ACM and WM, showed good time performance. However, WM was slightly slower than our proposed method due to the initial requirement of distinguishing the foreground from the background. DRLSE had a poor time performance of 16.01 seconds per nodule. It might be attributed to two reasons: the first was that DRLSE was realized with the version of full domain implementation rather than the version of narrowband implementation, and had higher computational cost [45]. The second reason was that the author(s) provided DRLSE source code in MATLAB, different from the Microsoft Visual Studio C# we used to code our algorithm and perform the comparison study. The transition efficiency between MATLAB and the Microsoft Visual Studio C# had been reported as a problem [48], [49].

Speckle poses a challenge to computer-aided systems for ultrasound images but also provides useful information about tissue characteristics [9]-[14], [15]-[16]. Koundal *et al.* applied Gaussian filter in clustering and contour evolution to reduce the impact of speckle [14]. Similarly, Savelonas *et al.* utilized local binary pattern operator which represented regional textural properties instead of single pixel [12]. These approaches averaged out speckle but meanwhile erased the information carried by the speckles. On the contrary, our method did not smooth out the speckles in every step of the algorithm, and, therefore, the speckle information was kept and was also an integral part of our proposed segmentation algorithm.

62

The gold-standard for performance validation was defined by a single experienced radiologist. The reason for not using multiple radiologists was that unlike other medical decisions it is not straightforward to pool together different boundaries delineated by different radiologists. Averaging different manually delineated boundaries may not be appropriate because the discrepancies are usually due to different viewpoints about the same nodule image instead of the observation errors. For example, for lobulated nodules radiologists usually have different viewpoints on whether a certain adjacent mass should be included as part of the nodule. To avoid such issue, we used the boundary manually delineated by a single experienced radiologist as the gold-standard in this study. Nevertheless, the gold-standard boundary established by only one experienced radiologist was still a limitation of this study.

6.1.2 Conclusion

In conclusion, we proposed a novel and semi-automatic method based on V-R statistics for thyroid nodules boundary segmentation on non-preprocessed ultrasound images. The results showed that our proposed method was reliable and performed excellently, and was much more effective than other standardized methods.

6.1.3 Future Research

We will try to incorporate the capability to automatically detect tumor from thyroid ultrasound images into current system. This will allow the system to become fully automatic and will no longer require manual input by operator.

6.2 Breast tumor detection in 3D ultrasound imaging

6.2.1 Discussion

The purpose of this study was to propose an intuitive, simple, efficient, and automatic method for detecting tumor on ABUS images. Therefore, achieving 100% sensitivity with

as low as possible FPs was an important objective in this study.

The results in Fig. 5-8 showed that our proposed method perfectly detected benign and malignant clinical breast tumors. The classifier of 3D filter ideally identified the tumor and correctly excluded the hypoechogenic regions at the bottom and lateral sides of image.

Fig. 5-9 showed examples that our proposed method could correctly detect tumor, but also misinterpreted shadows located in the lateral (left column) and rib (right column) regions. These conditions mislead our proposed system and resulted in false positives. The shortcoming may be solved by the inclusion of more potent features in the future.

In Table 5-11, 2D filter played an auxiliary role to enhance the performance of 3D merge. Therefore, only the basic features, intensity-related features, were used to filter out candidates in order to further improve the performance of 3D merge.

On the other hand, 3D filter was the most important part of our proposed method considering four types of features to reduce the FPs. The results in Table 5-11 showed that our classifier in 3D filter had an excellent filtering ability. It could filter 88.24%, 88.26%, 88.26%, 87.57%, and 88.79% false positives from previous step in overall, normal, abnormal, benign, and malignant cases, respectively, with a rapid processing time as stated above. Among them, the 3D filter had an identical filtering ability for normal and abnormal passes, proving the generalization of our proposed system. Besides, the filtering ability in malignant cases was slightly better than in benign cases, ensuring the clinical usefulness of our proposed system. Furthermore, the final performance as measured with FPs at 100% sensitivity were also outstanding and impartial, regardless of type of data. The difference between normal and abnormal passes was 0.6 FPs per pass, and the difference between benign and malignant cases was 0.83 FPs per tumor. Both differences were minimal and statistically insignificant (Table 5-12), showing that our

64

system was stable and reliable throughout all conditions, and the four types of chosen features were suitable regardless of normal/abnormal or benign/malignant ABUS images.

Sensitivity analysis regarding tunable parameters showed that performance varied only with different values of the percentile of grayscale value, but not with other parameters. It was observed that smaller value for percentile of grayscale value resulted in worse performance and even caused problems in the procedure. Too small value for percentile of grayscale value might filter out too many eligible square cells in the step of 2D merge, resulting in two possible issues. First, the result of 2D merge could not achieve 100% sensitivity, causing 2D filter and 3D filter to be skipped since both steps demanded 100% sensitivity. Second, such over-filtering might erase important information about tissue characteristics, so that the subsequent steps did not have enough information to detect correctly. On the other hand, the choice of value for other parameters had nearly no influence on overall performance. This showed that the step of 2D merge was crucial for the whole detection algorithm.

In the comparison with previous publications presented in Table 5-18 [35], [38], [39], the most relevant figures were FPs at 100% sensitivity, since the goal of CADe system is to correctly detect lesions without miss while minimize erroneous detections. It showed that our proposed method with 6.71 FPs outperformed previously published methods at 100% sensitivity. Besides, the method proposed by Chiang *et al* [39] had lower FPs at 90% sensitivity was obviously lower than all other methods, but accompanied by extremely largest FPs at 100% sensitivity.

The overall processing time of our proposed CADe system was 99.2094 seconds per pass, slower than two among three methods listed in Table 5-18. However, it was mainly attributed to the step of 2D merge, while the other three steps were less than 1 second in sum. Since the original ABUS images contained a large amount of image information and

65

had a large number of image slices, the step of 2D merge was the time-limiting step. Consequently, shortening the processing time for extracting image information at the beginning of the system will phenomenally improve the overall efficiency, and this will be our main future work.

The uniqueness of our proposed method were simplicity and intuitiveness. Compared to the previously proposed methods, our method implemented binary classifiers and did not involve any complex algorithms. If a system is built on complex algorithms, understanding and interpreting how the system works may be difficult for end-users, and even for developers themselves. The lack of transparency and accountability of such "black-box" system may lead to severe consequences [61]. Besides, with the exception of 2D merge, the steps of 2D filter, 3D merge, and 3D filter were very effective and efficient and required no parameters. Therefore, simplifying the step of 2D merge will also be our future objective. On the other hand, the limitation of our CADe system was that it was not able to further sketch contour on detected mass. For further computer-aided diagnosis, tumor contour is an important feature for classifying benign and malignant masses.

Though the sample size used in our study, 176 passes, was not small compared to relevant papers [33]-[39] in the literature, there was still room for further improvement on the credibility and reliability of the study results by enriching the dataset in the future.

6.2.2 Conclusion

In conclusion, we proposed an intuitive, simple, efficient, and automatic CADe system to detect breast tumors in ABUS images. The results showed that our proposed method was reliable and efficient with 6.71 FPs per pass at 100% sensitivity, much more effective than previously proposed methods.
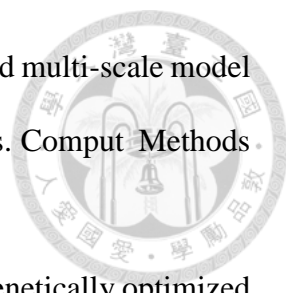
6.2.3　Future Research

　　We will increase the sample size involved in the research to strengthen the credibility of our proposed CADe system. Moreover, we will combine boundary segmentation capability, as described in the first topic of this dissertation, into current CADe system.
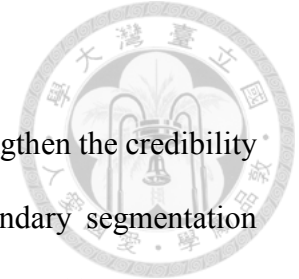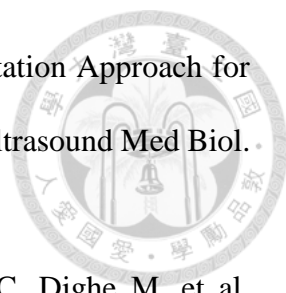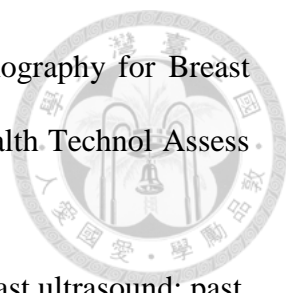
Reference
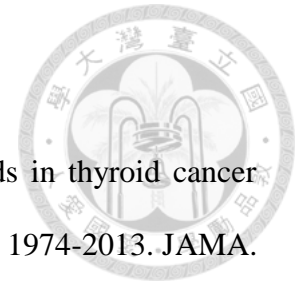
[1] Lim H, Devesa SS, Sosa JA, Check D, Kitahara CM. Trends in thyroid cancer incidence and mortality in the United States, 1974-2013. JAMA. 2017;317(13):1338-1348.

[2] Davies L, Welch HG. Current thyroid cancer trends in the United States. JAMA Otolaryngol Head Neck Surg. 2014;140(4):317-322.

[3] Baker SG. The central role of receiver operating characteristic (ROC) curves in evaluating tests for the early detection of cancer. J Natl Cancer Inst. 2003;95(7):511-515.

[4] Nabhan F, Ringel MD. Thyroid nodules and cancer management guidelines: comparisons and controversies. Endocr Relat Cancer. 2017;24(2):R13-R26.

[5] Chen K-Y, Chen C-N, Wu M-H, Ho M-C, Tai H-C, Kuo W-H, Huang W-C, Wang Y-H, Chen A, Chang K-J. Computerized quantification of ultrasonic heterogeneity in thyroid nodules. Ultrasound Med Biol. 2014;40(11):2581-2589.

[6] Wu M-H, Chen C-N, Chen K-Y, Ho M-C, Tai H-C, Wang Y-H, Chen A, Chang K-J. Quantitative analysis of echogenicity for patients with thyroid nodules. Sci Rep. 2016;6:35632.

[7] Chi J, Walia E, Babyn P, Wang J, Groot G, Eramian M. Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network. J Digit Imaging. 2017;30(4):477-486.

[8] Fish SA. Incidental Thyroid Nodules Detected on CT, MRI, or PET-CT Correlate Well with Subsequent Ultrasound Evaluation. Clin Thyroidol. 2017;29(3):107-109.

[9] Tsantis S, Dimitropoulos N, Cavouras D, Nikiforidis G. A hybrid multi-scale model for thyroid nodule boundary detection on ultrasound images. Comput Methods Programs Biomed. 2006;84(2-3):86-98.

[10] Iakovidis DK, Savelonas MA, Karkanis SA, Maroulis DE. A genetically optimized level set approach to segmentation of thyroid ultrasound images. Appl Intell. 2007;27(3):193-203.

[11] Maroulis DE, Savelonas MA, Iakovidis DK, Karkanis SA, Dimitropoulos N. Variable background active contour model for computer-aided delineation of nodules in thyroid ultrasound images. IEEE Trans Inf Technol Biomed. 2007;11(5):537-543.

[12] Savelonas MA, Iakovidis DK, Legakis I, Maroulis D. Active contours guided by echogenicity and texture for delineation of thyroid nodules in ultrasound images. IEEE Trans Inf Technol Biomed. 2009;13(4):519-527.

[13] Chiu L-Y, Chen A. A variance-reduction method for thyroid nodule boundary detection on ultrasound images. In 2014 IEEE International Conference on Automation Science and Engineering; 2014 Aug 18-22; New Taipei, Taiwan. IEEE; 2014. p. 681-685.

[14] Koundal D, Gupta S, Singh S. Automated delineation of thyroid nodules in ultrasound images using spatial neutrosophic clustering and level set. Applied Soft Computing. 2016;40:86-97.

[15] Madabhushi A, Metaxas DN. Combining Low-, High-Level and Empirical Domain Knowledge for Automated Segmentation of Ultrasonic Breast Lesions. IEEE Trans Med Imaging. 2003;22(2):155-169.

[16] Shan J, Cheng HD, Wang Y. Completely Automated Segmentation Approach for Breast Ultrasound Images Using Multiple-Domain Features. Ultrasound Med Biol. 2012;38(2):262-275.

[17] Cosgrove D, Barr R, Bojunga J, Cantisani V, Chammas MC, Dighe M, et al. WFUMB Guidelines and Recommendations on the Clinical Use of Ultrasound Elastography: Part 4. Thyroid. Ultrasound Med Biol. 2017;43(1):4-26.

[18] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA Cancer J Clin 2017;67(1):7-30.

[19] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global Cancer Statistics, 2012. CA Cancer J Clin 2015;65(2):87-108.

[20] Coleman C. Early detection and screening for breast cancer. Semin Oncol Nurs 2017;33(2):141-155.

[21] Eisenbrey JR, Dave JK, Forsberg F. Recent technological advancements in breast ultrasound. Ultrasonics 2016;70:183-190.

[22] Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography physical examination and breast US and evaluation of factors that influence them: An analysis of 27825 patient evaluations. Radiology 2002;225(1):165-175.

[23] Okello J, Kisembo H, Bugeza S, Galukande M. Breast cancer detection using sonography in women with mammographically dense breasts. BMC Med Imaging 2014;14:41.

[24] Zonderland HM, Coerkamp EG, Hermans J, van de Vijver MJ, van Voorthuisen AE. Diagnosis of breast cancer: contribution of US as an adjunct to mammography. Radiology 1999;213(2):413-422.

[25] Health Quality Ontario. Ultrasound as an Adjunct to Mammography for Breast Cancer Screening: A Health Technology Assessment. Ont Health Technol Assess Ser 2016;16(15):1-71.

[26] Brem RF, Lenihan MJ, Lieberman J, Torrente J. Screening breast ultrasound: past, present, and future. Am J Roentgenol 2015;204(2):234-240.

[27] Golatta M, Baggs C, Schweitzer-Martin M, Domschke C, Schott S, Harcos A, et al. Evaluation of an automated breast 3D-ultrasound system by comparing it with hand-held ultrasound (HHUS) and mammography. Arch Gynecol Obstet 2015;291(4):889-895.

[28] Kelly KM, Dean J, Comulada WS, Lee SJ. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. Eur Radiol 2010;20(3):734-742.

[29] Gubern-Mérida A, Tan T, van Zelst J, Mann RM, Karssemeijer N. Automated linking of suspicious findings between automated 3D breast ultrasound volumes. In Tourassi GD, Armato SG, editors. Published Proceedings of SPIE Medical Imaging 2016 - Medical Imaging 2016: Computer-Aided Diagnosis; February 27-March 3, 2016; San Diego, CA. https://doi.org/10.1117/12.2214945.

[30] van Zelst JCM, Tan T, Platel B, de Jong M, Steenbakkers A, Mourits M, et al. Improved cancer detection in automated breast ultrasound by radiologists using Computer Aided Detection. Eur J Radiol 2017;89:54-59.

[31] Ma J, Wu F, Jiang T, Zhao Q, Kong D. Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks. Int J CARS. 2017;12:1895-1910.

[32] Chang K-J, Chen W-H, Chen A, Chen C-N, Ho M-C, Tai H-C, Wu M-H, Tsai P-W. inventors; AmCad BioMed Corporation, assignee. Method for retrieving a
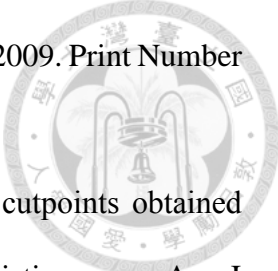
tumor contour of an image processing system. US patent 8,374,892. February 12, 2013.

[33] Ikedo Y, Fukuoka D, Hara T, Fujita H, Takada E, Endo T, et al. Development of a fully automatic scheme for detection of masses in whole breast ultrasound images. Med Phys 2007;34(11):4378-4388.

[34] Chang RF, Chang-Chien KC, Takada E, Huang CS, Chou YH, Kuo CM, et al. Rapid image stitching and computer-aided detection for multipass automated breast ultrasound. Med Phys 2010;37(5):2063-2073.

[35] Moon WK, Shen YW, Bae MS, Huang CS, Chen JH, Chang RF. Computer-aided tumor detection based on multi-scale blob detection algorithm in automated breast ultrasound images. IEEE Trans Med Imaging 2013;32(7):1191-1200.

[36] Tan T, Platel B, Mus R, Tabár L, Mann RM, Karssemeijer N. Computer-aided detection of cancer in automated 3-D breast ultrasound. IEEE Trans Med Imaging 2013;32(9):1698-1706.

[37] Moon WK, Lo CM, Chen RT, Shen YW, Chang JM, Huang CS, et al. Tumor detection in automated breast ultrasound images using quantitative tissue clustering. Med Phys 2014;41(4):042901.

[38] Lo CM, Chen RT, Chang YC, Yang YW, Hung MJ, Huang CS, et al. Multi-dimensional tumor detection in automated whole breast ultrasound using topographic watershed. IEEE Trans Med Imaging 2014;33(7):1503-1511.

[39] Chiang TC, Huang YS, Chen RT, Huang CS, Chang RF. Tumor detection in automated breast ultrasound using 3-D CNN and prioritized candidate aggregation. IEEE Trans Med Imaging 2019;38(1):240-249.

[40] Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Notification [510(k)] Submissions. U.S. Food and Drug

Administration website. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/computer-assisted-detection-devices-applied-radiology-images-and-radiology-device-data-premarket. Accessed May 3, 2020.
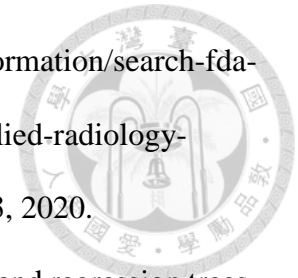
[41] Breiman L, Friedman JH, Stone CJ, Olshen RA. Classification and regression trees. CRC press; 1984.

[42] Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. Int J Comput Vision. 1988;1(4):321-331.

[43] Vincent L, Soille P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans Pattern Anal Mach Intell. 1991;13(6):583-598.

[44] Chang C-Y, Lei Y-F, Tseng C-H, Shih S-R. Thyroid segmentation and volume estimation in ultrasound images. IEEE Trans Biomed Eng. 2010;57(6):1348-1357.

[45] Li C, Xu C, Gui C, Fox MD. Distance regularized level set evolution and its application to image segmentation. IEEE Trans Image Process. 2010;19(12):3243-3254.

[46] Li C. Level set for image segmentation. MathWorks [document on the Internet]; 2013, Jun 3 [cited 2018 Oct 12]. Available from: https://www.mathworks.com/matlabcentral/fileexchange/12711-level-set-for-image-segmentation.

[47] Aby PK, Jose A, Jose B, Dinu LD, John J, Sabarinath G. Implementation and optimization of embedded face detection system. In Proceedings of 2011 International Conference on Signal Processing, Communication, Computing and Network Technologies; 2011 Jul 21-22; Thuckafay, India. IEEE; 2011. p. 250-253.

[48] Zhang Y, Ma Y-J, Yuan Q-L. The method and implementation of mixed programming with Visual C# and Matlab. Journal of Northwest Normal University (Natural Science). 2008;6:012.

[49] Hong L, Cai J. The application guide of mixed programming between MATLAB and other programming languages. In 2010 the 2nd International Conference on Computer and Automation Engineering; 2010 Feb 26-28; Singapore, Singapore. IEEE; 2010. vol. 3, p. 185-189.

[50] Billah MS, Mahmud TB. Semi-Automatic Segmentation and Ultrasonic Characterization of Solid Breast Lesions. B.S. [thesis]. Dhaka: Bangladesh University of Engineering and Technology; 2017. Available from: arXiv preprint. arXiv:1703.08238.

[51] Acharya T, Ray AK. Image Processing: Principles and Applications. Hoboken, NJ: John Wiley & Sons Inc.; 2005

[52] Lai YR, Tsai PC, Yao CY, Ruan SJ. Improved local histogram equalization with gradient-based weighting process for edge preservation. Multimed Tools Appl 2017;76(1):1585-1613.

[53] ContrastEnhancer.java. ImageJ: Image Processing and Analysis in JAVA. https://imagej.nih.gov/ij/developer/source/ij/plugin/ContrastEnhancer.java.html . Accessed November 2, 2019.

[54] Fine RE, Staren ED. Updates in breast ultrasound. Surg Clin North Am 2004;84(4):1001-1034.

[55] Liberman L, Menell JH. Breast imaging reporting and data system (BI-RADS). Radiol Clin N Am 2002;40(3):409-430.

[56] Shan J, Alam SK, Garra B, Zhang Y, Ahmed T. Computer-aided diagnosis for breast ultrasound using computerized BI-RADS features and machine learning methods. Ultrasound Med Biol 2016;42(4):980-988.

[57] Manly BFJ, Alberto JAN. Multivariate statistical methods: a primer. 4th ed. New York, NY: Chapman and Hall/CRC; 2016

[58] ABVS Workplace [operator manual]. Muenchen: Siemens AG; 2009. Print Number SPB7-270.620.50.02.24.

[59] Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol 2006;163(7):670-675.

[60] Chakraborty DP, Winter LHL. Free-Response Methodology: Alternate Analysis and a New Observer-Performance Experiment. Radiology 1990;174:873-881.

[61] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1(5):206-215.