Classification of Emphysema Patterns in Computed Tomography Based On Gabor Filter

by

Tengku Mohd Syamim Bin Tengku Azis

16788

Dissertation submitted in partial fulfilment of

the requirements for the

Bachelor of Engineering (Hons)

(Electrical and Electronic)

MAY 2015

Universiti Teknologi PETRONAS

32610 Bandar Seri Iskandar

Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Classification of Emphysema Patterns in Computed Tomography based on Gabor Filter

by

Tengku Mohd Syamim Bin Tengku Azis

16788

A project dissertation submitted to the

Department of Electrical and Electronics Engineering

Universiti Teknologi PETRONAS

In partial fulfillment of the requirement for

Bachelor of Engineering (Hons)

(Electrical and Electronic)

Approved by,

_____

DR. NORASHIKIN BINTI YAHYA

PROJECT SUPERVISOR

UNIVERSITI TEKNOLOGI PETRONAS

BANDAR SERI ISKANDAR, PERAK

MAY 2015

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

_____

TENGKU MOHD SYAMIM BIN TENGKU AZIS

ABSTRACT

Emphysema is a type of chronic obstructive pulmonary disease (COPD) affecting millions of people worldwide. Patients with emphysema typically have breathing difficulty. Early detection using Computed Tomography (CT) scan image can save many of the emphysema patients life. Furthermore, it helps the medical practitioners in planning suitable treatments for patients. The CT scan of human lungs are commonly taken from 3 different directions; center, bottom and top. The images obtained from different slices are then used by radiologist to identify normal or abnormal tissues. Computer-aided diagnosis (CAD) has becomes part of routine clinical work for assisting radiologist in detection of abnormal tissue in many screening sites and hospitals. One of the main processing technique in CAD is texture classification and analysis. In this research, a Gabor-based emphysema classification algorithm is proposed. Gabor filter offer the advantage of multi-resolution and multi-orientation properties and is optimal for measuring local spatial frequencies. In essence, the Gabor transform is performed by applying Gaussian masks prior to the discrete wavelet transform. The extracted feature from the Gabor filter is in the form of local energy calculated at different scale and orientation. The proposed emphysema classification algorithm involves four aspects, image pre-processing, feature extraction, matching (classification), and decision making. In the classification stage, the k-NN classifier is used to classify the CT images to two different classes which are Normal Tissue (NT) and Abnormal Tissue; Centrilobular Emphysema (CLE) and Paraseptal Emphysema (PSE). The proposed algorithm is evaluated using k-fold cross validation technique and its performance is shown to produce low misclassification rate of 0.01%.

ACKNOWLEDGEMENT

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

COPD	Chronic obstructive pulmonary disease
CT	Computed Tomography
NT	Normal Tissue
CLE	Centrilobular Emphysema
PSE	Paraseptal Emphysema
PLE	Panlobular Emphysema
FT	Fourier Transform
IFT	Inverse Fourier Transform
k-NN	k-Nearest Neighbor
FDR	Fisher's Discriminant Ratio

CHAPTER 1: INTRODUCTION

In this chapter, we first of all present the brief information about emphysema and its different classes. Secondly, details related to the project are presented, namely the problem statement, objectives, scope of study, relevancy of the research topic and the feasibility of the project.

1.1 Background

Emphysema is a disease affecting human respiratory system. It is one of the Chronic obstructive pulmonary disease (COPD) and resulted in limitation of airflow to lungs [2]. Nowadays, COPDs become a serious cause of death and health problem in the world. Based on the latest statistic, this disease is the fourth leading cause of death in the United States and by the time of year 2020, it will rank as the fifth burdening disease [3]. This disease causing the elastic quality of the tissue is reduced where pitting in pressure is the factor. Due to this problem, it is difficult for air to pass through the air paths. This problem will cause expansion to the lungs which will lie in front of the heart and will damaged the area that has been covered by lungs [1]. In addition, the number of air vesicles are decrease but the size are increase. In this case the blood vessels become narrowed although they are not fully damaged.

In the other research, T. Purushotham et al. [4] said that the tiny air sacs (alveoli) located at the end of the air path is damaged and reach deeply into the lungs. The air is being let out of the air sacs through the airways due to the difficulty of air to reach it. This problems cause the air sacs damaged and dysfunction where the process of exchanging oxygen for carbon dioxide become less efficient. In other words, human will have some difficulties in breathing. As mentioned earlier, Emphysema has three different classes or subtypes or patterns which is CLE, PLE, and PSE. Each of these classes has different definition and effects. This will be explained later in this research paper.

1.1.1   Centrilobular Emphysema (CLE)

Centrilobular Emphysema is also known as Centriacinar Emhysema where it is the most common type of pulmonary emphysema [2]. Before it started to spread, it begins in the respiratory bronchiole. This subtypes of emphysema is the lungs that have thin, small, multiple and spotty lucencies of walls. In the region where the proximal respiratory bronchiole at the center of the secondary pulmonary lobule, the process of this type of disease begins. Figure 1 shows the illustration of the Centrilobular Emphysema.

There are two types of changes which are early and late changes. For the early changes, the whole round of the centrilobular artery may have the presence of evenly distributed holes which seen as small round black in the central portion of the secondary pulmonary nodule. This result show that the attenuating is low compared to others. As for the late changes, this changes cause by the early changes where it become attached to each other. Due to this changes, the border between the infected area and the normal lungs can be observed clearly. Other than that, compression occurred around the lungs parenchyma as the enlargement of the dilated airspace. The quality of blood flow and the exchange of oxygen will be reduced.
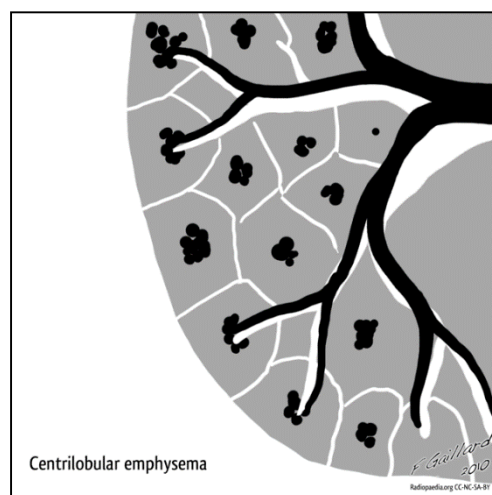
Figure 1: Illustration of Centrilobular Emphysema [23]

1.1.2 Paraseptal Emphysema (PSE)

This type of emphysema is lucencies in a single layer, multiple with visible thin walls. This type of emphysema also have different name as per its effect. By characterized by damaged of distal airway structures, alveolar ducts, and alveolar sacs, it also known as distal acinar airway emphysema [2]. Kathryn et al. [7] said that below the pleural surface involving the distal-most alveoli where it change its state to morphologic pattern. The other types of emphysema may exist with this type of emphysema. This emphysema typically occurs at the upper lungs along the dorsal surface where it tend to be limited in magnitude.
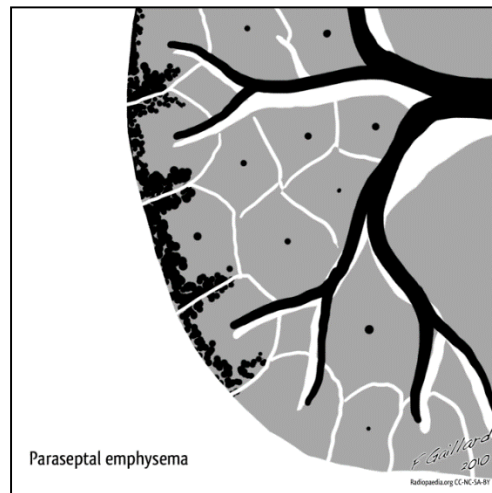
Figure 2: Illustration of Paraseptal Emphysema [23]

1.1.3 Panlobular Emphysema (PLE)

Panlobular Emphysema will cause the lungs to be lucent and have a small size pulmonary vessels. This type of emphysema is the highest risk of damaging of the lungs whereas Takahashi et al. [2] said from alveoli to lower half of the lungs can be destroyed uniformly by this type of emphysema.

Panlobular Emphysema or Panacinar Emphysema can be defined as permanent damaged of the alveoli which is the airspaces and respiratory bronchioles. Figure 3 shows the illustration of the Panlobular Emphysema

3

where the surface of the lungs looks darker than the others with high attenuation.

Figure 3: Illustration of Panlobular Emphysema [23]

1.2 Problem Statement

The CT scan of lungs carries important information that can be critical in diagnosis of emphysema. For a given image, a radiologist has to analyze and evaluate comprehensively in a short time.

Automatic detection and quantification of emphysema is commonly used to assist radiologist in CT images and thereby reduce the time consuming analysis of the emphysema.

1.3 Objective and Scope of Study

The proposed research objectives and scope of study are outlined in this section the objectives of this research are:

1. To develop Gabor filter-based emphysema classification for lungs CT scan.
2. To evaluate the accuracy of the proposed method/algorithm.

This research work will not be completed without covering 3 major topics; emphysema and its classification techniques, Gabor filter, and theories related to

pattern recognition. First of all, the formulation of the work was based on the issues related to classification of emphysema in lungs CT scan. The emphysema, which is a type of COPD disease, can be grouped into three classes, CLE, PLE and PSE. Several proposed classification algorithms are also reviewed in order to assess the suitability of a Gabor-based technique. Secondly, theories related to the Gabor filter are also reviewed for better understanding of its basic operation. Lastly, come concepts related to pattern recognitions such as matching technique, classification technique and validation technique are covered which is important in validating the proposed technique.

1.4 Relevancy of the Project

This project involves an image processing which is included in Data Signal Processing, core subject in 4 years of undergraduate study in Electrical and Electronic Engineering. Technical knowledge in MATLAB programming is very important in this project. Other than that, time management and soft skills is very crucial in completing this project where planning and presentation to evaluators are required. This research project is focusing on classification emphysema patterns therefore research on the disease need to be conducted.

1.5 Feasibility of the Project

This work is to be completed in 2 semesters. In the first semester, the focus is more on problem formulation, literature review and finalizing the methodology. In the second semester, the work focuses on the development the algorithm, validation, testing and evaluation.

CHAPTER 2: LITERATURE REVIEW

In this chapter, we first review some of the published emphysema classification techniques. Secondly, the fundamental of Gabor filter, including its mathematical foundations are presented. The Gabor filter are to be used for extracting feature from the lungs CT scan. In essence, the feature extraction by Gabor filter involves multiplying the Fourier transform (FT) on an image with Gaussian masks at different scales and orientations. Then, these products (at i-scale and j-orientation) are inverse Fourier transformed (IFT). The feature of the image is expressed as energy values (at i-scale and j-orientation) of the IFT image in the previous step.

2.1 Emphysema

The most common technique that have been used before this is the density mask. This technique is designated to read the amount of air that are in a Computed Tomography (CT) image and the percentages of pixel values can be computed which is smaller than a previous selected values [8]. Each of the attenuation in the CT images has its own values. The attenuation is measured in Hounsfield unit (HU) scale. This HU scale is used to detect pathological changes in lungs where it is linked to physical density of lungs tissue [6]. For each every person have different attenuation values who had emphysema compared to a healthy person with normal tissue (NT). There are some technique to quantify the emphysema attenuation values.

Normally, -910 HU is the value which it is taken as standard threshold. In the other research where Madani et al. [9] and Litmanovich et al. [11] said that the HU correspondence with emphysema is larger where the threshold values are between -960 to -980. In other perspective where Emam et al. [10] said that areas of the lungs have reading values lower than the threshold which is -910 HU are considered to be emphysematous. These values shows that there is an abnormal activities where the air spaces increase in space. As for conclusion of this density mask technique, it is a good method to classify each class of emphysema but the

disadvantages is when there is calibration and noise [8]. The averaging effect will cause problem to it during differentiating the patterns of emphysema.

The other method is called the texture analysis. It is used to analyze the texture pattern in lungs CT images. Nava et al. [8] said that this technique is can identify the changes occur in the lungs parenchyma and any status change related to emphysema. On the other hand, a combined textural features technique is used in some research that give out an effective result to distinguish the emphysema patterns. By using local binary patterns (LBPs) to classify each class of emphysema, Sørensen et al. [12] has come out with 95.2% classification accuracy with the set of 168 based on the three classes of emphysema which are normal tissue, centrilobular emphysema, and paraseptal emphysema.

Costa et al. [13] proposed an efficient texture analysis extraction technique and it is called Segmentation-based Fractal Texture Analysis (SFTA). In his research, the input image is decomposed into one set of binary images. In order to define segmented texture patterns, the set of binary images is computed. This SFTA used to classify 3258 emphysema patches where the size are 64 x 64 pixels.

Another alternatives is used for emphysema classification is density estimation of local histograms. In [14] kernel density estimation technique was used to deal with the inherent deficiency of the local intensity histograms retrieved from regions of interest (ROI). The technique is evaluated using leave-one-subject-out classification experiments and full lungs analysis.

In this approach, a technique is proposed called Gabor filter for the emphysema classification. The Gabor filter will act as a feature extraction method for assessment of emphysema in CT images. This approach will use the Gabor transform by applying the Gaussian masks prior to the discrete transform. The extracted feature from the Gabor filter will be in the form of local energy calculated at every scale and orientation. The methodology section is included to explain more about the technique has been used in this research.

2.2 Gabor Filter

Gabor function was introduced by Hungarian born electrical engineer Dennis Gabor in 1946 [15]. Features extraction mostly in texture-based image analysis and face recognition are frequently used the Gabor functions for segmentation and classification. The shape of the used wavelet can determined the information of the images and it also can be translated into words of a wavelet transform [16]. A spectral energy density that focused around the certain frequency and defined position are provided by two-dimensional Gabor transform and correlation of images.

Basically, the use of the Gabor filter in image processing is the point detection [16]. Approaches which very common that people use this Gabor filter or transform is the edge detection from the image feature. The other approaches is combine several filters with responses at different orientation which is called corner detection.

Gabor function is consist of two elements which are Gaussian masks and the Fourier transform (FT). FT is a common tool that been used to analyze frequency properties of a signal. By characterizing by sine and cosines functions, the signal is breaks into different representation. As for the Gaussian masks, it is a kind of FT but in different function. It is an eigenfunction of the FT. Gaussian masks or filter changes the signal by using convolution where it keep the signal to have a step function input with no overshoot and minimize the fall and rise time.

For 1D case, a complex exponential with local value of x = 0 of the Gabor function. Bařina et al. [16] said when dilation is fixed form kernel of a Gabor transform, there is a shift of the Gaussian masks and dilation of complex exponential function. For this research, this experiment will be only focused more on the 2D case.

Nava et al. [8] stated that the very minimum volume in the time-frequency space is occupied with the 2D Gabor functions which are the band-pass filter. The lowest bound of the ambiguity principle represented by the volume, the (x,y) and (u,v) is related to spatial and frequency variables: $(\Delta x)(\Delta y)(\Delta u)(\Delta v) \geq \frac{1}{16\pi^2}$. The definition of Gabor functions is the combination of Gaussian function and complex sinusoid and produce non-orthogonal basis set. The 2D Gabor function in spatial domain is stated as below:

$$g(x,y) = Ke^{-\frac{1}{2}\left[\frac{(x-x_0)^2+\gamma^2(y-y_0)^2}{\alpha^2}\right]+i(2\pi[u_0(x-x_0)+v_0(y-y_0)]+\emptyset)} \qquad (1)$$

Where $K = \frac{\gamma}{2\pi\alpha^2}$ is a normalizing constant, (x_0,y_0) = center of the function, (u_0,v_0) = central frequency and $\emptyset$ = phase of sinusoidal signal. The (γ, α) in the normalizing constant give function to control the bandwidth of the Gaussian wavelet throughout the X-axis and Y-axis. Function above can be divided into two parts after done the expansion which are function that responds only with a maximum in zero, even-symmetry function and odd-symmetry function. These two types of functions is used to detect the salient edge and responds to zero-crossing respectively [17]. By applying the Euler's formula to the previous equation, the equation will be:

$$g(x,y) = Ke^{-\frac{1}{2}(x^2+\gamma^2y^2)}[\cos(2\pi u_0 x) + i\sin(2\pi u_0 x)] \qquad (2)$$
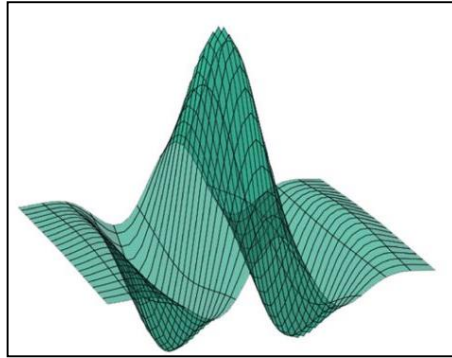
Figure 4: Even-symmetry function (Real part) [8]
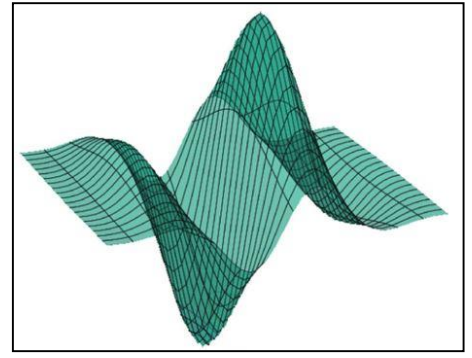
Figure 5: Odd-symmetry function (Imaginary part) [8]

The equation is $g(x,y) = g_e(x,y) + ig_o(x,y)$ where:

$$g_e(x,y) = Ke^{-\frac{1}{2}(\frac{x^2+\gamma^2y^2}{\alpha^2})}\cos(2\pi u_0 x) \quad \text{[Even-symmetry function]} \qquad (3)$$

$$g_o(x,y) = Ke^{-\frac{1}{2}(\frac{x^2+\gamma^2y^2}{\alpha^2})}\sin(2\pi u_0 x) \quad \text{[Odd-symmetry function]} \qquad (4)$$

Nava et al. [8] stated that in order to select the orientation and frequency response of Gabor functions, the functions are filtered in the Fourier domain. For the g_e, its FT is $\mathbb{F}\{g_e\} = \frac{1}{2}[\hat{G}(u,v) + \hat{G}(-u,-v)]$ whereas for g_o, its FT is $\mathbb{F}\{g_o\} = \frac{1}{2}[-i\hat{G}(u,v) + i\hat{G}(-u,-v)]$. The g_e is a symmetric function about the Y-axis and g_o is a symmetric function around the origin.

Gabor filter also has been used in many researches and applications. One of them is the classification of tumor images in brain. It is used to extract the pattern features of the magnetic resonance (MR) tumor images to distinguish between primary central nervous system lymphoma (PCNSL) and glioblastoma multiforme (GBM) [21]. Extracting the specific features on the contrast-enhanced TI-weighted MR images require the Gabor filter transform with eight orientations and several frequencies.

Other than that, Sara et al. [5] said in order to extract global and local features of an image, Gabor filter need to be applied on entire image and non-overlapping sub-images with equivalent size. Then she used the k-NN, Support Vector Machine (SVM) and ORL database to get the face recognition rate. Meanwhile, Kishore et al. [22] used different method but same technique where he used the Gabor-Feature-based DFT Shifting (GFDS) and Skin-detection-based Background Removal. The function of GFDS is to neutralize the image variations for example scale, pose and location whereas the function of Skin-detection-based Background Removal is to learn the shape of the face and the camera distance in order to remove the complex background.

CHAPTER 3: METHODOLOGY

In this chapter, we present the development of the algorithm involving image preprocessing, feature extraction using Gabor filter or classical features, and matching technique using k-NN classifier. The evaluation of the algorithm uses *k*-fold cross validation technique whereby the performance is measured in terms of misclassification rate. We also present the Fisher discriminant ratio, which is a test for selecting the most discriminative features for the purpose of reducing the number of features. Reducing the number of features offers the advantage of lowering the computational time of running the algorithm.

3.1 Proposed Method

Features and patterns extraction of the emphysema images is an important technique in identifying the different class of emphysema. Gabor filter will be used in this experiment to extract the emphysema patterns and features from lungs CT scan images. Small differences between each classes of the emphysema can be detected by the different directions and frequencies of Gabor filter.

A general block diagram summarizing the proposed algorithm is illustrate in Figure 6. In the first stage, the regions of interest (ROI) is selected from the input lungs CT scan. Next is the Gabor filer is used to extract the specific features of the emphysema sample images. Finally the classification stage, the k-NN classifier is used to identify each classes of the emphysema. The trained database is inserted during the classification stage to compare the features and patterns of the database with the test set.

```
Lungs CT scan  →  Pre-processing  →  Feature Extraction  →  k-NN classifier  →  Decide
                                                                    ↑
                                                            Trained Database
```
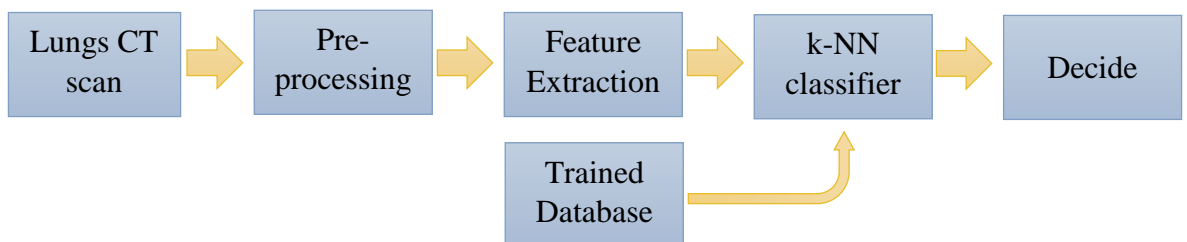
Figure 6: The proposed method

3.2 Dataset

In this research, a dataset by the Prof. Dr. Bruijne and Dr. Sørensen will be used which is the Bruijne and Sørensen (BS) dataset and MATLAB. The dataset consists size of 61 x 61 pixels of 168 patches. It is interpreted in 25 subjects which divided into three group healthy non-smokers, smokers without COPD and smokers with average or severe COPD. These patterns have different number in patches and subject for each group, NT (59 patches from 8 subjects), CLE (50 patches from 7 subjects) and PSE (59 patches from 10 subjects). The NT represents the healthy non-smokers whereas the CLE and PSE patches represents both smokers with or without COPD.

3.3 Feature Extraction

3.3.1   Gabor Filter

A rotated Gaussian function with certain angle, θ and frequency units, u_0 shifted along the X-axis is used if both of the odd and real parts are used which represents as below:

$$\hat{G}(u, v) = e^{-2\pi^2\alpha^2\left[(\tilde{u}-u_0cos\theta)^2+\frac{1}{2}(\tilde{v}+u_0sin\theta)^2\right]} \tag{5}$$

Where $(\tilde{u}, \tilde{v}) = (ucos\theta, -usin\theta)$. The condition where about one octave apart is the simple receptive fields of frequency bandwidths. Filtering of an image should consider the radial bandwidth formula where each filter having half-amplitude bandwidth, B_u linked to its central frequency by using the formula below:

$$\alpha = \frac{\sqrt{\log(2)}(2^{Bu}+1)}{\sqrt{2}\pi u_0(2^{Bu}-1)} \tag{6}$$

There are some disadvantages using this type of filtering where small number of central frequency will cause the filter act as low-pass filters instead of band-pass filters. Due to this some information of the image will be loss. Moreover, Nava et al. [18] said if beyond its upper limit, $u \le \frac{1}{2}$, the value of

α will be large where it will affect the filtering (artifacts). In determining the best angular bandwidth, B$_\theta$, the formula below is used:

$$\frac{\alpha}{\gamma} = \frac{\sqrt{\log(2)}}{\sqrt{2}\pi u_0 \tan\left(\frac{B_\theta}{2}\right)} \tag{7}$$

Where the axisymmetric filters is considered and γ is set to value 1. By using the formula, eight orientations of filter bank can be obtained where B$_\theta = \frac{\pi}{8}$.
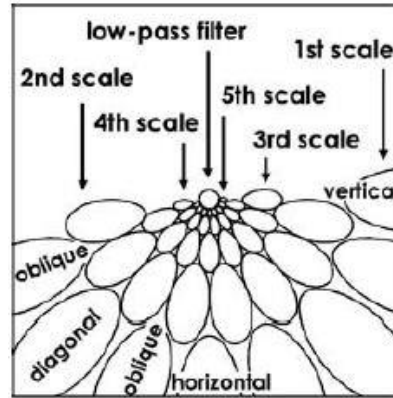
Figure 7: Gaussian mask of 5-scale and 8-orientation Gabor filter [24]

In the feature extraction of an image, similar patterns and characteristics need to be identified which are common to a specific class. Characteristics in specific frequency bands and orientations can extracted by the Gabor filter by using complex Gabor coefficients as shown below:

$$G_{(s,\theta)}(x,y) = \sqrt{E^2_{(s,\theta)}(x,y) + O^2_{(s,\theta)}(x,y)} \tag{8}$$

Where, $E_{(s,\theta)} = I * g_{e(s,\theta)}$ and

$O_{(s,\theta)} = I * g_{o(s,\theta)}$

The "$*$" symbol shows the convolution of the image, I. The $g_{e(s,\theta)}$ and $g_{o(s,\theta)}$ indicates the even-symmetry and odd-symmetry filters at the scale, s and orientation, θ.
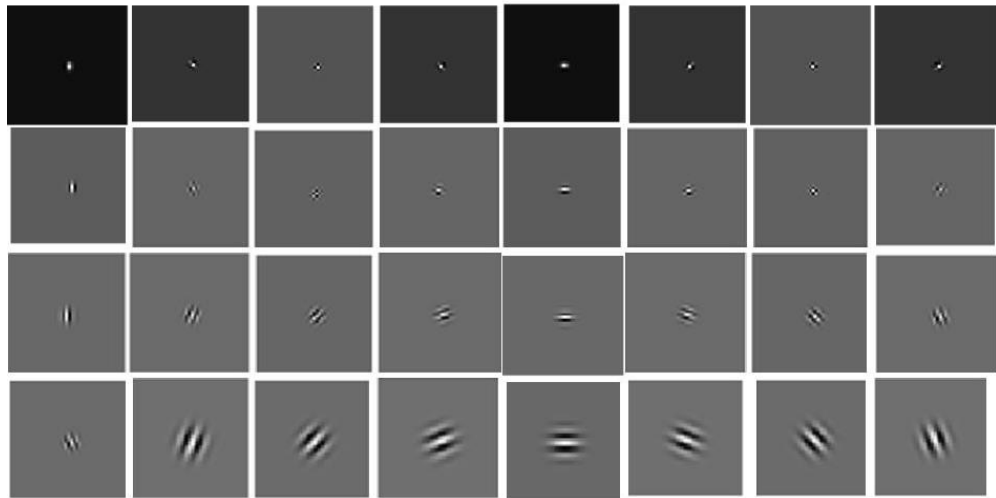
Figure 8: Gabor filter with different scales (from top to bottom) and eight directions [21]

Figure 9 shows the illustration of the Gabor filter technique.
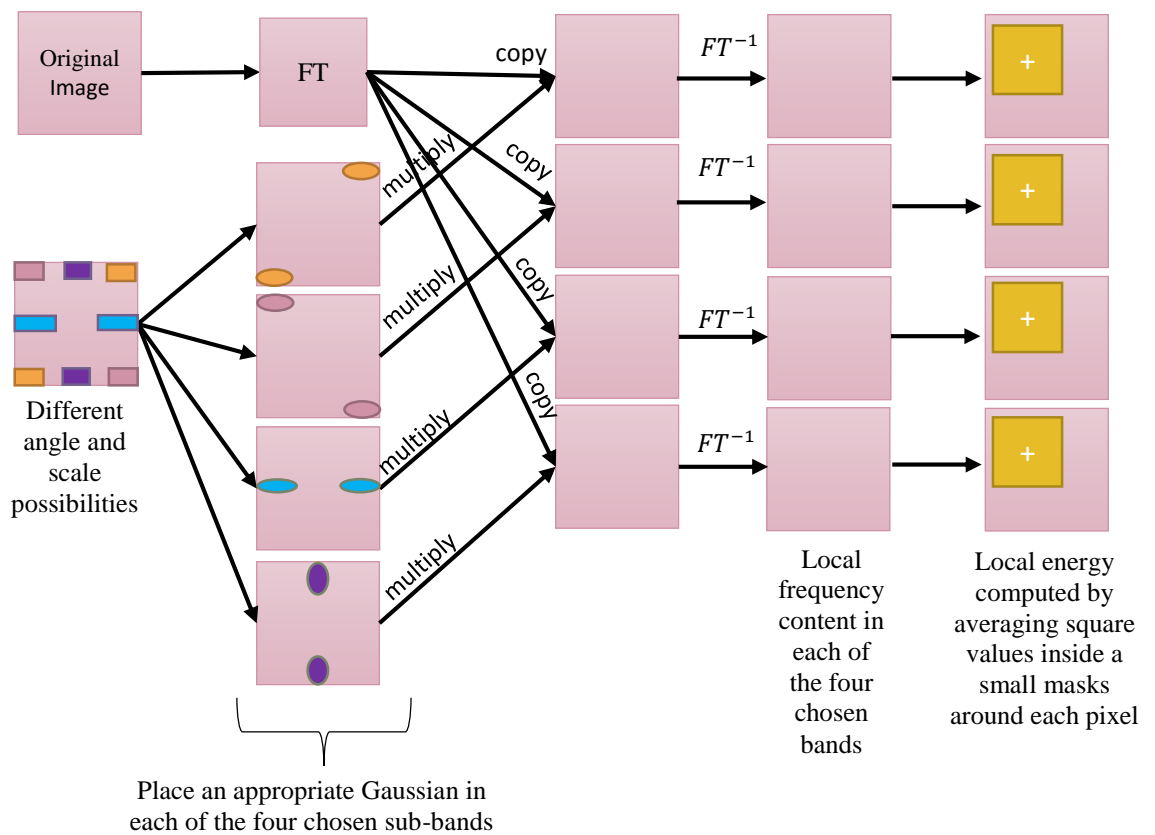
Figure 9: Feature extraction using Gabor filter

Before the energy plot is generated, the patches/sample images has to undergo some stages of filter such as Fourier transform (FT) and Inverse Fourier Transform (IFT) after multiply the images with Gaussian masks function. As mentioned before, the local energy is computed by averaging square values inside a small masks around each pixel.

(a)   (b)
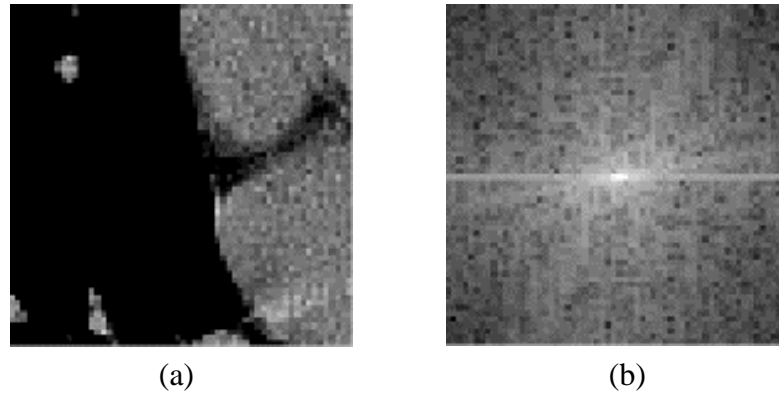
Figure 10: (a) A Paraseptal Emphysema (PSE) patches/sample images to be filtered and (b) the magnitude of its Fourier transform

0°   45°   90°   135°

(a)

0°   45°   90°   135°

(b)

0°   45°   90°   135°

(c)
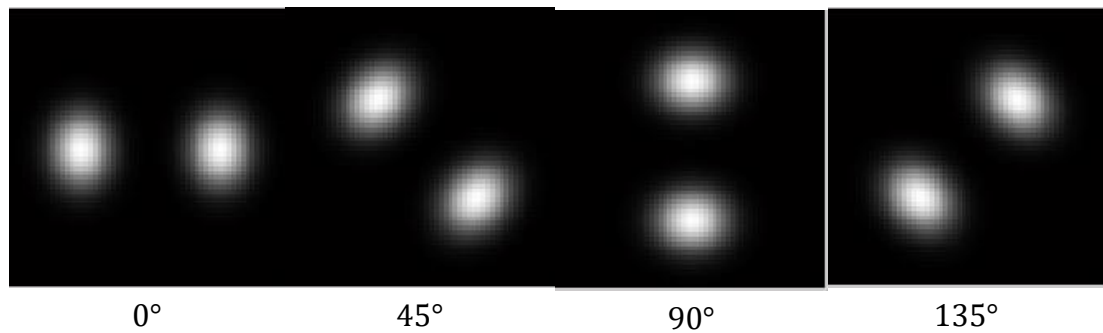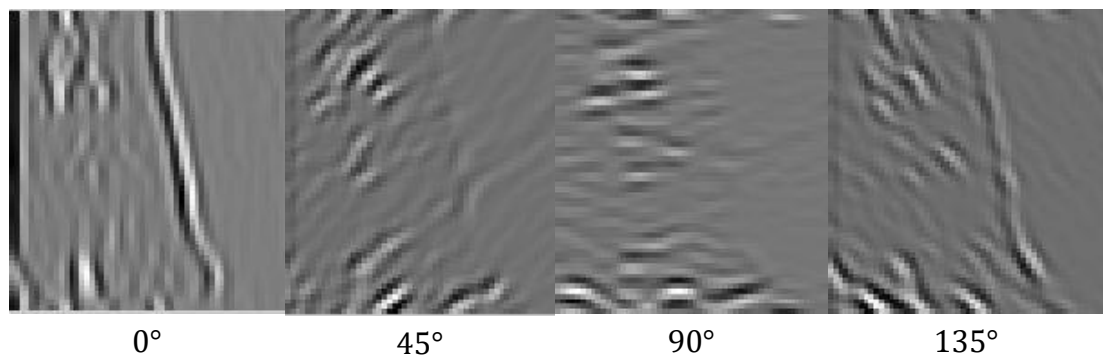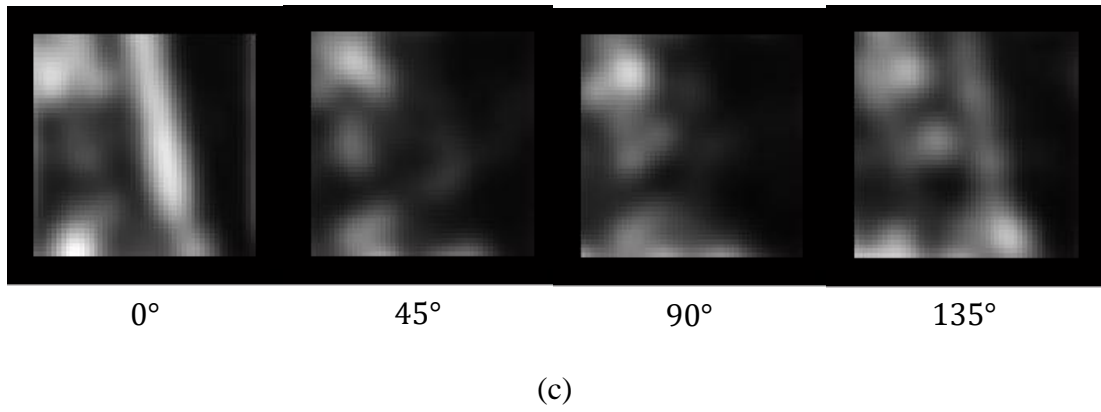
Figure 11: (a) Top row: magnitude of the Fourier transform of image in 16(a) multiplied the Gaussian masks for scale 1. (b) Second row: inverse Fourier transforms of the panels above. (c) Third row: local energy features computed from the panels above.

0°   45°   90°   135°

(a)

0°   45°   90°   135°

(b)

| 0° | 45° | 90° | 135° |

(c)
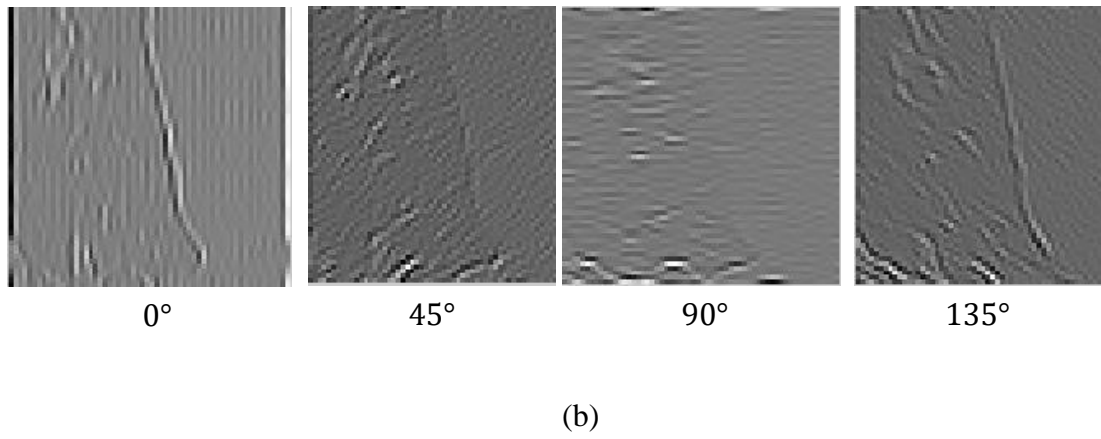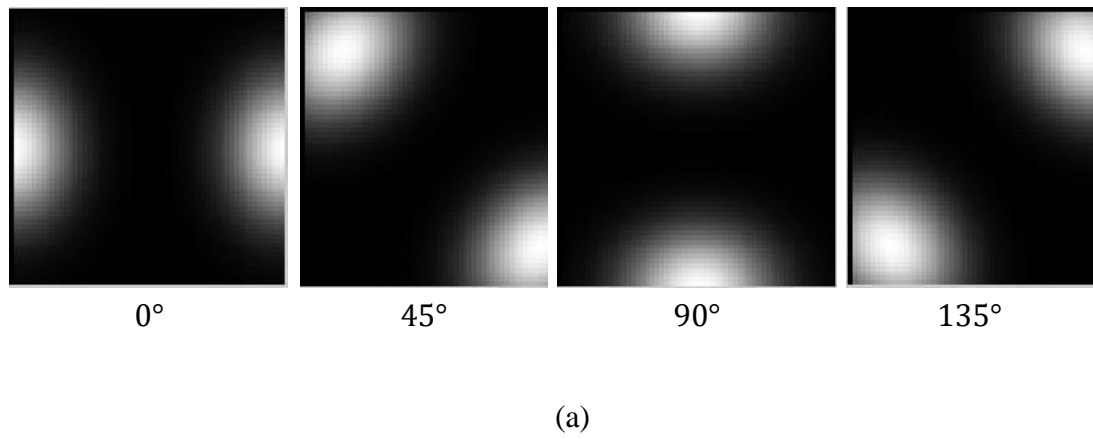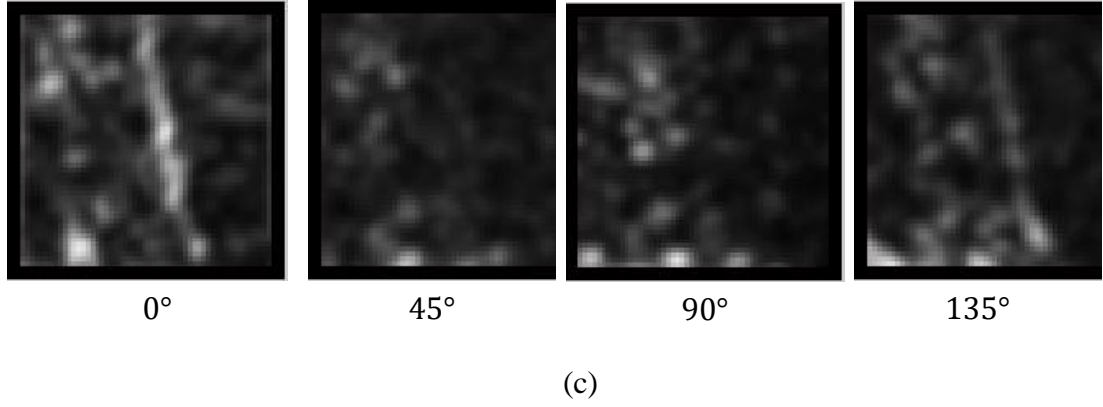
Figure 12: (a) Top row: magnitude of the Fourier transform of image in Figure 16(a) multiplied the Gaussian masks for scale 2. (b) Second row: inverse Fourier transforms of the panels above. (c) Third row: local energy features computed from the panels above.

Figure 10 - 12 shows the patches is undergo the FT, multiplication of Gaussian masks and IFT at scale 1 orientation 4 and scale 2 orientation 4.

3.3.2   Classical Features

By utilizing a classical features extraction in this project, only certain features of the image samples are extracted. This technique only involving the calculation of the important features that mostly contain in the grayscale image.

The features that has been chosen are mean, variance, kurtosis, skewness and power. Below is the features' equation that are used

$$mean = \bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}, \tag{9}$$

The X_i is the signal value at i-th instant thus the average values within the data set are calculated, where N is the total data.

$$variance = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}, \tag{10}$$

Equation above is defined from the dispersion within the data set which is the variance.

$$kurtosis = N \frac{\sum_{i=1}^{N} (X_i - \bar{X})^4}{[\sum_{i=1}^{N} (X_i - \bar{X}^2)]^2}, \tag{11}$$

17

Determining the "peakedness" of data distribution and frequency of extreme values is defined as the kurtosis.

$$skewness = \sqrt{N}\,\frac{\sum_{i=1}^{N}(X_i - \bar{X})^3}{[\sum_{i=1}^{N}(X_i - \bar{X}^2)]^{\frac{3}{2}}}, \tag{12}$$

Skewness is the degree of symmetry in the probability distribution of a data set.

$$power = \frac{\sum_{i=1}^{N}(X_i)^2}{N}, \tag{13}$$

The average of the data energy can be obtained from the power equation.

3.4 k-nearest neighbor (k-NN) Classifier

The other add-on technique for this approach is the k-nearest neighbor (k-NN) classifier. This k-NN is an algorithm which is used for classification, pattern recognition and statistical estimation. Based on the similarity measurement, k-NN can stores available cases and classifies the previous cases to new cases. G. Bhattacharya et al. [19] said that this performance of k-NN algorithm can be measured by the right distance measure, similarity measure for voting and suitable value/choice for parameter k. The performance of the k-NN classifier is increased when the distance measure and the similarity function is improved [20]. Majority vote of its neighbors is how each case is classified where the case is allocated to the most common class among its measured distance function of the k-NN. Given a situation where k = 1 thus the case will be allocated to the class of its nearest neighbor. Below is the distance functions of the k-NN classifier:

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \qquad \text{- Euclidean} \tag{14}$$

$$\sum_{i=1}^{k}|x_i - y_i| \qquad \text{- Manhattan} \tag{15}$$

$$\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{\frac{1}{q}} \quad \text{- Minkowski} \tag{16}$$

All of this distance functions are only can be used for continuous variables. In this research, the focus is more on the Euclidean formula. Other than that, 2 types

of set which contain 50 number sets will be used. There will be 25 number of training sets and the other 25 number will be the test set. The best result data is when choosing the optimal value for k. Basically, if the value of k is set to be large, it will give accurate results due to reduction of noise. Example of using this algorithm is shown in Figure 13 and Figure 14:
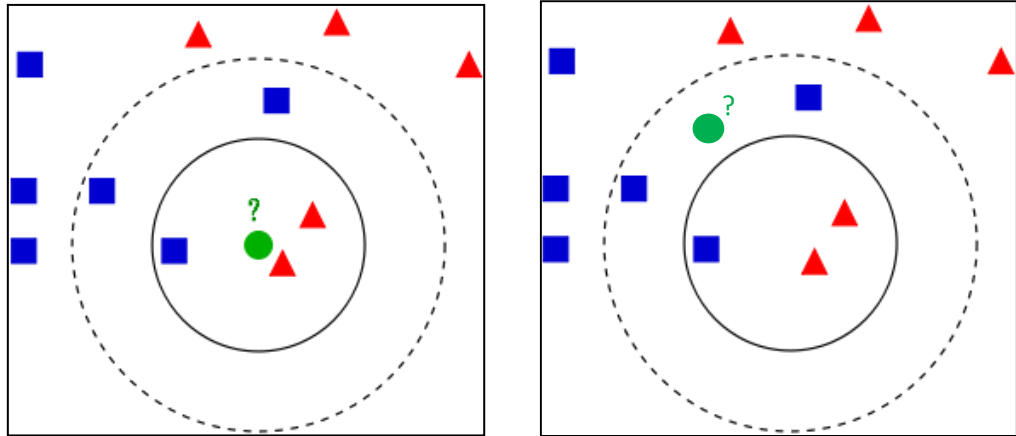
Figure 13: Test sample within the solid line circle [25] Figure 14: Test sample within the dotted line circle [25]

The green circle or the test sample should be categorized either to the blue squares or blocks in the first class or to the red triangles in the second class. Figure 13 shows a case if value of $k = 3$, it is within the solid line circle where the test sample is classified to the second class due to the greater number of red triangles compare to blue squares within the solid line circle. Another case in Figure 14 shows a case if value of $k = 5$, it will be within the dotted line where the test sample is classified to first class due to the greater number of blue squares within the dotted line circle.

3.5 Cross Validation

This is a technique where the misclassification rate is calculated in the total number of the sample images. It is a way for a model to a hypothetical validation set to be predicted when explicit validation set is not available.

3.5.1 Holdout Method

The holdout method is the most common and simple type of cross validation technique. The data sets are divided into two sets which are training set and testing/sample set. By using the training set only, the function approximator are fits into a function where it will guess the output values for the data located in the testing set. The mean absolute test set error will be calculated which is used to evaluate the model. The disadvantage of this technique is the evaluation may have a high variance. Thus this technique is used as for foundation of the cross-validation technique.
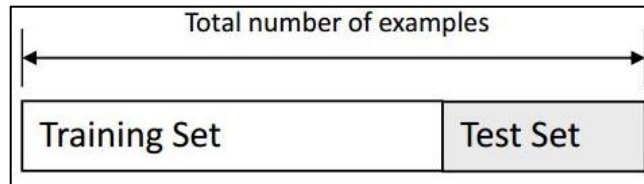
Figure 15: Illustration of holdout method [26]

3.5.2 k-fold Cross Validation

In this project, the improved technique for the holdout technique is used which is the *k*-fold cross-validation technique. The data set is separated into *k* subsets and the *k* times of holdout technique is repeated. Each time of the experiment, the *k* subsets are used as the test set and the *k-1* subsets are combined together to form a training set. Every data will be have a chance to be in the test set and *k-1* times in a training. Thus the average error across all of the *k* trials are computed. As the value of *k* is increased, the variance of the resulting estimate is reduced. The error is estimated as the average error rate on test examples.

$$E = \frac{1}{K}\sum_{i=1}^{K} E_i \tag{17}$$

Figure 16 shows the illustration of using *k*-fold cross validation where for each of *k* experiments, *k-1* folds for training is used and different fold for testing.
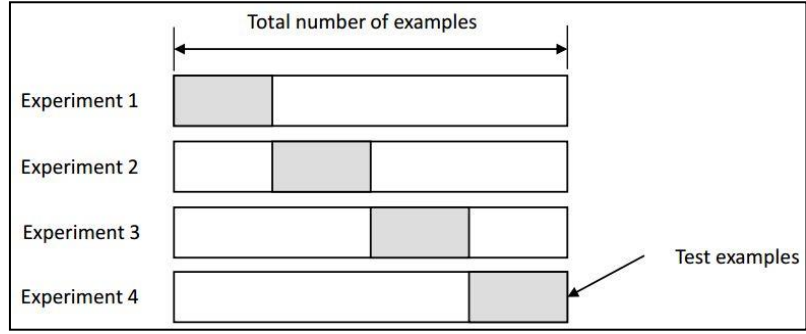
Figure 16: Illustration of k-fold cross validation [26]

3.6 Feature Reduction

A feature reduction technique is also used in this project. The purpose of using this technique is to select the data that are have higher number in discriminatory information from both Gabor filter and classical feature extraction results in order to avoid the classifier performing badly through choosing poor choice of features and to reduce computational time. Highly number of informative features is selected which an attempt to put of classes in the feature space far apart from each other and to locate the data points within each of the class close to each other.

3.6.1 Fisher's Discriminant Ratio

Fisher's Discriminant Ratio (FDR) is one of the feature reduction technique that a have been chosen in this project. Its function is to compute the discriminatory power of individual features between two selected classes. It is independent of the type of class distribution. The FDR can defined as

$$FDR = \frac{(m_1 - m_2)^2}{(\sigma_1^2 + \sigma_2^2)} \qquad (18)$$

21

3.7 Key Milestone

Figure 17 is the project/key milestone for FYP I.

FYP I project/key milestone

Week 1	Week 6	Week 9	Week 13	Week 14
•Title Selection/Proposal	•Extended Proposal Submission	•Viva: Proposal Defence and Progress Evaluation	•Draft Report Submission	•Final Report Submission
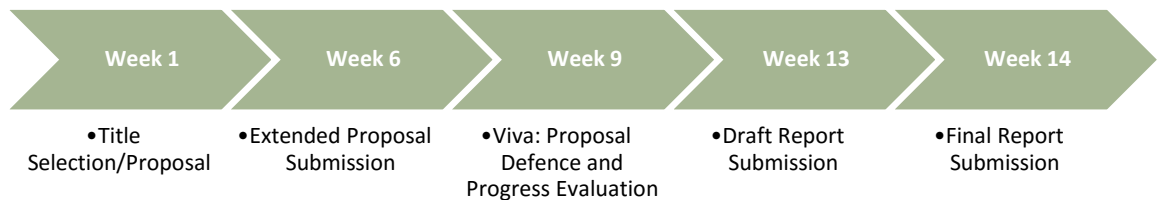
Figure 17: FYP 1 Key Milestone

As per shown in the Figure 17, in week 1, the title of FYP I is posted for each students to choose for their project. Between week 2 and week 5, the draft of the report can be submitted and checked by respective supervisor(s) for improvement. In week 6, the students need to submit their Extended Proposal to their respective supervisor(s) and through Turnitin website. This website is used to detect plagiarisms and notify lecturer in charge where each students must not exceed FYP requirement where it is 25% plagiarisms.

In week 9, there will be a Viva where the Proposal Defense and Progress Evaluation event is take place. This event require the students to present their project feasibility and identify the tools available in the university and the period require for the project to be done. During week 13 and week 14, the Draft Report and Final Report FYP I need to be submitted to the respective supervisor(s) and through Turnitin.

Figure 18 is the project/key milestone for FYP II.

FYP II project/key milestone

Week 8	Week 11	Week 13	Week 14	Week 15
•Progress Report Submission	•ELECTREX/PRE-SEDEX	•Draft Report Submission	•Final Report/Technical Report Submission	•Viva

Figure 18: FYP 2 Key Milestone

As shown in the Figure 18, in week 8, the submission of the Progress Report must be submitted through the Turnitin website where the plagiarism must not exceed 25%. The progress report is the report for which the students continued the project from the FYP 1. In week 11, the students must prepare the poster for presentation to the internal evaluators. During this presentation, the students suppose already their results for the project.

During week 13, the draft of the final report can be submitted and checked by the student's respective supervisor(s) for improvement. As for the week 14, the Final Report/Technical Report must be submitted through the Turnitin and supervisor(s). During week 15, there will be last Viva which will be evaluated by the external evaluator(s) and one internal evaluator.

3.8 Gantt Chart

Table 1 is the Gantt chart for FYP I.

Table 1: Gantt chart for FYP 1

No	Activity / Time	Semester Week													
		2015													
		Jan		Feb				March				April			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Project Title Research	█	█												
2	Research and prepare extended proposal		█	█	█	█	█								
3	**Submission of Extended Proposal**						▓								
4	Algorithm design and preparation of proposal defence						█	█							
5	**Oral presentation to supervisor and external examiner**									▓					
6	Development of algorithm									█	█	█			
7	Demonstrate algorithm and prepare interim report												█	█	
8	**Submission of Draft Interim Report**													▓	
9	**Finalisation and submission of Interim Report**														▓

Table 2 is the Gantt chart for FYP II.

Table 2: Gantt chart for FYP 2

No	Activity / Time	Semester Week														
		2015														
		May		June				July					August			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Project Work Continues	▓	▓	▓	▓	▓	▓	▓	▓							
2	**Submission of Progress Report**								�							
3	Project Work Continues								▓	▓	▓	▓	▓	▓	▓	
4	**ELECTREX /Pre-SEDEX**											▓				
5	**Submission of Draft Final Report**													▓		
6	**Submission of Dissertation**														▓	
7	**Submission of Technical Report**														▓	
8	**Viva**															▓
9	**Submission of Project Dissertation**															▓

25

CHAPTER 4: RESULTS AND DISCUSSIONS

In this chapter, we present the results related to testing and evaluation of the algorithm. Two main sections are presented, namely 4.1 Gabor-based classification technique, and 4.2 Classical-based classification technique. In both sections, we first calculate the misclassification rate using all the features generated by Gabor or classical features, and then we investigate the possibility of reducing the number of features. The reduced features are selected based on Fisher discriminant ratio which ranks the features from the most prominent to the least prominent. In all experiments, the misclassification rate are calculated using k-fold cross validation technique.

4.1 Gabor-based classification technique

In this experiment, a histogram of energy is created based on the sample of images. Each images will produced different number of energy where it represent the textures extracted from the images. A few test has been conducted for the texture extraction by using MATLAB.

In order for the histogram to be generated, an emphysema database is required. The emphysema database that contain 168 number of images where it contains 59 of NT, 50 of CLE and 59 of PSE are used. The size of the patches or sample images are in the order of 61×61 pixel.

Features that has been extracted using Gabor Filter function are expressed in the forms of energy value is based on the 2 scale and 4 orientation $(0°, 45°, 90°, 135°)$. The energy value are plot at decreasing order where it start from the left side. (a) Normal Tissue (NT), (b) Centrilobular Emphysema (CLE), (c) Paraseptal Emphysema (PSE).
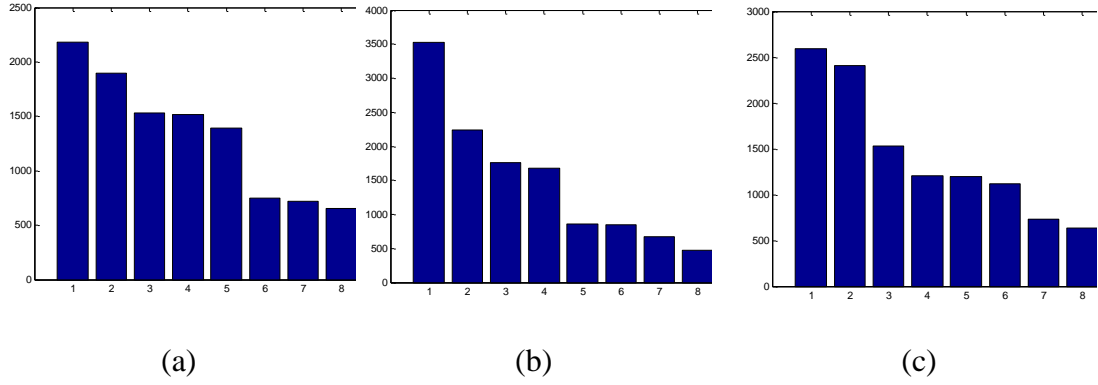
Figure 19: Energy Plot of 3 classes of Emphysema (a) Normal Tissue (NT)
(b) Centrilobular Emphysema (CLE) (c) Paraseptal Emphysema (PSE)

By using the data obtained from the histogram of the local energy value from Gabor feature extraction, the value for each of the data is saved in separate file. The local energy values are classified into the total number (normal and abnormal) of image samples and total number abnormal tissue image samples with all of the 8 features that have been filtered. From the total number data energy values, it is distributed and classified into 3 classes of emphysema according to its corresponding number of samples (NT = 59, CLE = 50, PSE = 59).

Table 3: Local Energy values classified into Normal Tissue (NT) class

No. of Samples	Type of Features							
	Energy 1	*Energy 2*	*Energy 3*	*Energy 4*	*Energy 5*	*Energy 6*	*Energy 7*	*Energy 8*
1	1529.52	2178.81	1894.32	1517.21	752.72	722.17	1391.68	654.39
2	3179.01	948.73	2450.04	1797.61	1115.01	497.42	1438.45	469.57
3	786.35	575.41	980.89	1457.39	322.15	251.08	351.07	423.83
4	1025.69	288.21	1953.70	707.88	479.32	172.62	786.66	253.79
5	1164.10	1005.34	1048.97	974.73	580.11	443.93	470.09	362.90

Table 4: Local Energy values classified into Centrilobular Emphysema (CLE) class

No. of Samples	Type of Features							
	Energy 1	*Energy 2*	*Energy 3*	*Energy 4*	*Energy 5*	*Energy 6*	*Energy 7*	*Energy 8*
1	1755.22	2237.61	3522.54	864.94	849.85	669.38	1682.91	474.42
2	904.80	772.64	1744.77	1062.63	409.79	356.21	798.42	408.73
3	1299.92	920.25	2727.21	870.12	479.13	218.01	1130.16	272.04
4	2253.74	3205.43	1439.98	400.27	879.03	425.40	622.83	135.15
5	821.52	691.54	1206.83	850.77	427.74	440.29	847.40	446.38

Table 5: Local Energy values classified into Paraseptal Emphysema (PSE) class

No. of Samples	Type of Features							
	Energy 1	*Energy 2*	*Energy 3*	*Energy 4*	*Energy 5*	*Energy 6*	*Energy 7*	*Energy 8*
1	2593.22	1117.27	2404.34	1200.16	1204.55	632.79	1534.71	735.31
2	2077.64	7289.54	4505.22	1058.17	1057.45	2115.54	2330.82	641.76
3	2075.68	663.02	1375.53	490.75	928.99	270.92	701.93	254.55
4	1576.38	638.25	1157.15	844.97	718.68	200.28	337.18	194.92
5	2969.60	1158.71	1600.94	964.53	1083.22	540.75	1131.89	509.63

By using the k-NN classifier, the each of local energy data is grouped into their respective group which in this case the Normal Tissue (NT) should be Group 1, Centrilobular Emphysema (CLE) should be Group 2 and lastly Paraseptal Emphysema (PSE) should be Group 3 as mentioned before. The purpose of doing this experiment is to train the database to recognize the pattern and values of local energy for each classes of the emphysema.

4.1.1 Misclassification rate

In this section, the misclassification rate of the data is calculated using the k-fold cross validation technique. The Table 6 shows the k values is removed from the total number of examples in certain experiments as the test set and the misclassification rate between the normal tissue and abnormal tissue for Gabor feature extraction data.

Table 6: Misclassification rate of the Gabor features normal tissue and Gabor features abnormal tissue using k-fold cross validation

K-FOLD CROSS VALIDATION TEST (Normal Tissue vs. Abnormal Tissue)	
No. Test Set (k)	**Misclassification Rate (%)**
2	1.19
4	2.38
6	3.57
8	4.76
10	5.95
12	7.14
14	8.33
16	9.52
18	10.71
20	11.90

The Table 7 shows the k values is removed from the total number of examples in certain experiments as the test set and the misclassification rate between the classes in the abnormal tissue group which are Centrilobular Emphysema (CLE) & Paraseptal Emphysema (PSE) for Gabor feature extraction data.

Table 7: Misclassification rate of the Gabor features abnormal tissue classes
(CLE & PSE) using k-fold cross validation

K-FOLD CROSS VALIDATION TEST (CLE vs. PSE)	
No. Test Set (k)	**Misclassification Rate (%)**
2	1.83
4	3.67
6	5.50
8	7.34
10	9.17
12	11.01
14	12.84
16	14.68
18	16.51
20	18.35

4.1.2 Feature Reduction

In this experiment, the features selection is done by using the Fisher's Discriminant Ratio (FDR) where it rank all of the features selected into the most discriminant features that can be used to classify the images samples. The results for the Gabor features data is recorded as below which features that have most discriminant values is rank from left to the least discriminant values to the right.

Table 8: Features selection based on the ranking of each features between
Normal Tissue (NT) and Centrilobular Emphysema (CLE) for Gabor features

Type of Features	Energy 5	Energy 1	Energy 6	Energy 4	Energy 2	Energy 7	Energy 8	Energy 3
Most Discriminant Values	0.22	0.22	0.15	0.11	0.09	0.09	0.07	0.05

Table 9: Features selection based on the ranking of each features between
Normal Tissue (NT) and Paraseptal Emphysema (PSE) for Gabor features data

Type of Features	Energy 3	Energy 7	Energy 5	Energy 1	Energy 8	Energy 2	Energy 4	Energy 6
Most Discriminant Values	0.26	0.15	0.14	0.04	0.04	0.03	0.03	0.02

Table 10: Features selection based on the ranking of each features between Centrilobular
Emphysema (CLE) and Paraseptal Emphysema (PSE) for Gabor features data

Type of Features	Energy 5	Energy 1	Energy 3	Energy 7	Energy 4	Energy 2	Energy 6	Energy 8
Most Discriminant Values	0.95	0.88	0.44	0.37	0.21	0.20	0.18	0.13

Three most discriminant features are selected from each experiments which involve Gabor-based features which reduced the number of features. Then the cross validation technique is applied to the reduced features. The results generated from the reduced features data are the same as the full features data for Gabor-based classification technique. The data generated can be refer back in Table 6 – 7.

4.2 Classical-based classification technique

This classical feature extraction is where five features of the image samples which are mean, variance, skewness, kurtosis and power are extracted. The purpose of using this additional technique is as a comparison purpose with the Gabor filter results. For this technique, the data also have been classified into 3 classes emphysema according to its corresponding number of samples (NT = 59, CLE = 50, PSE = 59) by using k-NN classifier.

Table 11: Classical features value from Normal Tissue (NT) class

No. of Samples	Type of Features				
	Mean	*Variance*	*Skewness*	*Kurtosis*	*Power*
1	-692.99	69727.21	4.18	24.77	549948.53
2	-287.01	267789.48	0.68	2.89	350097.00
3	-746.44	24722.61	3.46	16.26	581892.95
4	-457.55	172681.13	0.39	1.27	381988.55
5	-744.79	30338.77	4.85	31.85	585052.13

Table 12: Classical features value from Centrilobular Emphysema (CLE) class

No. of Samples	Type of Features				
	Mean	*Variance*	*Skewness*	*Kurtosis*	*Power*
1	-679.54	114621.96	2.91	12.25	576378.42
2	-744.24	44133.61	2.82	10.78	598016.51
3	-759.05	63684.59	2.52	8.51	639823.95
4	-759.09	97793.31	4.12	21.24	673985.52
5	-746.68	50992.92	2.44	8.90	608520.56

Table 13: Classical features value from Paraseptal Emphysema (PSE) class

No. of Samples	Type of Features				
	Mean	*Variance*	*Skewness*	*Kurtosis*	*Power*
1	-291.03	217926.80	-0.03	1.48	302570.19
2	-241.23	251433.29	0.34	2.09	309558.82
3	-477.24	167729.27	0.65	1.58	395445.02
4	-302.86	190715.19	-0.19	1.31	282390.06
5	-249.60	227015.65	-0.04	1.40	289257.41

4.2.1 Misclassification Rate

The Table 14 shows the k values is removed from the total number of examples in certain experiments as the test set and the misclassification rate between the normal tissue and abnormal tissue for classical feature extraction data.

Table 14: Misclassification rate of the classical features normal tissue and classical features abnormal tissue using k-fold cross validation

K-FOLD CROSS VALIDATION TEST (Classical Normal Tissue vs. Classical Abnormal Tissue)	
No. Test Set (k)	**Misclassification Rate (%)**
2	1.19
4	2.38
6	3.57
8	4.76
10	5.95
12	7.14
14	8.33
16	9.52
18	10.71
20	11.90

The Table 15 shows the 'k' values is removed from the total number of examples in certain experiments as the test set and the misclassification rate between the classes in the abnormal tissue group which are Centrilobular Emphysema (CLE) & Paraseptal Emphysema (PSE) for classical feature extraction data.

Table 15: Misclassification rate of the classical features abnormal tissue classes (CLE & PSE) using k-fold cross validation

K-FOLD CROSS VALIDATION TEST (Classical CLE vs. Classical PSE)	
No. Test Set (k)	**Misclassification Rate (%)**
2	1.83
4	3.67
6	5.50
8	7.34
10	9.17
12	11.01
14	12.84
16	14.68
18	16.51
20	18.35

As we can see from the Table 6, Table 7, Table 14 and Table 15, the results can be concluded that the misclassification rate of the k-fold cross validation for both techniques are increased as the number of test set (k) increased. This

is due to the number of the sample is reduced as the number of k is chosen incrementally and also the number of training set is reduced. From tables for both techniques, the same results are generated which both techniques produced the same rate of accuracy.

4.2.2 Feature Reduction

The results for the reduced feature for classical-based technique data is recorded as below which features that have most discriminant values is rank from left to the least discriminant values to the right.

Table 16: Features selection based on the ranking of each features between Normal Tissue (NT) and Centrilobular Emphysema (CLE) for classical features data

Type of Features	Power	Mean	Variance	Skewness	Kurtosis
Most Discriminant Values	1.71	1.04	0.46	0.22	0.15

Table 17: Features selection based on the ranking of each features between Normal Tissue (NT) and Paraseptal Emphysema (PSE) for classical features data

Type of Features	Power	Mean	Variance	Skewness	Kurtosis
Most Discriminant Values	25.88	1.81	1.35	1.12	0.90

Table 18: Features selection based on the ranking of each features between Centrilobular Emphysema (CLE) and Paraseptal Emphysema (PSE) for classical features data

Type of Features	Mean	Power	Variance	Skewness	Kurtosis
Most Discriminant Values	81.58	10.66	6.63	5.82	1.88

33

Three most discriminant features are selected from each experiments which involve classical-based features which reduced the number of features. Then the cross validation technique is applied to the reduced features. The results generated from the reduced features data are the same as the full features data. The data generated can be refer back in Table 14 – 15.

CHAPTER 5: CONCLUSION AND RECOMMENDATION

In this work, emphysema classification of lungs CT images based on Gabor filter is developed and tested. The Gabor filter are used to extract the feature of the CT sample images or patches. The feature are expressed in terms of energy values at 2-scale and 4-orientation and displayed in histogram format. In addition to Gabor, we also the classical features, that is mean, variance, kurtosis, skewness and power. The classical-based method are used for comparison with the Gabor-based method.

In order to reduce computational time, Fisher Discriminant Ratio (FDR) is calculated for both Gabor-based and classical-based method. This technique is to rank and choose the most discriminant features generated by Gabor filter or the classical features. Based on the FDR ranking, only the most discriminant features are selected and thereby reduce the computational time. In this experiment, full and reduced features for both methods are evaluated. The performance of the algorithm, measured in terms of misclassification rate, were evaluated by using existing emphysema database. The results show low misclassification rate for the Gabor-based method which indicate the good potential of Gabor-based emphysema classification.

It should be noted that the misclassification rate for both Gabor based and classical-based method is the same and therefore further investigation. Ideally, since Gabor filter offer the advantage of multi-resolution and multi-orientation properties, we should expect it to perform better that the classical-based method. So for future work, the Gabor-based and the classical-based method should be tested with other data set or in other applications. It is hope the advantage of multi-resolution and multi-orientation properties of Gabor filter will shows it benefit than the classical-based method.

REFERENCES

[1]   E.N. Hughes, R.E. Simpson, "Pulmonary Emphysema and Associated Problems in Anaesthesia. A Review," in Canadian Anaesthetists' Society Journal, pp. 341-342.

[2]   M. Takahashi, J. Fukuoka, N. Nitta, "Imaging of pulmonary emphysema: A pictorial review," in International Journal of COPD, 2008, pp. 193-204.

[3]   Rabe et al., "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease," in GOLD executive summary, Sep 2007, pp. 532-555.

[4]   T. Purushotham et al., "Pattern recognition diagnostics for Emphysema using phase only correlation technique," March 2010, pp. 40-43.

[5]   S. Nazari, M. Shahram Moin, "Face recognition using global and local Gabor features," 2013, pp. 1-4.

[6]   O. M. Mets, P. A. Jong, B. van Ginneken, H. A. Gietema, J. W. Lammers, "Quantitative computed tomography in COPD: possibilities and limitations," 2012, pp. 133-145.

[7]   D. Kathryn, F. Carol, "Pulmonary histologic changes in Marfan Syndrome – A case series and literature review," in Anatomic Pathology, 2011, pp. 857-863.

[8]   R. Nava, B. Escalante-Ramírez, G. Cristóbal, R. San José Estépar, "Extended Gabor approach applied to classification of emphysematous patterns in computed tomography," in Med Biol Eng Comput, 2014, pp. 393-403.

[9]   A. Madani, J. Zanen, V. de Maertelear, P. A. Gevenois, "Pulmonary emphysema: objective quantification at multi-detector row CT-comparison with macroscopic and microscopic morphometry," in Radiology, pp. 1036-1043.

[10]   M. Emam, J. F. Renaud de la Faverie, N. Gharbi, "Characterization of lungs's emphysema distribution: Numerical assessment of disease development," 2010, pp. 464-469.

[11]   D. Litmanovich, P. M. Boiselle, A. Bankier, "CT of pulmonary emphysema – current status, challenges, and future directions," 2009, pp. 537-51.

[12]   L. Sørensen, S. B. Shaker, M. de Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," in IEEE Trans Med Imaging, 2010, pp. 559-569.

[13]   A. F. Costa, G. Humpire-Mamani, A. J. M. Traina, "An efficient algorithm for fractal analysis of textures," in 25th Conference on graphics, patterns, and images (SIBGRAPI), 2012, pp. 39-46.

[14]   C. S. Mendoza, G. R. Washko, J. C. Ross, A. A. Diaz, D. A. Lynch, J. D. Crapo, E. K. Silverman, B. Acha, C. Serrano, R. S. Estépar, "Emphysema quantification in a multi-scanner HRCT cohort using local intensity distributions," in 9th IEEE International symposium on biomedical imaging, 2012, pp. 474-477.

[15]   D. Gabor, "Theory of communication – Part III: Radio and Communication Engineering," in Journal of the Institution of Electrical Engineers, 1946, pp. 429-457.

[16]   B. David, Z. Pavel, "Gabor wavelets in image processing", pp. 1-5.

[17]   R. Redondo, F. Šroubek, S. Fischer, G. Cristóbal, "Multifocus image fusion using the log-Gabor transform and multisize windows technique", in Inf Fusion, 2009, pp. 163-171.

[18]   R. Nava, B. Escalante-Ramírez, G. Cristóbal, "Texture image retrieval based on log-Gabor features," in Progress in pattern recognition, image analysis, computer vision, and applications, Lecture Notes in Computer Science, vol 7441, pp. 414-421.

[19]   G. Bhattacharya, K. Ghosh, A. S. Chowdhury, "Test point specific k estimation for k-NN classifier," in 2014 22nd International Conference on Pattern Recognition, 2014, pp. 1478-1483.

[20]   G. Bhattacharya, K. Ghosh, A. S. Chowdhury, "An affinity-based new local distance function and similarity measure for k-NN algorithm," in Pattern Recognition Letters, vol. 33, 2012, pp. 356-363.

[21]   L. Yi-hui, M. Muftah, T. Das, L. Bai, K. Robson, D. Auer, "Classification of MR tumor images based on Gabor wavelet analysis," in Journal of Medical and Biological Engineering, 2010, pp. 22-28.

[22]   B. Kishore, V. L. Rana, K. Manikantan, S. Ramachandran, "Face recognition using Gabor-feature-based DFT shifting," pp. 1-8.

[23]   Dr Henry Knipe, D. Y. (2015). *Pulmonary emphysema*. Retrieved from Radiopaedia.org: http://radiopaedia.org/articles/pulmonary-emphysema

[24]   F. Sylvain, S. Filip, P. Laurent, R. Rafael, C. Gabriel, "Self-Invertible 2D Log-Gabor Wavelets," in International Journal of Computer Vision 75(2), 2007, pp. 231-246.

[25]   Wikipedia. (2015, August 12). *k-nearest neighbors algorithm*. Retrieved from Wikipedia The Free Encyclopedia: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

[26]   Gutierrez-Osuna, R. (2015, August 15). *Lecture notes; Pattern recognition;Cross-Validation.* Retrieved from perception sensing instrumentation: http://psi.cse.tamu.edu/teaching/lecture_notes/

APPENDICES

Coding for Gabor filter

```matlab
clc;close all;clear all

img=mean(double(imread('168.tiff')),3); %In this case image files should be in the
same Folder

img = img(1:60,1:60);

imgsize=size(img);

if (sum(mod(imgsize,2))~=0)

    disp('Warning: Image must be of even size');

    if mod(imgsize(1),2)~=0

      disp(sprintf('It has %d rows. I am ignoring the last row',imgsize(1)));

      img= img(1:(imgsize(1)-1),:);

    end

    if mod(imgsize(2),2)~=0

      disp(sprintf('It has %d columns. I am ignoring the last
column',imgsize(1)));

      img= img(:,1:(imgsize(2)-1));

    end

end

disp(sprintf('The size of the image is: %d x %d',imgsize(1),imgsize(2)));

SIZE=max(imgsize);

highest=pi*(SIZE-2)/SIZE;

disp(sprintf('w_max: %f\n',highest));

w0factor = 10;

L = 8;

disp('Computing...');

w=[w0factor*2*pi/min(imgsize)]; % w0

disp(sprintf('w_0 : %f',w(1)));

Ktilde= log(highest/w(1))/log(2.0);

K= round(Ktilde);

for i=1:K

    w=[ w w(i)*2 ];

end

rho= [ 0 ];

incw= [ w(1) ];
```

```matlab
for i=1:K

    rho= [ rho (w(i+1)+w(i))/2 ];

    incw= [ incw w(i+1)-w(i) ];

end

incphi= 2*pi/L;

disp(sprintf('Radial bands: %d',K));

disp('At radii:');

i=1:K;

rho(2:(K+1))

disp('with bandwidth limits:');

incw(2:(K+1))

disp('with radial standard deviations:')

(2.^(i-2)).*w(1)./sqrt(2*log(2))

disp('and angular standard deviations:')

rho(i+1).*tan(incphi/2)/sqrt(2*log(2))

disp(sprintf('\nAngular increment: %f\n',incphi));

disp(sprintf('%d angular bands at angles:',(L/2)));

i=1:(L/2);

(i-1).*incphi.*180/pi

max_index= K*L/2;


all_masks= zeros(imgsize(1),imgsize(2),max_index);

all_ifts= zeros(imgsize(1),imgsize(2),max_index);

disp('Computing DFT');

FT=fft2(double(img));

FT= fftshift(FT);

mask=zeros(SIZE);

alpha=2.5;  % alpha value for psi

ener_vector = [];

pos_vector  = [];

ang_vector  = [];

for pos=1:K

    for ang=0:L/2-1

        index= (pos-1)*(L/2)+ang+1;

        sigmax= 0.85*w(1)*2^(pos-2); % The value in the border must be 0.5

        sigmay= 0.85*rho(pos+1)*tan(incphi/2); % Idem

        anglemid= ang*incphi;

        for y=-SIZE/2+1:SIZE/2-1
```

```matlab
            wy= y*2*pi/SIZE;

            for x=0:SIZE/2-1

                wx= x*2*pi/SIZE;

                wr= sqrt(wx^2+wy^2);

                mask(y+SIZE/2+1,x+SIZE/2+1)= exp( -
((abs(wx*cos(anglemid)+wy*sin(anglemid))-rho(pos+1))^2)/(2*sigmax*sigmax)-((-
wx*sin(anglemid)+wy*cos(anglemid))^2)/(2*sigmay*sigmay) );

                mask(-y+SIZE/2+1,-x+SIZE/2+1)= mask(y+SIZE/2+1,x+SIZE/2+1);

            end

        end

mask(SIZE/2+1,SIZE/2+1)=0; % Remove DC component

        rows=SIZE-imgsize(1);

        cols=SIZE-imgsize(2);

        masksamesize=mask((1+rows/2):SIZE-rows/2,(1+cols/2):SIZE-cols/2);

        masksamesize(1,:)=0; % Ensures symmetry

        masksamesize(:,1)=0; % Idem

        all_masks(:,:,index)= masksamesize(:,:);

        FTfil= FT.*masksamesize;

disp(sprintf('Computing IFT... rho: %d  angle: %d',pos,ang))

        pos_vector= [ pos_vector pos];

        ang_vector= [ang_vector ang];

        IFT=real(ifft2(ifftshift(FTfil)));

        all_ifts(:,:,index)= IFT(:,:);

        energy=sum(sum(IFT.^2))/(imgsize(1)*imgsize(2));

        ener_vector= [ ener_vector energy];

        disp(sprintf('Energy: %f',energy));

    end

end

max_ift_global=max(max(max(all_ifts)));

pos_ang_ener = [pos_vector;ang_vector;ener_vector]';

% Computing local energies
disp('Computing local energies. Please wait...');

all_energies= zeros(imgsize(1),imgsize(2),max_index);

SIGMA=2*pi/(sqrt(2)*rho(2));

S=round(sqrt(-2*SIGMA*SIGMA*log(0.1)));

for pos=1:K

maskavg=normpdf(-S:S,0,SIGMA);

    maskavg=maskavg/sum(maskavg);

    for ang=0:L/2-1
```

```matlab
        index= (pos-1)*(L/2)+ang+1;

        IFT(:,:)= all_ifts(:,:,index);

        psi= abs((1-exp(-2*alpha*IFT/max_ift_global))./(1+exp(-
2*alpha*IFT/max_ift_global)));

        ener= conv2(maskavg, maskavg', psi, 'valid');

        all_energies((1+S):(imgsize(1)-S),(1+S):(imgsize(2)-S),index)=ener(:,:);

    end

    S=fix(S/2);

    SIGMA=SIGMA/2;

End

max_ener=max(max(max(all_energies)));

FTvis= log(abs(FT)+1);

maxvalue=max(max(FTvis));

minvalue=min(min(FTvis));

FTvis= round((FTvis-minvalue)*255/(maxvalue-minvalue));

[ordered, indexes]= sort(ener_vector);

ordered=fliplr(ordered);

indexes=fliplr(indexes);

idx_ener = [indexes; ordered]';

%%%Plot of energies of 20 bands used in Gabor expansion of image
figure;idx = 1:length(ordered);

bar(idx,ordered);

pos = 1; %%%% 1 to 2 scale

ang = 4;  %%% 0 to 7 orientation

[masksamesize,ftplt,iftplt,psiplt,ener] =
gfn(pos,ang,all_masks,all_ifts,all_energies,FT,maxvalue,minvalue,alpha);

figure;imshow(img/255); xlabel('Original image');

figure;imshow(FTvis/255); xlabel('Fourier Transform');

figure;imshow(masksamesize); xlabel(sprintf('Filter mask\n(Scale: %d
Orientation: %d)',pos,ang));

figure;imshow(ftplt); xlabel('FT filtered');

figure;imshow(iftplt); xlabel('IFT');

figure;imshow(psiplt); xlabel('psi');

figure;imshow(ener/max_ener); xlabel('local energies');
```

Coding for k-NN classifier for Gabor-based

```matlab
clear all; close all; clc

fet = zeros(1,8);

lim = 168;

for images=1:lim

str = strcat(int2str(images),'.tiff'); %% In this case image files should be in
same Folder

eval('I=imread(str);');

% eval('load(str);');

I  = mean(double(I),3);

imgsize=size(I);

if (sum(mod(imgsize,2))~=0)

    disp('Warning: Image must be of even size');

    if mod(imgsize(1),2)~=0

      disp(sprintf('It has %d rows. I am ignoring the last row',imgsize(1)));

      I= I(1:(imgsize(1)-1),:);

     end

end

K  = gaborenergy(I)';

fet=cat(1,fet,K);

end

Totalfet  = fet(2:lim+1,:);

Abfet = fet(61:lim+1,:);

fetNT = Totalfet(1:59,:);

fetCLE = Totalfet(60:109,:);

fetPSE = Totalfet(110:168,:);

c = [ones(59,1);2*ones(50,1);3*ones(59,1)];

group1 = knnclassify(fetNT, Totalfet, c);

group2 = knnclassify(fetCLE, Totalfet, c);

group3 = knnclassify(fetPSE, Totalfet, c);
```

Coding for *k*-fold cross validation for Gabor-based (full feature)

```matlab
clear all; close all; clc

load('Totalfet.mat') %save full features data in '.mat' format

g = [ones(59,1); 2*ones(50,1); 3*ones(59,1)];

c = cvpartition(g,'k',2); %k value is choose incrementally by 2

fun = @(xT,yT,xt,yt)(sum(~strcmp(yt,classify(xt,xT,yT))));

rate = sum(crossval(fun,Totalfet,g,'partition',c))...
            /sum(c.TestSize)*100
```

Coding for Fisher's Discriminant Ratio (FDR) for Gabor-based

```matlab
clear all; close all; clc

%load Gabor database
load('fetNT.mat');

load('fetCLE.mat');


X = fetNT';

Y = fetCLE';

feature_names={'Energy1','Energy2','Energy3','Energy4','Energy5','Energy6','Energy7','Energy8'};

%Calculate each feature's FDR
[NumOfFeatures,N]=size(X);

for i=1:NumOfFeatures

    FDR_value(i)=Fisher(X(i,:),Y(i,:));

end

%Sort features in descending FDR value
[FDR_value,feature_rank]=sort(FDR_value,'descend');

XY = FDR_value', feature_names(feature_rank)'

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%load Gabor database
load('fetNT.mat');

load('fetPSE.mat');
X = fetNT';
Z = fetPSE';

%Calculate each feature's FDR
[NumOfFeatures,N]=size(X);

for i=1:NumOfFeatures

    FDR_value1(i)=Fisher(X(i,:),Z(i,:));

end

%Sort features in descending FDR value
[FDR_value1,feature_rank1]=sort(FDR_value1,'descend');

XZ = FDR_value1', feature_names(feature_rank1)'

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%load Gabor database
load('fetCLE.mat');

load('fetPSE.mat');

Y = fetCLE';

Z = fetPSE';

%Calculate each feature's FDR
[NumOfFeatures,N]=size(Y);

for i=1:NumOfFeatures

    FDR_value2(i)=Fisher(Y(i,:),Z(i,:));

end

%Sort features in descending FDR value
[FDR_value2,feature_rank2]=sort(FDR_value2,'descend');

YZ = FDR_value2', feature_names(feature_rank2)'
```

Coding for *k*-fold cross validation for Gabor-based (reduced feature)

```
clear all; close all; clc

load('Totalfet.mat') %save reduced features of Gabor data in '.mat' format

Totalfet = [Totalfet(:,1),Totalfet(:,3),Totalfet(:,5)]; %take the most prominent
features

g = [ones(59,1); 2*ones(50,1); 3*ones(59,1)];

c = cvpartition(g,'k',2)

fun = @(xT,yT,xt,yt)(sum(~strcmp(yt,classify(xt,xT,yT))));

rate = sum(crossval(fun,Totalfet,g,'partition',c))...
            /sum(c.TestSize)*100
```

Coding for classical feature extraction

```
clear all; close all; clc

fet = [];

lim = 168;

for images = 1:lim

str = strcat(int2str(images),'.tiff'); %% In this case image files should be in
same Folder

eval('I=imread(str);');

image = double(I);

data1 = image(:);

    mn = mean(data1); % mean
    vr = var(data1); % variance

    sk = skewness (data1); % skewness

     k = kurtosis(data1); % kurtosis

    pwr = sum(data1.^2)/length(data1);% Power
```

```
cf = [mn vr sk k pwr];

fet = cat(1,fet,cf);

end

classicalfet = fet(1:lim,:);

classicalAbfet = fet(60:lim,:);

cNT = fet(1:59,:);

cCLE = fet(60:109,:);

cPSE = fet(110:168,:);
```

Coding for k-NN classifier for classical-based (full feature)

```
clear all; close all; clc

load('classicalfet.mat') %save full features of classical data in '.mat' format

cNT = classicalfet(1:59,:);

cCLE = classicalfet(60:109,:);

cPSE = classicalfet(110:168,:);

c = [ones(59,1);2*ones(50,1);3*ones(59,1)];

group1 = knnclassify(cNT, classicalfet, c);

group2 = knnclassify(cCLE, classicalfet, c);

group3 = knnclassify(cPSE, classicalfet, c);
```

Coding for Fisher's Discriminant Ratio (FDR) for classical-based

```
clear all; close all; clc

%load classical database
load('cCLE.mat');


load('cPSE.mat');

Y = cCLE';

Z = cPSE';

feature_names={'Mean','Variance','Skewness','Kurtosis','Power'};

%Calculate each feature's FDR
[NumOfFeatures,N]=size(Y);

for i=1:NumOfFeatures

    FDR_value(i)=Fisher(Y(i,:),Z(i,:));

end

%Sort features in descending FDR value
[FDR_value,feature_rank]=sort(FDR_value,'descend');

cCLEvscPSE = FDR_value', feature_names(feature_rank)'
```

```matlab
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%load classical database
load('cNT.mat');

load('cCLE.mat');

X = cNT';

Y = cCLE';

%Calculate each feature's FDR
[NumOfFeatures,N]=size(X);

for i=1:NumOfFeatures

    FDR_value1(i)=Fisher(X(i,:),Y(i,:));

end

%Sort features in descending FDR value
[FDR_value1,feature_rank1]=sort(FDR_value1,'descend');

cNTvscCLE = FDR_value1', feature_names(feature_rank1)'

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%load classical database
load('cNT.mat');

load('cPSE.mat');


%Calculate each feature's FDR
[NumOfFeatures,N]=size(X);

for i=1:NumOfFeatures

    FDR_value2(i)=Fisher(X(i,:),Z(i,:));

end

%Sort features in descending FDR value
[FDR_value2,feature_rank2]=sort(FDR_value2,'descend');

cNTvscPSE = FDR_value2', feature_names(feature_rank1)'
```

Coding for *k*-fold cross validation for classical-based (reduced feature)

```matlab
clear all; close all; clc

load('classicalfet.mat')

Totalfet = [classicalfet(:,1),classicalfet(:,2),classicalfet(:,5)]; %take the
most prominent features

group = [ones(59,1); 2*ones(50,1); 3*ones(59,1)];

c = cvpartition(group,'k',4);

fun = @(xT,yT,xt,yt)(sum(~strcmp(yt,classify(xt,xT,yT))));

rate = sum(crossval(fun,Totalfet,group,'partition',c))...
            /sum(c.TestSize)*100
```