私 立 東 海 大 學

資 訊 工 程 研 究 所

碩士論文

指導教授：黃育仁 博士

Advisor：Associate Professor Yu-Len Huang

三維超音波之乳房腫瘤惡性度估測

Tumor Grade Estimation for Breast Cancer in 3D Ultrasound Image

研究生：陳珮妮

Name：Pei-Ni Chen

中 華 民 國 一 百 零 二 年 七 月

July 2013

摘要

在顯微鏡觀看的異常癌細胞可經由腫瘤分級系統來區分，以期能預測腫瘤生長速度和擴散的可能性。乳腺癌腫瘤的分級信息是關於腫瘤的生長和擴散的速度，也可以用來預測它對治療反應是否良好。醫生可以依據腫瘤的分級和許多其他因素，如癌症分期，為病人制定個人治療方案，並預測病人的預後。在這項研究中，由三維(3D)超音波冠狀病理圖像(C-視圖成像)進行擷取圖像特徵分析和分類，協助醫生估計乳腺腫瘤級。本研究的目的是檢查腫瘤紋理特徵和腫瘤的惡性程度之間的相關性，從冠狀切面三維影像提取代表性的特徵指數，確立腫瘤的紋理變化和乳腺組織學分級之間的數值關係，本研究先利用去除斑點的前處理過程後，運用聲像雙邊濾波器提取16個紋理特徵，衡量的數量和空間分佈的特點進行評估，經由特徵決定腫瘤等級的高低。分類的結果顯示，所採用的特徵指標具有良好的類別分辨能力，非常有潛力可以用來預測腫瘤的惡性等級。

關鍵字： 乳癌，乳腺癌，三維超音波，病理分級，活組織切片檢查

ABSTRACT

Tumor grade is a system used to classify cancer cells in terms of how abnormal they look under a microscope and how quickly the tumor is likely to grow and spread. Tumor grade information of breast carcinoma is available about tumor growth and proliferation of speed, also can be used to predict whether it can make a good response to treatment. Physicians always consult the tumor grade and many other factors, such as cancer stage, to develop an individual treatment plan for the patient and try to forecast the patient's prognosis. In this study, the coronal plane image (C-view imaging) from three-dimensional (3D) ultrasography is performed for analysis and classification by capturing image characteristics to help physicians estimating the breast tumor grade. For the purpose of inspecting correlativity between the tumor texture on sonogram and the malignancy of tumor, this study extracts representative feature indices from coronal section of 3D sonogram such that the numerical relation between texture variation and histological grade of breast cancer can be established. After the proposed speckle reducing procedure, the features of architectural distortion can be stably extracted by applying bilateral filter on the sonogram. Sixteen indices are evaluated to measure the amount and spatial distribution of the features; these indices are then voted by decision trees to identify a tumor as either low or high grade. The performance of classification reveals the well discriminating abilities of these indices; notable results indicate that the proposed feature indices have mighty potential to predict tumor grades.

Keywords: tumor; breast cancer; 3D sonography; tumor grade; biopsy

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death among women. About 232,340 new cases of invasive breast cancer were identified in women for 2013. Breast cancer ranks second as a cause of cancer death in women. The early diagnostic is essential to increase the therapy efficacy [1]. The American Cancer Society (ACS) suggested that the 20-40 years-old women should have regular breast screening through the screening will be able to reduce breast cancer death case to 45% [2]. The most effective method to reduce the breast cancer death rate is early to discover and treat unite diseases of breast cancer. Therefore the high reliability test results are an important basis for early diagnosis.

In addition to mammogram, ultrasound (US) and breast magnetic resonance (MR) image have been used to improve diagnosis and reduce the number of unnecessary biopsies for patients who have palpable mass and inconclusive mammograms [3-5]. Three-dimensional (3D) ultrasography has been known as the inexpensive, portable, safe and widely available tool in breast cancer screening so far. Diagnostic ultrasound has no side effects on human body. Advanced ultrasound scanner is able to provide real-time and high-resolution medical imaging. Breast ultrasound examinations often

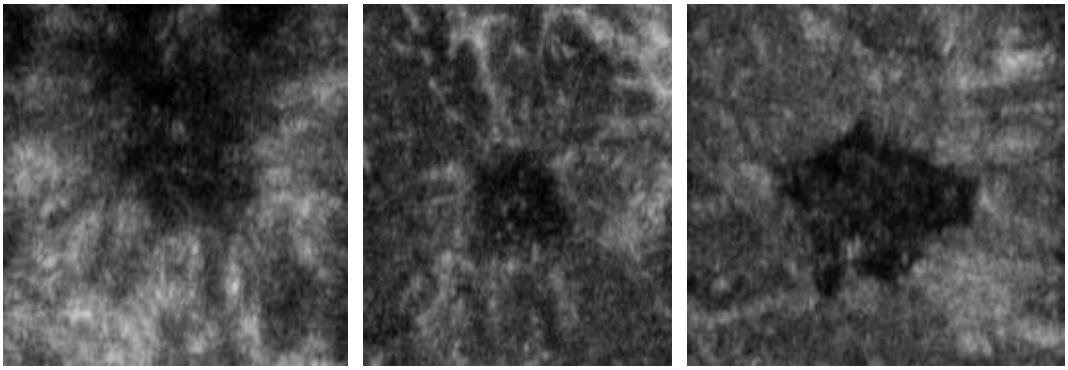
need to match in order to enhance the mammograms found the correct rate of lesions. In addition, the Oriental female breasts are denser than that of West female, which may affect the interpretation of the mammography, so the importance of breast ultrasound diagnosis of breast tumor is ineffable [6].

Histological grade is one of the best-established prognostic factors in breast cancer, which represents the morphological assessment of tumor biological characteristics. It has been shown to be able to generate important information related to the clinical behavior of breast cancers. For a large collection of examples of textures, see Brodatz 1966, feature extraction to compute a characteristic of a digital image able to numerically describe its textural properties. Texture classification is used to determine to which of a finite number of physically defined classes (such as normal and abnormal tissue) a homogeneous texture region belongs [7]. Texture analysis is concerned with the study of the variation in intensity of image elements (pixel) values acquired under certain conditions. From a medical imaging perspective, physical quantities at scales smaller than the scales of interest can be analyzed for proper classification [8-9]. Therefore, analysis of the ultrasound images could help doctors to determine the tumor characteristics. The number of tumor biopsy is in hopes of significant decreasing.

Physicians utilized the tumor grade and other factors, such as cancer stage and a

patient’s age and general health, to develop a treatment plan and to determine a patient’s prognosis (the likely outcome or course of a disease; the chance of recovery or recurrence). Generally, a lower grade indicates a better prognosis. A higher-grade cancer may grow and spread more quickly and may require immediate or more aggressive treatment [10]. The importance of tumor grade in planning treatment and determining a patient’s prognosis is greater for certain types of cancer, such as soft tissue sarcoma, primary brain tumors, and breast and prostate cancer.

This study aims to extract representative feature index from coronal section of 3D sonograms. The proposed method inspected the correlation between sonograms and the malignant tumor grade. The numerical relation was established to estimating the breast tumor grade by using the statistical features in 3D breast ultrasography [11]. Figure 1 shows three image slices of coronal section from three different cases. Those hyper-echoic lines in all three images indicate more or less a certain degree of architectural distortion, but masses are all solid and definitely visible.



(a) (b) (c)

Fig. 1. Example slices from coronal section of 3D sonogram: (a) Grade I case; (b) Grade II case; (c) Grade III case

CHAPTER 2

MATERIALS

2.1 Data Acquisition

One hundred and ninety one cases were obtained for this retrospective study from Changhua Christian Hospital, Changhua, Taiwan. All patients of 191 tumors have been diagnosed as infiltrating ductal carcinoma or mixed type cases. The 191 data composed of three parts: 23 cases are Grade I, 128 cases are Grade II, 40 cases are Grade III. Sonographic examinations were operated on the GE Voluson 730 3D ultrasound system prior to biopsy or surgery, and resolutions of the images have values varied in range from 0.013 to 0.029 cm/pixel among different cases. For the reason that strong converging pattern shall be revealed, only coronal section of 3D volumetric data was used in the study, and the slices of 2D images on coronal section were reconstructed from 3D volumetric data accordingly. The central nearest location of mass, also known as region of interest (ROI), was marked on three successive slices of coronal plane for each case by one experienced physician.

The reconstructed coronal view and corresponding ROIs were reviewed by another senior physician. To extract the region of interest in 3D ultrasound imaging, a physician with experience in breast ultrasound examination defined and manually

selected region of interest (ROI) including the central nearest location of tumor in the specific three slices, i.e. the center front, center and center behind slices in 3D ultrasound imaging. This study selected the central nearest location of the tumor images that physicians painted, and we used 69 slices Grade I and 120 slices Grade III of 3D ultrasound images as research data.

Table 1. The number of cases for tubule formation, nuclear pleomorphism and mitotic count that has used in this study

	Tubule	**Nuclear**	**Mitosis**
Score 1	14	9	130
Score 2	51	86	48
Score 3	126	96	13
Total	191	191	191

2.2 Tumor Grade

For all tumors, specimens were taken through biopsy or surgery after sonographic examinations, and each tumor received a series of histopathological result along with the corresponding sonogram. A dozen of indices were included in the histopathological results, but only the histological grade was analyzed for the

purpose of this study. The histological grade of breast carcinoma was scored according to the Nottingham-modified Bloom-Richardson grading system [12], which is based on the microscopic evaluation of morphological and cytological features of tumor cells. Three intermediate scores have been defined for tubule formation, nuclear pleomorphism, and mitotic count in the grading system. Each of the three scores is given one to three points, and an overall grading score is given by adding three scores together. The highest possible score is nine (3+3+3). The exact criteria for each component differ in each system and the systems are evolving as more detailed data becomes available. The corresponding grades are list as below.

Tubular Formation

The criteria of tubular formation are as follows:

- Score 1: >75% of tumor area forming glandular/tubular structures
- Score 2: 10% to 75% of tumor area forming glandular/tubular structures
- Score 3: <10% of tumor area forming glandular/tubular structures

Nuclear Grade

The criteria of nuclear grade are as follows:

- Score 1: Nuclei small with little increase in size in comparison with normal breast epithelial cells, regular outlines, uniform nuclear chromatin, little variation in size
- Score 2: Cells larger than normal with open vesicular nuclei, visible nucleoli, and moderate variability in both size and shape
- Score 3: Vesicular nuclei, often with prominent nucleoli, exhibiting marked variation in size and shape, occasionally with very large and bizarre forms

Mitosis Score

The mitotic count score criteria vary depending on the field diameter of the microscope used by the pathologist. The pathologist will count how many mitotic figures are seen in 10 high power fields. Using a high power field diameter of 0.50 mm, the criteria are as follows:

- Score 1: less than or equal to 7 mitoses per high power field
- Score 2: 8-14 mitoses per high power field
- Score 3: equal to or greater than 15 mitoses per high power field

Tumor Grade

The criteria of tumor grade are as follows:

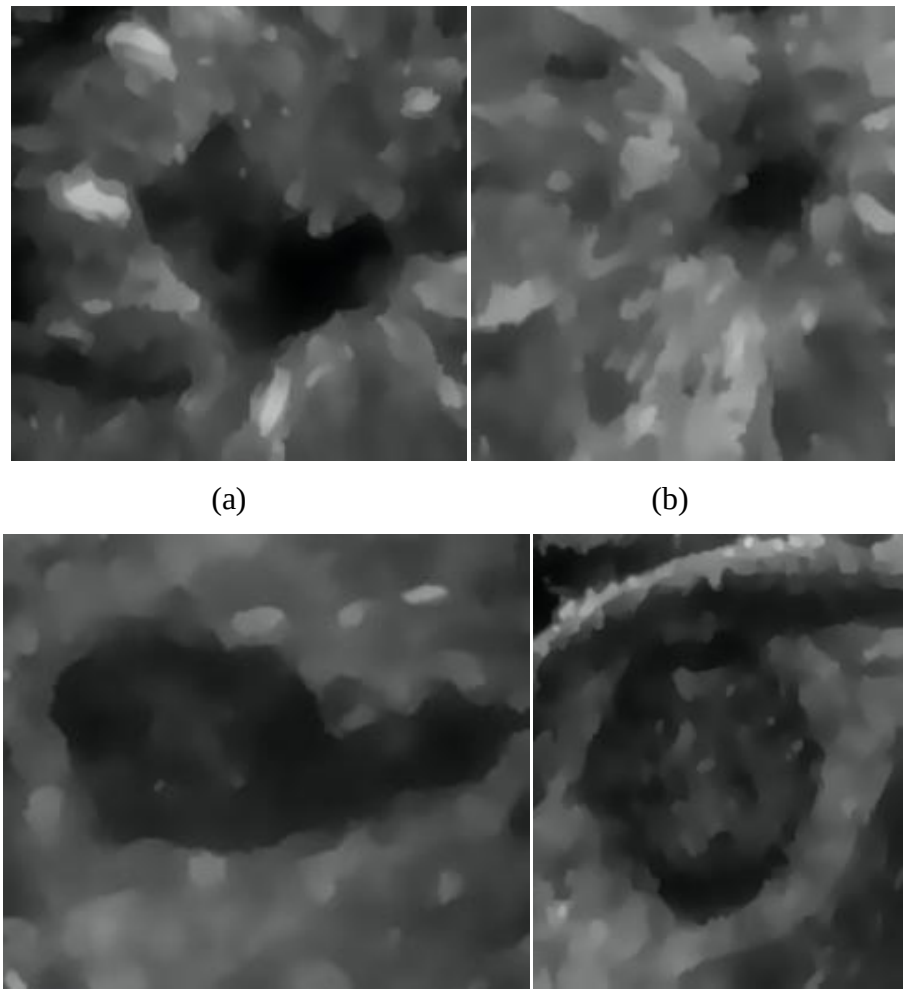
- Grade I: scores of 3, 4, or 5
- Grade II: scores of 6 or 7
- Grade III: scores of 8 or 9

CHAPTER 3

METHODS

3.1 Image Preprocessing

The reduced contrast resolution breast ultrasound images often contain noise and spots. Thus preprocessing is crucial to extracting the feature precisely for noisy images. Two kinds of noise, speckle noise and synthetic flaw, were found commonly existed in our source data. The speckle noise is generally recognized as inherent characteristic of ultrasound imaging, but the synthetic flaw is believed to be the side effect that one anatomical section is synthesized from the slices of another section. The coronal section performed in the proposed method was reconstructed from the sagittal slices as the model which shows in Fig. 2. The two chosen ROI slices in reconstructed coronal section were preprocessed for reducing the noise before feature extraction. To eliminating the speckled noise while also enhancing feature, the bilateral filter recursively improve contrast resolution on the coronal plane images, and eliminate the speckled noise while also enhancing the tumor texture feature.



(a) (b)

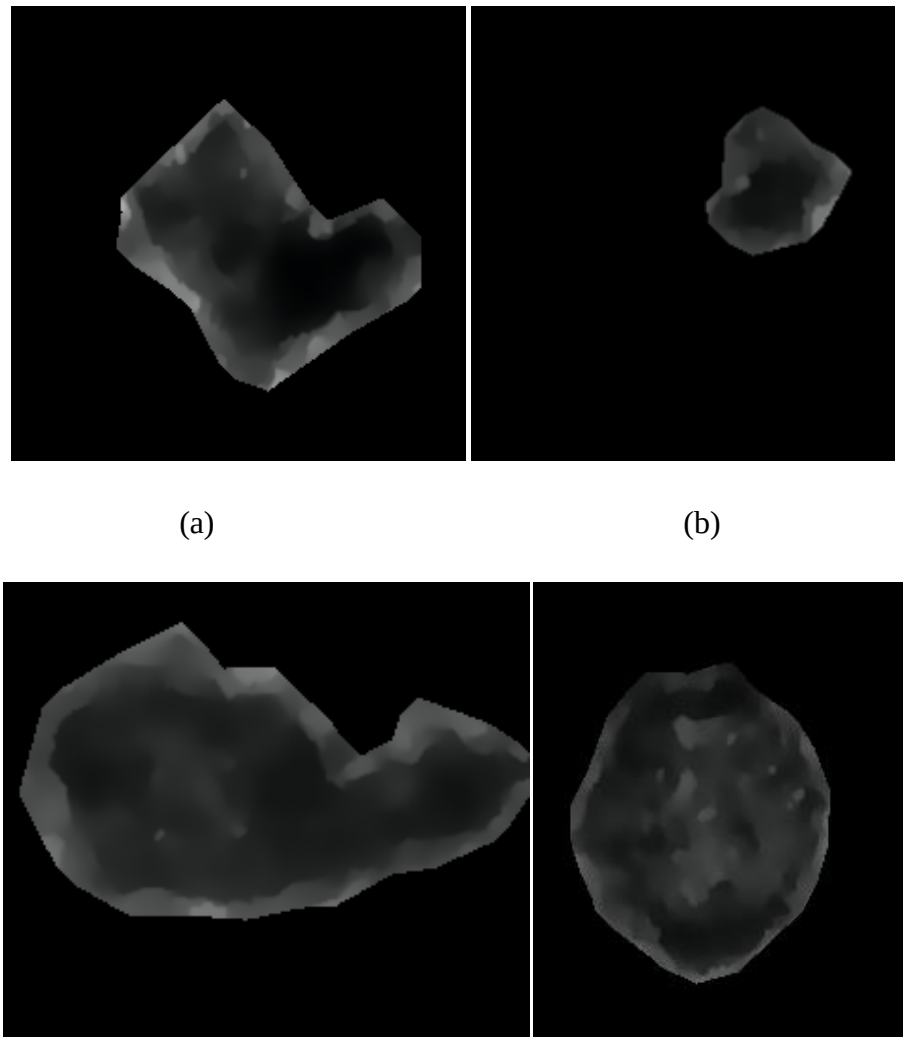
(c) (d)

Fig. 2. The example of the pre-processing image: (a)(b) the corresponding pre-processed images (Grade I tumor); (c)(d) the corresponding pre-processed images (Grade III tumor).

The bilateral filter's weight is based on a Gaussian distribution. Bilateral filter is a nonlinear filter, originally proposed to image processing to effectively noise level slider technology [13]. The main principle is Gauss-level slide handling in the spatial domain and the intensity domain at the same time. This study utilized the bilateral filter as the pre-processing procedure to eliminate the speckle noise while also to

enhance texture/boundary feature on the coronal plane images. The bilateral filter is similar to the K-means clustering and the mean shift algorithms in that it maintains a set of data points that are iteratively replaced by means. The filter's weight is based on a Gaussian distribution.

This study began with calculating the structure tensor and smooths it with a Gaussian filter. Local orientation and a measure for the anisotropic are then derived from the eigen-values and eigenvector of the smoothed structure tensor. Finally, a nonlinear filter performed the structure-aware smoothing. This nonlinear filter used weighting functions defined over an ellipse, whose shape was based on the local orientation and anisotropic. The filter response was defined as a weighed sum of the local averages, where more weight was given to those averages with low standard deviation. Then an experienced physician who was familiar with breast ultrasound interpretations manually determined contours of the tumor in the coronal plane image. Figure 3 shows the pre-processed images and the determined regions for the breast tumors with Grade I and III, respectively.



(a) (b)

(c) (d)

Fig. 3. The example of the image: (a)(b) the manually determined region of tumor images (Grade I tumor); (c)(d) the manually determined region of tumor images (Grade III tumor).

3.2 Feature Extraction

Textural and intensity statistical features were utilized to estimate the tumor grade. One of the simplest approaches for describing texture is to use statistical

moments of the intensity histogram of an image or region. Method which only utilized histograms in calculation would result in measures of texture that carry only information about distribution of intensities, but not about the relative position of pixels with respect to each other in that texture. The proposed method calculated a set of features related to the geometry of the boundary and the structure inside of the tumor. Moreover, texture is an important characteristics used to identify the objects from an image [14].

Textural features deal with the internal structure of the tumor which can be described using this co-occurrence matrix. Performing a statistical approach such as co-occurrence matrix would help to provide valuable information about the relative position of the neighboring pixels in an image. The gray value was calculated relationship between two relative positions of the pixels under different distances and different angles. The proposed method got a co-occurrence matrix from calculating the number of times occurrences with statistics. This pixel by pixel in the spatial relationship improved the disadvantage of grayscale intensity value charts has nothing to do with the space.

The co-occurrence method provides a second-order approach for generating texture features. Given a spatial window within the image, the co-occurrence method finds the conditional joint probabilities, C_{ij} of all pairwise combinations of gray

levels given the inter-pixel displacement vector (δ_x, δ_y), which represents the separation of the pixel pairs in the *x*- and *y*-directions respectively. The set of gray level co-occurring probabilities (GLCP) is defined as

$$C_{ij} = \frac{P_{ij}}{\sum_{i,j=0}^{G-1} P_{ij}} \quad (1)$$

where P_{ij} represents the frequency of occurrence between two gray levels, *i* and *j*, for a given displacement vector (δ_x, δ_y), for the specified window size. G is the number of quantized gray levels. Traditionally, the probabilities are stored in a gray level co-occurrence matrix (GLCM), where index (*i*, *j*) in the matrix represents the probability C_{ij}. Statistics are applied to the GLCM to generate texture features which are assigned to the center pixel of the image window.

The performed parameters were angular second moment (energy), inertia, correlation, the absolute value, the inverse difference, sum of entropy and the maximum probability. For quantitatively describing the intensity statistical features of the image, the useful features of the image were obtained from the intensity histogram. Statistical features used in this study included mean, sum variance, skewness, kurtosis, energy and entropy. This study found that the texture and statistical features are valuable to distinguish grade level from malignant lesions [15-16]. These features are defined as follows:

(1) *Contrast*:

This feature returns a measure of the power difference between a pixel and its neighbor over the total image. Description of two different grayscale values generated contrast in co-occurrence matrix. Contrast would be more strongly due to pixels changes more in the tumor tissue. The normal tissue changes lower, so the value of contrast is lower than the tumor tissue. The *Contrast* is defined as

$$\sum_{i,j} |i\text{-}j|^2 p(i,j). \tag{2}$$

(2) *Correlation*:

This feature returns a measure of how linked a pixel is to its neighbor over the whole image. *Correlation* is described a probe into the relation of *X*-axis and *Y*-axis in the co-occurrence matrix. The overall image pixels are similar, then the result values of correlation will greater. The *Correlation* is defined as

$$\sum_{i,j} \frac{\left(i\text{-}\mu_i\right)\left(j\text{-}\mu_j\right) p(i,j)}{\sigma_i \sigma_j}. \tag{3}$$

(3) *Energy*:

This feature shows the sum of squared elements in the GLCM. The *Energy* is defined as

$$\sum_i \sum_j p(i,j)^2. \tag{4}$$

(4) *Sum Entropy*:

This feature is a scalar value representing the irregularity of grayscale image that can be used to characterize the texture of the input image. *Entropy* is used to

assess the amount data of hidden information. The grayscale values of intensity change more, then the more information will be impliedly in the tumor region, so you can get larger entropy. The *Sum Entropy* is defined as

$$-\sum_i \sum_j p(i,j)\log(p(i,j)). \tag{5}$$

(5) *Maximum probability*:

The *Maximum probability* is defined as

$$\max_{i,j} p(i,j). \tag{6}$$

(6) *Sum variance*:

Variance is used to measure the number of image texture features and the average deviation volume. When the tumor imaging has larger change, it will have larger variances. The *Sum variance* is defined as

$$\sum_i \sum_j \left(i + \sum_i \sum_j p(i,j) \cdot log\left(p(i,j)\right)\right)^2 P_{i+j}(i). \tag{7}$$

(7) *Skewness*:

The *Skewness* is defined as

$$\sum_i \sum_j \left(i + \sum_i \sum_j p(i,j) \cdot log\left(p(i,j)\right)\right)^3 P_{i+j}(i). \tag{8}$$

(8) *Kurtosis*:

The *Kurtosis* is defined as

$$\sum_i \sum_j \left(i + \sum_i \sum_j p(i,j) \cdot log\left(p(i,j)\right)\right)^4 P_{i+j}(i) - 3. \tag{9}$$

(9) *Inertia*:

Inertia is measured relative position between the common pixel size of gray value in occurrence. When the image gray value differences, the relative position gray level value of the pixels are very minor difference. The *Inertia* is defined as

$$\sum_i \sum_j (i-j)^2 p(i,j). \quad (10)$$

(10) *Homogeneity*:

It returns a value that measures the closeness of the division of elements in the GLCM to the GLCM crosswise. Tumor tissue compared to normal tissue grayscale value changes should be relatively minor, so the homogeneous has high performance. The *Homogeneity* is defined as

$$\sum_{i,j} \frac{p(i,j)}{1+|i-j|} \ . \quad (11)$$

(11) *Angular Second Moment*:

The *Angular Second Moment* is defined as

$$\sum_i \sum_j p(i,j)^2. \quad (12)$$

(12) *Sum of Squares Variance*:

Variance squares are only a matrix X axis or Y-axis direction seeking the sum of square variance changes. If there were a matrix X axis distribution is more dense, the resulting value is larger, otherwise smaller. The *Sum of Squares Variance* is defined as

$$\sum_i \sum_j (i-\mu)^2 p(i,j). \quad (13)$$

(13) *Inverse Difference Moment*:

In general, the resulting value is greater when the inverse difference moment is not in big changing group of pixels. The *Inverse Difference Moment* is defined as

$$\sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j). \quad (14)$$

(14) *Sum average*:

Average is described the average of the probability value occurred in X-axis or Y-axis on the matrix. When the image grayscale values are concentrated on a certain value, the resulting of average value is larger.

$$\sum_i \sum_j i P_{i+j}(i,j). \quad (15)$$

In addition to the different levels of feature extraction, this study also evaluated features within certain distance from the defined ROI. As advised by the senior physician, there might already provide enough information to represent feature of textures in the range of 1-3 mm around the ROI. To this end, the representativeness of extracted features in the different ranges was examined as well. Four different ranges around ROI were extracted, i.e. full, 1mm, 2mm and 3mm around ROI.

3.3 Classification

The K-means clustering algorithm was first used by James MacQueen in 1967.

K-means clustering aims to partition the n observations into *k* sets ($k \leq n$) S = {S_1, S_2 ,……, S_k } so as to minimize the within-cluster sum of squares (WCSS): $\arg\min \sum_{i-1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$. Figure 4 illustrates the pseudo-code of K-means algorithm.

The K-means algorithm is an unsupervised classification method that calculates initial class means evenly distributed in the data space and then iteratively clusters the pixels into the nearest class using a minimum distance technique. The main idea is to define *k* centroids (one for each cluster) which are placed in a cunning way due to different location may cause different result. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. The early group is done when no point is pending. Then *k* new centroids are needed to re-calculate as barycenter of the clusters resulting from the previous step. After the method obtains these *k* new centroids, a new binding has to be done between the same data set points and the nearest new centroid. These steps are repeated until any reassignment of cases would make the clusters more internally variable or externally similar. This study utilized the K-means clustering algorithm to classify and estimate breast tumor in 3D sonography to a corresponding grade. Take note that the output value of the K-means is either 0 or 1. When the output value of a suspicious tumor region is 1, the system would classify the tumor in the ultrasound image as Grade I. Conversely, when the

output value is 0, the tumor would be diagnosed as Grade III [17-18].

```
K-means(k, x, clusters)
Randomly assign the mean values m1, m2,..., mk
    do until there are no changes in m1, m2,..., mk
        for  i ← 1 to n
            do if xi is most close to mj
                then clusters[i] ← j
                    Replace mj with the recalculated mean value in cluster j
        end
    end
return clusters
```

Fig. 4. Pseudo-code of K-means clustering algorithm

CHAPTER 4

RESULTS

4.1 Grade Estimation

In all 189 obtained cases, there were 69 and 120 cases identified as Grade I and III, respectively. To evaluate how well the indices are correlated to histological grades, both Pearson's and Spearman's correlation coefficient were calculated. Student's t distribution was tested for the significance. Also, bagging decision tree was used to classify the indices to histological grades. Performance of the classifier is evaluated by receiver operating characteristic (ROC) curve and the value of area under curve (AUC). Table 2 shows the classification experiment, accuracy is 84.66% (68/80), the sensitivity is 78.26% (35/40), the specificity is 88.33% (33/40), the positive predictive value is 79.41% (35/42) and the negative predictive value is 87.60% (33/38).

Table 3 shows the final feature cluster centrality parameter results of grade model with 189 slices observations. The initial cluster centers are the variable values of the *k* well-spaced observations. The final cluster centers are computed as the mean for each variable within each final cluster. The final cluster centers reflect the characteristics of the typical case for each cluster. According to each feature classify two different groups of the grade systems, each group representing different scores.

We can determine the new sample data for which group by the final cluster center's characteristic parameters. For example, we can calculate Euclidean distance between two group's center to classified Grade I or Grade III.

Table 2. Classification of breast tumor grade by proposed K-means system

	Grade I		Grade III	
K-means Output = 1	TP	54	FP	15
K-means Output = 0	FN	14	TN	106
Total	68		121	

TN = true-negative; FN = false-negative;
FP = false-positive; TP = true-positive;
Accuracy = (TP+TN) / (TP+TN+FP+FN);
Sensitivity = TP / (TP+FN);
Specificity = TN / (TN+FP);
PPV = TP / (TP+FP);
NPV = TN / (TN+FN).

From the experimental results, the proposed method using the K-means algorithm with the textural and statistical features is feasible for coronal plane imaging classification. The results indicated that the proposed features can be utilized as potential predictors for tumor grades. As the proposed classification system is useful to estimate the breast tumor grade which may be the useful information for physicians to prognoses the effect of treatments for malignant lesions.

Table 3. The final cluster centers of tumor grade classification

	Cluster	
Feature	Grade I	Grade III
Mean	.030	.034
Inverse	.038	.054
Skewness	.023	.464
Kurtosis	2.617	2.911
Energy	6.754	6.558
Entropy	1.565	3.394
Mean (2*mm*)	.023	.018
Standard (2*mm*)	.056	.051
Contrast	.031	.040
Correlation	.956	.961
Energy	.716	.522
Homogeneity	.993	.988

Although histological grade of breast cancer has been recognized for a long period of time, and its prognostic value has been validated in multiple independent studies, there are still some concerns regarding the incorporation of grade into routine breast cancer staging systems. In this study, we assessed the prognostic significance of histological grade in a series of operable breast cancer graded using the Nottingham histological grading system, uniformly treated in a single institution.

4.1 TNM Estimation

To gain a general understanding of how features work, full feature set was used in classification models of the three experiments. Table 4 shows the results of classification in TNM model with 573 slices of observations from different feature selecting procedures.

Table 4. Classification of breast tumor TNM by proposed K-means system

	Tubule	Nuclear	Mitosis
Accuracy	78.87%	81.98%	79.04%

Table 5-7 show the final feature cluster centrality parameter results of TNM model with 191 observations. The tumor itself is not marked, different ranges of ROI perimeter marked with brackets at the rear features. The cluster 1, 2 and 3 represents a fraction of TNM system’s score levels. From the table we can know the characteristics of each group coefficient. The new sample data for which group was determined by the final cluster center's characteristic parameters. The relationship between each feature and scores would be known, for example entropy score 1 is greater than score 3 in Table 5.

Table 5. The final cluster centers of tubule classification

Feature	Cluster 1	Cluster 2	Cluster 3
Contrast	.108	.216	.164
Entropy	7.522	3.914	2.529
Sum of Squares Variance	.215	.081	.062
Inverse Difference Moment	5.986	15.457	8.759
Angular Second Moment	8.333	4.571	2.782
Homogeneity	.993	.986	.989
Kurtosis	2.617	3.991	2.661

Table 6. The final cluster centers of nuclear classification

Feature	Cluster 1	Cluster 2	Cluster 3
Angular Second Moment	.741	.790	.657
Standard	.0775	.081	.097
Kurtosis	4.839	5.050	5.772
Kurtosis (1*mm*)	8.610	42.025	123.564
Sum average (1*mm*)	.035	.073	.039
Entropy	.098	.198	.205
Inertia	.005	.033	.017
Mean	.352	.133	.244
Contrast	.105	.192	.226
Skew (2*mm*)	4.769	3.704	5.021
Correlation (2*mm*)	.958	.956	.958
Intertia (2*mm*)	.007	.027	.018
Inverse Difference Moment (2*mm*)	.154	.086	.062

Table 7. The final cluster centers of mitosis classification

Feature	Cluster 1	Cluster 2	Cluster 3
Skew	1.608	1.297	1.518
Kurtosis (1*mm*)	5.504	4.331	5.144
Entropy	9.353	3.373	3.720
Angular Second Moment	.094	.024	.037
Inertia	.276	.063	.099
Homogeneity (1*mm*)	.991	.989	.993
Mean (1*mm*)	.303	.187	.384
Sum average (1*mm*)	.064	.174	.111
Sum average (2*mm*)	.083	.023	.034
Inertia (2*mm*)	.303	.081	.121
Energy	.632	.664	.611
Kurtosis	3.991	2.955	2.285

CHAPTER 5

CONCLUSION AND DISCUSSION

This study proposed a system to differentiate between tumors with Grade I and Grade III. This study reduced to any amount of noises in breast tumor, that preprocessed by using a bilateral filter. It is an ideal enhancing method for sonogram with both efficiency in processing time and effectiveness in noise reduction. From the experimental results, the proposed method using the K-means algorithm with the proposed textural and statistical features is possible for coronal plane image classification. The results indicated that the proposed features can be utilized as potential predictors for tumor grades. As the proposed classification system is useful to estimate the breast tumor grade which maybe the useful information for physicians to propose the effect of treatments for malignant lesions. Physicians utilize tumor grade and other factors, such as cancer stage and a patient's age and general health, to develop a treatment plan and to determine a patient's prognosis (the likely outcome or course of a disease; the chance of recovery or recurrence). Generally, a lower grade indicates a better prognosis. A higher-grade cancer may grow and spread more quickly and may require immediate or more aggressive treatment.

Grading systems differ depending on the type of cancer. In Grade I tumors, the tumor cells and the organization of the tumor tissue appear close to normal. These

tumors tend to grow and spread slowly. In contrast, the cells and tissue of Grade III tumors do not look like normal cells and tissue. Grade III tumors tend to grow rapidly and spread faster than tumors with a lower grade. We establish the texture features model of grade and TNM systems. The performances of the three-one model exhibited the capabilities to identify tumor grades. Combination of these two classification models was not sufficient yet to form a complete system, though. The motive for classifying in two models had the origin from bias quantities in tumor grades, where the amount of Grade II cases was about four times larger than each of the other grades; they were the main causes of the false negatives and false positives in both models. Tumor grades are not easily classified. In consequence of this observation, other classification models might be worth of trying. We hope the system can classify the score of tubule, nuclear and mitosis. The proposed method can score by TNM classification to determine the relationship between their grade system.

Our results indicated that the proposed indices of tumor feature had mighty potential to predict tumor grades. Although the histological grade has been extensively studied and is a widely accepted index to the malignancy of breast cancer, it is certainly not the only one in the breast cancer research community. We are aware of the existence of other prognosis factors as well; for example, subgroup of the gene-expression profiles. It might be really helpful if a classification model can be

trained by the response of both histological grade and the tumor subgroup [19-20].

References

[1] Rotten, D., J. M. Levaillant, and L. Zerat. "Analysis of normal breast tissue and of solid breast masses using three-dimensional ultrasound mammography. "*Ultrasound in obstetrics & gynecology* 14.2 (1999): 114-124.

[2] American Cancer Society. Cancer Facts & Figures 2013. Atlanta, Ga: American Cancer Society; 2013

[3] Chou, Y. H., Tiu, C. M., Hung, G. S., Wu, S. C., Chang, T. Y., & Chiang, H. K. (2001). Stepwise logistic regression analysis of tumor contour features for breast ultrasound diagnosis. *Ultrasound in medicine & biology*, *27*(11), 1493.

[4] Berg, W. A., Gutierrez, L., NessAiver, M. S., Carter, W. B., Bhargavan, M., Lewis, R. S., & Ioffe, O. B. (2004). Diagnostic Accuracy of Mammography, Clinical Examination, US, and MR Imaging in Preoperative Assessment of Breast Cancer1. *Radiology*, *233*(3), 830-849.

[5] M.C. De Nunzio, A.J. Evans, S.E. Pinder, I. Davidson, A.R.M. Wilson, L.J. Yeoman, C.W. Elston, and I.O. Ellis, "Correlations between the mammographic features of screen detected invasive breast cancer and pathological prognostic factors," *The Breast*, vol. 6, no. 3, pp. 146-149, 1997.

[6] M.C. Alexander, B.C. Yankaskas, and K.W. Biesemier, "Association of stellate mammographic pattern with survival in small invasive breast tumors," *American Journal of Roentgenology*, vol. 187, no. 1, pp. 29-37, 2006.

[7] Petrou, M., & Sevilla, P. G. *Image processing: dealing with texture*. Vol. 10. Chichester: Wiley, 2006.

[8] J.M. Ko, M.J. Nicholas, J.B. Mendel, and P.J. Slanetz, "Prospective assessment of computer-aided detection in interpretation of screening mammography," *American Journal of Roentgenology*, vol. 187, no. 6, pp. 1483-1491, 2006.

[9] D. Rotten, J.M. Levaillant, and L. Zerat, "Analysis of normal breast tissue and of solid breast masses using three-dimensional ultrasound mammography," *Ultrasound in Obstetrics and Gynecology*, vol. 14, no. 2, pp. 114-124, 1999.

[10] S.F. Huang, R.F. Chang, D.R. Chen, and W.K. Moon, "Characterization of spiculation on ultrasound lesions," *Medical Imaging, IEEE Transactions on*, vol. 23, no. 1, pp. 111-121, 2004.

[11] C.W. Elston and I.O. Ellis, "Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up," *Histoplane*, vol. 19, pp. 403-410, 1991.

[12] Tomasi, C., and Manduchi, R. "Bilateral filtering for gray and color images."

Computer Vision, 1998. Sixth International Conference on. IEEE, 1998.

[13] Soh, L. K., and Tsatsoulis, C. "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices." *Geoscience and Remote Sensing, IEEE Transactions on* 37.2 (1999): 780-795.

[14] Haralick, R. M., Shanmugam, K.,and Dinstein, I. H. "Textural features for image classification." *Systems, Man and Cybernetics, IEEE Transactions on* 6 (1973): 610-621.

[15] Cheng, H. D., Shan, J., Ju, W., Guo, Y., & Zhang, L. "Automated breast cancer detection and classification using ultrasound images: A survey." *Pattern Recognition* 43.1 (2010): 299-317.

[16] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. Burlington, MA, USA: Academic Press, 2008.

[17] J. MacQueen, "Some methods for classification and analysis of multivariate observations." *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. No. 281-297. 1967.

[18] Hartigan, J. A., & Wong, M. A. "Algorithm AS 136: A k-means clustering algorithm." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28.1 (1979): 100-108.

[19] C. Lin, S.Y. Chien, L.S. Chen, S.J. Kuo, T.W. Chang and D.R. Chen, "Triple negative breast carcinoma is a prognostic factor in Taiwanese women," *BMC Cancer*, vol. 9, no. 1, p. 192, 2009.

[20] L. Carey, E. Winer, G. Viale, D. Cameron, and L. Gianni, "Triple-negative breast cancer: disease entity or title of convenience?," *Nature Reviews Clinical Oncology*, vol. 7, no. 12, pp. 683-692, 2010.