TEXTURE ANALYSIS OF MLO AND CC MAMMOGRAMS FOR BREAST CANCER DIAGNOSIS

A PROJECT REPORT

Submitted by

RAMYA.V **Reg. No.:1010107073**

RENUKA.S **Reg. No.:1010107074**

SHANMUKA PRIYA.R **Reg. No.:1010107088**

in partial fulfilment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

ELECTRONICS AND COMMUNICATION

ENGINEERING

KUMARAGURU COLLEGE OF TECHNOLOGY

COIMBATORE-641049

(An Autonomous Institution Affiliated to Anna University, Chennai)

APRIL 2015

KUMARAGURU COLLEGE OF TECHNOLOGY

COIMBATORE-641049

(An Autonomous Institution Affiliated to Anna University, Chennai)

BONAFIDE CERTIFICATE

Certified that this project reporttitled**"TEXTURE ANALYSIS OF MLO AND CC MAMMOGRAMS FOR BREAST CANCER DIAGNOSIS"** is the bonafied work of "**RAMYA.V,RENUKA.S,SHANMUKA PRIYA.R"** who carried out the project work under my supervision.

SIGNATURE

Ms.S.Sasikala, M.Tech., (Ph.D).,
Associate Professor/E.C.E
Kumaraguru College of Technology
Coimbatore.

SIGNATURE

Dr.RajeswariMariappan M.E., Ph.D.,
HEAD OF THE DEPARTMENT
Electronics & Communication Engineering
Kumaraguru College of Technology
Coimbatore.

The candidates with Register numbers 1010107073,1010107074 and 1010107088 are examined by us in the project viva-voce examination held on ……

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

First we would like to express our praise and gratitude to the Lord, who has showered his grace and blessing enabling us to complete this project in an excellent manner. He has made all things in beautiful in his time.

We express our sincere thanks to our beloved Joint Correspondent, **Shri. Shankar Vanavarayar** for his kind support and for providing necessary facilities to carry out the project work.

We would like to express our sincere thanks to our beloved Principal, **Dr.R.S.Kumar M.E., Ph.D.,** who encouraged us with his valuable thoughts.

We would like to express our sincere thanks and deep sense of gratitude to our HOD,**Dr.RajeswariMariappan M.E., Ph.D.,**for her valuable suggestions and encouragement which paved way for the successful completion of the project.

We are greatly privileged to express our deep sense of gratitude to the Project Coordinator **Ms. A. Kalaiselvi M.E.,(Ph.D)**,Assistant Professor,for her continuous support throughout the course.

In particular, We wish to thank and express our everlasting gratitude to the Supervisor **Ms. S.Sasikala M.Tech., (Ph.D)**, Associate Professor for her expert counselling in each and every steps of project work and we wish to convey our deep sense of gratitude to all teaching and non-teaching staff members of ECE Department for their help and cooperation.

Finally,we thank our parents and our family members for giving us the moral support in all of our activities and our dear friends who helped us to endure our difficult times with their unfailing support and warm wishes.

ABSTRACT

Mammogram is an X-ray image of breast used to screen breast cancer. Mammograms play a key role in early breast cancer detection and helps to decrease the breast cancer mortality rate . Mammogram can be used either for screening or for diagnostic purposes. Mammogram depends on age of the person to be tested. The analysis of the affected breast image must be done as early as possiible to know the percentage of affected area for further treatment. In this paper, an approach is proposed to develop a SVM classification system for cancer detection from digital mammograms. The proposed system consists of three steps. The first step is to identify the Region of Interest (ROI) of 256×256 pixel. Second step is the feature extraction is done by Gabor filter, Steerable pyramid, Greylevel co-occurrence matrix (GLCM) method. Based on the data obtained, it concluded that texture analysis could distinguish malignant and benign images with considerably good accuracy. The third step is the classification process, SVM (Support Vector Machine) technique uses the image content to classify normal and malignant masses.

LIST OF FIGURES

LIST OF ABBREVATIONS

GLCM - **G**rey **L**evel **C**o-occurance **M**atrix

HRT - **H**ormone **R**eplacement **T**herapy

SVM - **S**upport **V**ector **M**achine

FDA - **F**ood and **D**rug **A**dministration

PEM - **P**ositron **E**mission **M**ammography

MRI - **M**agnetic **R**esonance **I**maging

INV - **I**nverse **D**ifference

INN - **I**nverse **D**ifference **N**ormalized

CHAPTER 1

INTRODUCTION

1.1 OBJECTIVE

The aim of this paper is to

- Detect accuracy level
- To classify between normal and cancerous mass

1.2 INTRODUCTION TO BREAST CANCER AND MAMMOGRAM

1.2.1Breast cancer

Breast cancer is a kind of cancer that develops from breast cells. Breast cancer usually starts off in the inner lining of milk ducts or the lobules that supply them with milk. A malignant tumor can spread to other parts of the body. A breast cancer that started off in the lobules is known as *lobular carcinoma*, while one that developed from the ducts is called *ductal carcinoma*. The vast majority of breast cancer cases occur in females. Breast cancer is the most common invasive cancer in females worldwide. It accounts for 16% of all female cancers and 22.9% of invasive cancers in women. 18.2% of all cancer deaths worldwide, including both males and females, are from breast cancer. Breast cancer rates are much higher in developed

nations compared to developing ones. There are several reasons for this, with possibly life-expectancy being one of the key factors - breast cancer is more common in elderly women; women in the richest countries live much longer than those in the poorest nations. The different lifestyles and eating habits of females in rich and poor countries are also contributory factors is what experts believe. Below are the causes of breat cancer as recent studies had revealed

Getting older - the older a woman gets, the higher is her risk of developing breast cancer; age is a risk factor. Over 80% of all female breast cancers occur among women aged 50+ years (after the menopause).

Genetics - women who have a close relative who has/had breast or ovarian cancer are more likely to develop breast cancer. If two close family members develop the disease, it does not necessarily mean they shared the genes that make them more vulnerable, because breast cancer is a relatively common cancer. The majority of breast cancers are not hereditary. Women who carry the BRCA1 and BRCA2 genes have a considerably higher risk of developing breast and/or ovarian cancer. These genes can be inherited. TP53, another gene, is also linked to greater breast cancer risk.

A history of breast cancer - women who have had breast cancer, even non-invasive cancer, are more likely to develop the disease again, compared to women who have no history of the disease.

Having had certain types of breast lumps - women who have had some types of benign (non-cancerous) breast lumps are more likely to develop cancer later on. Examples include atypical ductal hyperplasia or lobular carcinoma in situ.

Dense breast tissue - women with more dense breast tissue have a greater chance of developing breast cancer.

Estrogen exposure - women who started having periods earlier or entered menopause later than usual have a higher risk of developing breast cancer. This is because their bodies have been exposed to estrogen for longer. Estrogen exposure begins when periods start, and drops dramatically during the menopause.

Obesity - post-menopausal obese and overweight women may have a higher risk of developing breast cancer. Experts say that there are higher levels of estrogen in obese menopausal women, which may be the cause of the higher risk.

Height - taller-than-average women have a slightly greater likelihood of developing breast cancer than shorter-than-average women. Experts are not sure why.

Alcohol consumption - the more alcohol a woman regularly drinks, the higher her risk of developing breast cancer is. The Mayo Clinic says that if a woman wants to drink, she should not exceed one alcoholic beverage per day.

Radiation exposure - undergoing X-rays and CT scans may raise a woman's risk of developing breast cancer slightly. Scientists at the Memorial Sloan-Kettering Cancer Center found that women who had been treated with radiation to the chest for a childhood cancer have a higher risk of developing breast cancer.

HRT (hormone replacement therapy) - both forms, combined and estrogen-only HRT therapies may increase a woman's risk of developing breast cancer slightly. Combined HRT causes a higher risk.

Certain jobs - French researchers found that women who worked at night prior to a first pregnancy had a higher risk of eventually developing breast cancer. Canadian researchers found that certain jobs, especially those that bring the human body into contact with possible carcinogens and endocrine disruptors are linked to a higher risk of developing breast cancer. Examples include bar/gambling, automotive plastics manufacturing, metal-working, food canning and agriculture. They reported their findings in the November 2012 issue of *Environmental Health*.

Cosmetic implants may undermine breast cancer survival - women who have cosmetic breast implants and develop breast cancer may have a higher risk of dying prematurely form the disease compared to other females, researchers from Canada reported in the *BMJ (British Medical Journal)* (May 2013 issue). The team looked at twelve peer-reviewed articles on observational studies which had been carried out in Europe, the USA and Canada. Experts had long-wondered whether cosmetic breast implants might make it harder to spot malignancy at an early stage, because they produce shadows on mammograms. In this latest study, the authors found that a woman with a cosmetic breast implant has a 25% higher risk of being diagnosed with breast cancer when the disease has already advanced, compared to those with no implants. Women with cosmetic breast implants who are diagnosed with breast cancer have a 38% higher risk of death from the disease, compared to other patients diagnosed with the same disease who have no implants, the researchers wrote. After warning that there were some limitations in the twelve studies they looked at, the authors concluded "Further investigations are warranted

into the long term effects of cosmetic breast implants on the detection and prognosis of breast cancer, adjusting for potential confounders."

1.2.2Treatment

A multidisciplinary team will be involved in a breast cancer patient's treatment. The team may consists of an oncologist, radiologist, specialist cancer surgeon, specialist nurse, pathologist, radiologist, radiographer, and reconstructive surgeon. Sometimes the team may also include an occupational therapist, psychologist, dietitian, and physical therapist. The team will take into account several factors when deciding on the best treatment for the patient, including:

- The type of breast cancer
- The stage and grade of the breast cancer - how large the tumor is, whether or not it has spread, and if so how far
- Whether or not the cancer cells are sensitive to hormones
- The patient's overall health
- The age of the patient
- The patient's own preferences.

The main breast cancer treatment options may include:

- Radiation therapy (radiotherapy)
- Surgery
- Biological therapy (targeted drug therapy)

- Hormone therapy
- Chemotherapy.

1.2.3Surgery

- **Lumpectomy -** surgically removing the tumor and a small margin of healthy tissue around it. In breast cancer, this is often called breast-sparing surgery. This type of surgery may be recommended if the tumor is small and the surgeon believes it will be easy to separate from the tissue around it. British researchers reported that about one fifth of breast cancer patients who choose breast-conserving surgery instead of mastectomy eventually need a reoperation.
- **Mastectomy -** surgically removing the breast. *Simple mastectomy* involves removing the lobules, ducts, fatty tissue, nipple, areola, and some skin. *Radical mastectomy* means also removing muscle of the chest wall and the lymphnodes in the armpit.
- **Sentinel node biopsy -** one lymph node is surgically removed. If the breast cancer has reached a lymph node it can spread further through the lymphatic system into other parts of the body.
- **Axillary lymph node dissection -** if the sentinel node was found to have cancer cells, the surgeon may recommend removing several nymph nodes in the armpit.

- **Breast reconstruction surgery** - a series of surgical procedures aimed at recreating a breast so that it looks as much as possible like the other breast. This procedure may be carried out at the same time as a mastectomy. The surgeon may use a breast implant, or tissue from another part of the patient's body.

Controlled doses of radiation are targeted at the tumor to destroy the cancer cells. Usually, radiotherapy is used after surgery, as well as chemotherapy to kill off any cancer cells that may still be around. Typically, radiation therapy occurs about one month after surgery or chemotherapy. Each session lasts a few minutes; the patient may require three to five sessions per week for three to six weeks. The type of breast cancer the woman has will decide what type of radiation therapy she may have to undergo. In some cases, radiotherapy is not needed. Radiation therapy types include:

- **Breast radiation therapy** - after a lumpectomy, radiation is administered to the remaining breast tissue
- **Chest wall radiation therapy** - this is applied after a mastectomy
- **Breast boost** - a high-dose of radiation therapy is applied to where the tumor was surgically removed. The appearance of the breast may be altered, especially if the patient's breasts are large.
- **Lymph nodes radiation therapy** - the radiation is aimed at the axilla (armpit) and surrounding area to destroy cancer cells that have reached the lymph nodes
- **Breast brachytherapy** - scientists at UC San Diego Moores Cancer Center revealed that patients with early-stage breast cancer in the milk ducts which has not spread, seem to benefit from undergoing breast brachytherapy with a strut-based applicator. This 5-day treatment is given to patients after they have undergone lumpectomy surgery. The researchers found that women who received strut-based breast brachytherapy had lower recurrence rates, as well as fewer and less severe side effects.

- **Side effects of radiation therapy** may include fatigue, lymphedema, darkening of the breast skin, and irritation of the breast skin.

1.2.5Chemotherapy

Medications are used to kill the cancer cells - these are called *cytotoxic* drugs. The oncologist may recommend chemotherapy if there is a high risk of cancer recurrence, or the cancer spreading elsewhere in the body. This is called adjuvant chemotherapy. If the tumors are large, chemotherapy may be administered before surgery. The aim is to shrink the tumor, making its removal easier. This is called neo-adjuvant chemotherapy. Chemotherapy may also be administered if the cancer has metastasized - spread to other parts of the body. Chemotherapy is also useful in reducing some of the symptoms caused by cancer. Chemotherapy may help stop estrogen production. Estrogen can encourage the growth of some breast cancers.

Side effects of chemotherapy may include nausea, vomiting, loss of appetite, fatigue, sore mouth, hair loss, and a slightly higher susceptibility to infections. Many of these side effects can be controlled with medications the doctor can prescribe. Women over 40 may enter early menopause.

Hormonetherapy(hormoneblockingtherapy)

Used for breast cancers that are sensitive to hormones. These types of cancer are often referred to as ER positive (estrogen receptor positive) and PR positive (progesterone receptor positive) cancers. The aim is to prevent cancer recurrence. Hormone blocking therapy is usually used after surgery, but may sometimes be used beforehand to shrink the tumor. If for health

reasons, the patient cannot undergo surgery, chemotherapy or radiotherapy, hormone therapy may be the only treatment she receives. Hormone therapy will have no effect on cancers that are not sensitive to hormones. Hormone therapy usually lasts up to five years after surgery.

The following hormone therapy medications may be used:

Tamoxifen - prevents estrogen from binding to ER-positive cancer cells. Side effects may include changes in periods, hot flashes, weight gain, headaches, nausea, vomiting, fatigue, and aching joints. A biomarker in breast cancer patients who do not respond, or who have become resistant to Tamoxifen has been discovered by researchers at the University of Manchester, England. They say that their discovery will help doctors decide which patients are suitable or not for adjuvant (complementary) hormone therapy with Tamoxifen. Biomarker may predict breast cancer recurrence after Tamoxifen - scientists from the Cancer Center and Department of Pathology at Massachusetts General Hospital, Boston, say that it may be possible to predict which women will have a higher risk of cancer recurrence after completing tamoxifen treatment. The biomarker measures the ratio of gene expression in the HOXB13 and IL17BR genes.

Aromatase inhibitors - this type of medication may be offered to women who have been through the menopause. It blocks aromatase. Aromatase helps estrogen production after the menopause. Before the menopause, a woman's ovaries produce estrogen. Examples of aromatase inhibitors include letrozole, exemestane, and anastrozole. Side effects may include nausea, vomiting, fatigue, skin rashes, headaches, bone pain, aching joints, loss of libido, sweats, and hot flashes.

Ovarian ablation or suppression - pre-menopausal women produce estrogen in their ovaries. Ovarian ablation or suppression stop the ovaries from producing estrogen. Ablation is done either through surgery or radiation therapy - the woman's ovaries will never work again, and she will enter the menopause early. A luteinising hormone-releasing hormone aganist drug called Goserelin will suppress the ovaries. The patient's periods will stop during treatment, but will start again when she stops taking Goserelin. Women of menopausal age (about 50 years) will probably never start having periods again. Side effects may include mood changes, sleeping problems, sweats, and hot flashes.

1.2.6Biological treatment (targeted drugs)

Trastuzumab (Herceptin) - this monoclonal antibody targets and destroys cancer cells that are HER2-positive. Some breast cancer cells produce large amounts of HER2 (growth factor receptor 2); Herceptin targets this protein. Possible side effects may include skin rashes, headaches, and/or heart damage.

Lapatinib (Tykerb) - this drug targets the HER2 protein. It is also used for the treatment of advanced metastatic breast cancer. Tykerb is used on patients who did not respond well to Herceptin. Side effects include painful hands, painful feet, skin rashes, mouth sores, extreme tiredness, diarrhea, vomiting, and nausea.

Bevacizumab (Avastin) - stops the cancer cells from attracting new blood vessels, effectively causing the tumor to be starved of nutrients and

oxygen. Side effects may include congestive heart failure, hypertension (high blood pressure), kidney damage, heart damage, blood clots, headaches, mouth sores. Although not approved by the FDA for this use, doctors may prescribe it "off-label". Using this drug for breast cancer is controversial. In 2011, the FDA said that Avastin is neither effective nor safe for breast cancer. Swiss researchers found that Avastin offers only a modest benefit regarding disease progression in women with advanced stage breast cancer. They added that it has no impact on survival.

Lowdoseaspirin-Research carried out on laboratory mice and test tubes suggests that regular low-dose aspirin may halt the growth and spread of breast cancer. Scientists from the Veterans Affairs Medical Center in Kansas City and the University of Kansas Medical Center explained that their tests on cancer lines and in mice showed that aspirin not only slowed the growth of cancer cells and shrank tumors considerably, but also stopped metastasis (cancer spreading to new sites). Their research involved assessing aspirin's effects on two types of cancer, including the aggressive "triple-negative" breast cancer which is resistant to most current treatments. Cancer campaigners cautioned that although the current results show great promise, this research is at a very early stage and has yet to be shown to be effective on humans.

Figure 1: Breast cancer symbol

The goal of breast cancer awareness campaigns is to raise the public's "brand awareness" for breast cancer, its detection, its treatment, and the need for a reliable, permanent cure. Increased awareness has increased the number of women receiving mammograms, the number of breast cancers detected, and the number of women receiving biopsies. Overall, as a result of awareness, breast cancers are being detected at an earlier, more treatable stage. Awareness efforts have successfully utilized marketing approaches to reduce the stigma associated with the disease. Generally speaking, breast cancer awareness campaigns have been highly effective in getting attention for the disease. Breast cancer receives significantly more media coverage than other prevalent cancers. Breast cancer advocacy uses the pink ribbon and the color pink as a concept brand to raise money and increase screening. The breast cancer brand is strong: people who support the "pink brand" are members of the socially aware niche market, who are in favor of improved lives for women, believe in positive thinking, trust biomedical science to be able to solve any problem if given enough money, and prefer curative treatments to prevention. The brand ties together fear of cancer, hope for early identification and successful treatment, and the moral goodness of women with breast cancer and anyone who visibly identifies themselves with breast cancer patients. This brand permits and even encourages people to substitute conscientious consumption and individual

symbolic actions, like buying or wearing a pink ribbon, for concrete, practical results, such as collective political action aimed at discovering non-genetic causes of breast cancer. The establishment of the brand and the entrenchment of the breast cancer movement have been uniquely successful, because no countermovement opposes the breast cancer movement or believes that breast cancer is desirable.

1.2.7 Mammogram

Mammography is the process of using low-energy X-rays (usually around 30 kVp) to examine the human breast, which is used as a diagnostic and screening tool. The goal of mammography is the early detection of breast cancer, typically through detection of characteristic masses and/or microcalcifications.

Like all X-rays, mammograms use doses of ionizing radiation to create images. Radiologists then analyze the images for any abnormal findings. It is normal to use lower-energy X-rays (typically Mo-K) than those used for radiography of bones. Ultrasound, ductography, positron emission mammography (PEM), and magnetic resonance imaging (MRI) are adjuncts to mammography. Ultrasound is typically used for further evaluation of masses found on mammography or palpable masses not seen on mammograms. Ductograms are still used in some institutions for evaluation of bloody nipple discharge when the mammogram is non-diagnostic. MRI can be useful for further evaluation of questionable findings as well as for screening pre-surgical evaluation in patients with known breast cancer to detect any additional lesions that might change the surgical approach, for instance from breast-conserving lumpectomy to mastectomy. Other procedures being investigated include tomosynthesis.

For the average woman, the U.S. Preventive Services Task Force recommended (2009) mammography every two years in women between the ages of 50 and 74. The American College of Radiology and American Cancer Society recommend yearly screening mammography starting at age 40. The Canadian Task Force on Preventive Health Care (2012) and the European Cancer Observatory (2011) recommends mammography every 2–3 years between 50 and 69. These task force reports point out that in addition to unnecessary surgery and anxiety, the risks of more frequent mammograms include a small but significant increase in breast cancer induced by radiation. The Cochrane Collaboration (2013) concluded that the trials with adequate randomisation did not find an effect of mammography screening on total cancer mortality, including breast cancer, after 10 years. The authors of systematic review write: "If we assume that screening reduces breast cancer mortality by 15% and that overdiagnosis and overtreatment is at 30%, it means that for every 2000 women invited for screening throughout 10 years, one will avoid dying of breast cancer and 10 healthy women, who would not have been diagnosed if there had not been screening, will be treated unnecessarily. Furthermore, more than 200 women will experience important psychological distress including anxiety and uncertainty for years because of false positive findings." The authors conclude that the time has come to re-assess whether universal mammography screening should be recommended for any age group.They thus state that universal screening may not be reasonable. The Nordic Cochrane Collection, which in 2012 reviews updated research to state that advances in diagnosis and treatment make mammography screening less effective today. They state screening is “no longer effective.” They conclude that “it therefore no longer seems reasonable to attend” for breast cancer screening at any age, and warn of misleading information on the internet.

Mammography has a false-negative (missed cancer) rate of at least 10 percent. This is partly due to dense tissues obscuring the cancer and the fact that the appearance of cancer on mammograms has a large overlap with the appearance of normal tissues. A meta-analysis review of programs in countries with organized screening found 52% over-diagnosis.

CHAPTER 2

IMAGE PROCESSING

2.1 INTRODUCTION OF IMAGE PROCESSING:

Preprocessing algorithms frequently form the first processing step after capturing the image. Image preprocessing typically denotes a processing step transforming a source image into a new image which is fundamentally similar to the source image, but differs in certain aspects, e.g. improved contrast. Preprocessing is also defined as changing the brightness of individual image pixels. The preprocessing functions can be divided into two basic groups, depending on what the resulting brightness value of a pixel in the output image.

Pixel operations:

Pixel operations compute the brightness of a pixel in the output image exclusively from the brightness of the corresponding pixel in the source image. This group also encompasses image arithmetic functions which combine several images, because these functions only use a single pixel on a fixed position from each of the source images. Pixel operations can be further divided into homogeneous and inhomogeneous pixel operations. Homogeneous operations use the same transformation function for each pixel, for inhomogeneous ones the transformation function depends on the location of the pixel in the image

Local operations:

Local operations take a certain neighborhood of the current pixel into account when computing the brightness of the corresponding output image pixel. An example is the mean value filter, which sets the brightness of the output image pixel

to the average brightness of a small neighborhood of the corresponding point in the source image.

In imaging science, image processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it. Image processing usually refers to digital image processing, but optical and analog image processing also are possible. This article is about general techniques that apply to all of them. The acquisition of images (producing the input image in the first place) is referred to as imaging. Closely related to image processing are computer graphics and computer vision. In computer graphics, images are manually made from physical models of objects, environments, and lighting, instead of being acquired (via imaging devices such as cameras) from natural scenes, as in most animated movies. Computer vision, on the other hand, is often considered high-level image processing out of which a machine/computer/software intends to decipher the physical contents of an image or a sequence of images (e.g., videos or 3D full-body magnetic resonance scans). In modern sciences and technologies, images also gain much broader scopes due to the ever growing importance of scientific visualization (of often large-scale complex scientific/experimental data). Examples include microarray data in genetic research, or real-time multi-asset portfolio trading in finance.

2.2 Goal of image preprocessing:

- Enhance the visual appearance of images.
- Improve the manipulation of datasets.

2.3 Pre-processing involves

- Image resampling

  Reduce or increase the number of pixels of the dataset.

- Greyscale contrast enhancement

  Improve the visualization by brightening the dataset.

- Noise removal

  Noises are removed using filters

- Mathematical operations

  To enhance particular features by arithmetical operation or by morphological operations

- Manual correction

  Fine tune of an image is done by editing.

CHAPTER 3

TECHNIQUES OF FEATURE EXTRACTION

3.1 FEATURE EXTRACTION

The features are important for every classification algorithms. A typical mammogram contains a vast amount of heterogeneous information that depicts different tissues, vessels, ducts, chest skin, breast edge, the film, and the Xray machine characteristics. Feature extraction is done by different techniques.

3.1.1 GLCM

The GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image**.** Grey-Level Co-occurrence Matrix texture measurements have been the workhorse of image texture since they were proposed by Haralick in the 1970s. To many image analysts, they are a button you push in the software that yields a band whose use improves classification - or not. The original works are necessarily condensed and mathematical, making the process difficult to understand for front-line image analyst. This GLCM texture was developed to help such people, and it has been used extensively world-wide since 1999.Texture is one of the important characteristics used in identifying objects or regions of interest in an image. Texture contains important information about the structural arrangement of surfaces. The textural features based on gray-tone spatial dependencies have a general applicability in image classification. The three fundamental pattern elements used in human interpretation of images are spectral, textural and contextual features. Spectral features describe the average tonal variations in various bands of the visible and/or infrared portion of an

electromagnetic spectrum. Textural features contain information about the spatial distribution of tonal variations within a band. The fourteen textural features proposed by Haralick contain information about image texture characteristics such as homogeneity, gray-tone linear dependencies, contrast, number and nature of boundaries present and the complexity of the image. Contextual features contain information derived from blocks of pictorial data surrounding the area being analyzed. Haralick first introduced the use of co-occurrence probabilities using GLCM for extracting various texture features. GLCM is also called as *Gray level Dependency Matrix*. It is defined as "A two dimensional histogram of gray levels for a pair of pixels, which are separated by a fixed spatial relationship." GLCM of an image is computed using a displacement vector d, defined by its radius δ and orientation θ.

The features are important for every classification algorithms. Here texture features of images are extracted. The GLCMs features are stored in ai × j ×n matrix, where n is the number of GLCMs calculated usually due to the different orientation and displacements used in the algorithm. Usually the values i and j are equal to 'NumLevels' parameter of the GLCM computing function. Note that matlab quantization values belong to the set {1,...,NumLevels} and not from {0,...,(NumLevels-1)} as provided.

GLCM Directions of Analysis:

1. Horizontal (0)
2. Vertical (90)
3. Diagonal:
    a.) Bottom left to top right (-45)

b) Top left to bottom right (-135)

Denoted as P0, P45, P90, & P135 Respectively.

Ex. P0(i , j).

Creating a texture image:

The result of a texture calculation is a single number representing the entire window. This number is put in the place of the centre pixel of the window, then the window is moved one pixel and the process is repeated for calculating a new GLCM and a new texture measure. In this way an entire image is built up of texture values.

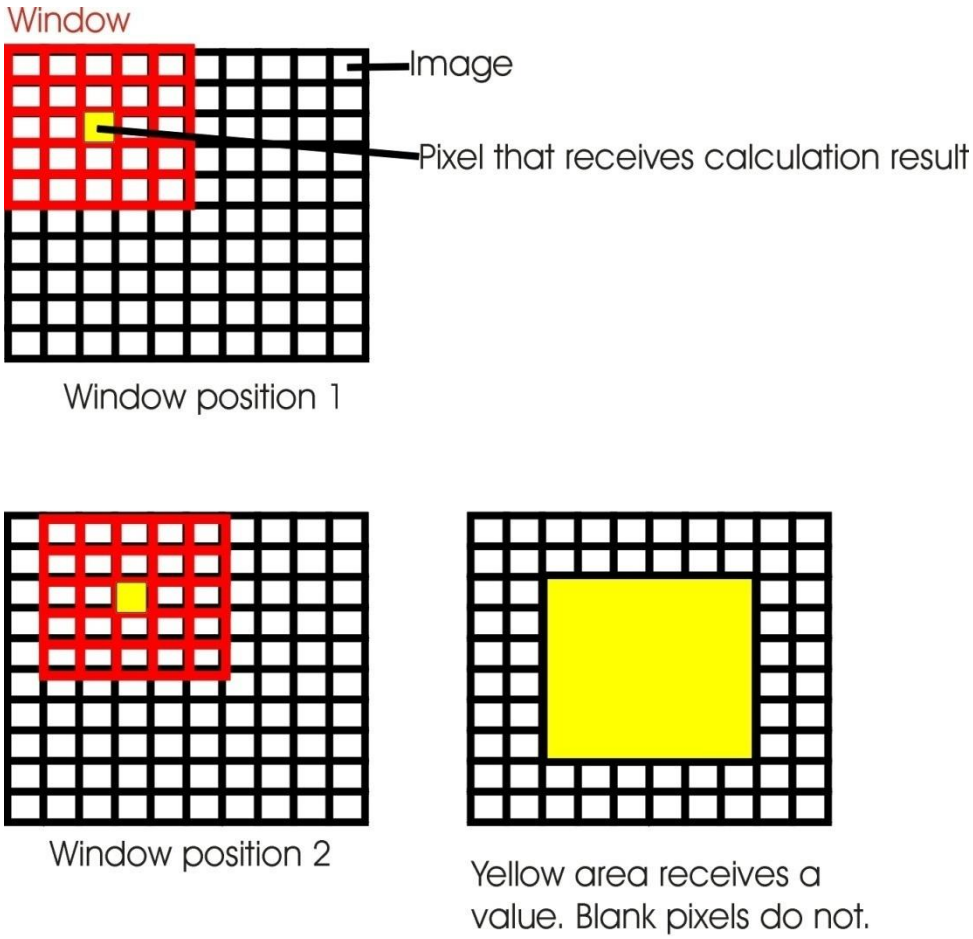


Figure 2: GLCM windows

Extracted GLCM features :

- Autocorrelation
- Contrast
- Correlation
- Correlation
- Cluster Prominence
- Cluster Shade
- Dissimilarity
- Energy

- Entropy
- Homogeneity
- Homogeneity
- Maximum probability
- Sum of squares
- Sum average
- Sum variance
- Sum entropy
- Difference variance
- Difference entropy
- Information measure of correlation1 and 2
- Inverse difference (INV)
- Inverse difference normalized (INN)
- Inverse difference moment

MATHEMATICAL FORMULAS :

Correlation:

$$\sum_{i,j=0}^{N-1} P_{i,j} \left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right]$$

Variance:

$$\sigma_i^2 = \sum_{i,j=0}^{N-1} P_{i,j}\left(i-\mu_i\right)^2 \quad \sigma_j^2 = \sum_{i,j=0}^{N-1} P_{i,j}\left(j-\mu_j\right)^2$$

Standard Deviation:

$$\sigma_i = \sqrt{\sigma_i^2} \qquad \sigma_j = \sqrt{\sigma_j^2}$$

Mean:

$$\mu_i = \sum_{i,j=0}^{N-1} i\left(P_{i,j}\right) \qquad \mu_j = \sum_{i,j=0}^{N-1} j\left(P_{i,j}\right)$$

Entropy:

$$\sum_{i,j=0}^{N-1} P_{i,j}\left(-\ln P_{i,j}\right)$$

Homogenity:

$$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+\left(i-j\right)^2}$$

Dissimilarity:

$$\sum_{i,j=0}^{N-1} P_{i,j}|i - j|$$

Contrast:

$$\sum_{i,j=0}^{N-1} P_{i,j}\left(i-j\right)^2$$

Angular Moment:

$$\sum_{i,j=0}^{N-1} P_{i,j}{}^2$$

Energy:

$$Energy = \sqrt{ASM}$$

3.1.2 GABOR FILTER

Gabor features are a common choice for texture analysis. In image processing, a Gabor filter, shown in figure 2, named after Dennis Gabor, is a linear filter used for edge detection. Frequency and orientation representations of Gabor filters are similar to those of the human visual system, and they have been found to be particularly appropriate for texture representation and discrimination. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. Simple cells in the visual cortex of mammalian brains can be modeled by Gabor functions. Thus, image analysis with Gabor filters is thought to

be similar to perception in the human visual system. Its impulse response is defined by a sinusoidal wave (a plane wave for 2D Gabor filters) multiplied by a Gaussian function. Because of the multiplication-convolution property (Convolution theorem), the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function. The filter has a real and an imaginary component representing orthogonal directions. The two components may be formed into a complex number or used individually.

Complex

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right) exp\left(i\left(\frac{2\pi x'}{\lambda}+\psi\right)\right) \quad (1)$$

Real

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right) cos\left(\frac{2\pi x'}{\lambda}+\psi\right) \quad (2)$$

Imaginary

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right) sin\left(\frac{2\pi x'}{\lambda}+\psi\right) \quad (3)$$

$$x' = xcos\theta + ysin\theta \quad (4)$$

$$y' = -xcos\theta + ysin\theta \quad (5)$$

Where, λ is the wavelength of sinusoidal factor

$\theta$is the orientation of the normal to the parallel strips of Gabor functions

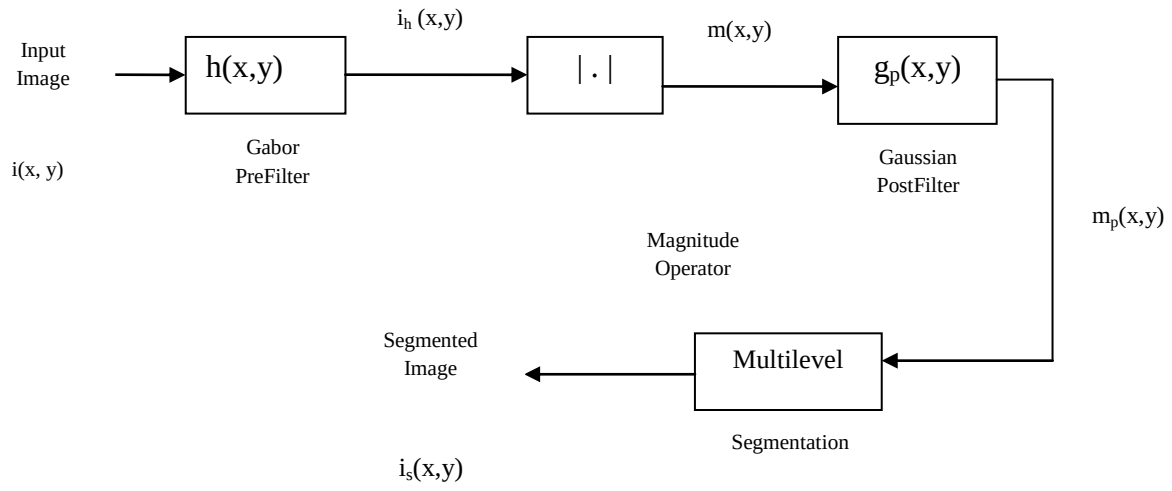


Figure 3: Overview of Gabor-Filter Scheme

A set of Gabor filters with different frequencies and orientations may be helpful for extracting useful features from an image. Gabor filters have been widely used in pattern analysis applications. For example, it has been used to study the directionality distribution inside the porous spongy trabecular bone in the spine. Gabor filters are directly related to Gabor wavelets, since they can be designed for a number of dilations and rotations. However, in general, expansion is not applied for Gabor wavelets, since this requires computation of bi-orthogonal wavelets, which may be very time-consuming. Therefore, usually, a filter bank consisting of Gabor filters with various scales and rotations is created.

The filters are convolved with the signal, resulting in a so-called Gabor space. This process is closely related to processes in the primary visual cortex. Jones and Palmer showed that the real part of the complex Gabor function is a good fit to the receptive field weight functions found in simple cells in a cat's striate cortex. The Gabor space is very useful in image processing applications such as optical character recognition, iris recognition and fingerprint recognition. Relations

between activations for a specific spatial location are very distinctive between objects in an image. Furthermore, important activations can be extracted from the Gabor space in order to create a sparse object representation.

3.1.3 STEERABLE PYRAMID

The Steerable Pyramid is a linear multi-scale, multi-orientation image decomposition that provides a useful front-end for image-processing and computer vision applications. We developed this representation in 1990, in order to overcome the limitations of orthogonal separable wavelet decompositions that were then becoming popular for image processing (specifically, those representations are heavily aliased, and do not represent oblique orientations well). Once the orthogonality constraint is dropped, it makes sense to completely reconsider the filter design problem (as opposed to just re-using orthogonal wavelet filters in a redundant representation, as is done in cycle-spinning or undecimated wavelet transforms).

A steerable pyramid, shown in figure 3, is an implementation of a multi-scale, multi-orientation band-pass filter bank used for applications including image compression, texture synthesis, and object recognition. It can be thought of as an orientation selective version of a Laplacian pyramid, in which a bank of steerable filters are used at each level of the pyramid instead of a single Laplacian of Gaussian filter. The basis functions of the steerable pyramid are Kth-order directional derivative operators (for any choice of K), that come in different sizes and K+1 orientations. As directional derivatives, they span a rotation-invariant subspace, and they are designed and sampled such that the whole

transform forms a tight frame. An example decomposition of an image of a white disk on a black background is shown to the right. This particular steerable pyramid contains 4 orientation sub bands, at 2 scales. The smallest sub band is the residual lowpass information.

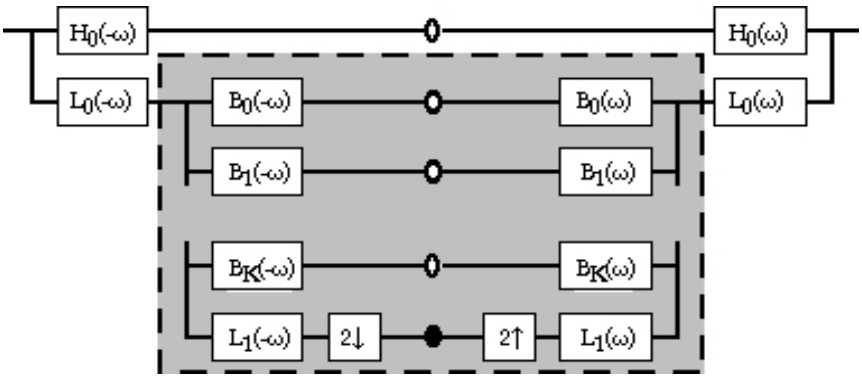


Figure 4: Steerable Pyramid

The block diagram for the decomposition (both analysis and synthesis) is shown to the right. Initially, the image is separated into low and highpass sub bands, using filters L0 and H0. The lowpass sub band is then divided into a set of oriented bandpass sub bands and a low(er)-pass sub band. This low(er)-pass sub band is subsampled by a factor of 2 in the X and Y directions. The recursive (pyramid) construction of a pyramid is achieved by inserting a copy of the shaded portion of the diagram at the location of the solid circle (i.e., the lowpass branch).

Advantages of Steerable Pyramid

The steerable pyramid performs

- polar-separable decomposition in the frequency domain,
- Allowing independent representation of scale and orientation.
- Since it is a tight frame, it obeys the generalized form of Parseval's Equality: The vector-length (L2-norm) of the coefficients equals that of the original signal.

ii. Applications of Steerable pyramid

- orientation analysis
- noise removal and enhancement
- transient detection
- texture representation and synthesis.

3.2 SVM CLASSIFICATION

Let us consider a supervised binary classification problem. If the training data are represented by $\{x_i, y_i\}, i = 1,2, \dots, N$ and $y_i \epsilon \{-1, +1\}$, where N is the number of training samples, $y_i = +1$ for class $\omega 1$ and $y_i = -1$ for class $\omega 2$. Suppose the two classes are linearly separable. This means that it is possible to find at least one hyperplane defined by a vectorwith a bias w_0 , which can separate the classes without error:

$$f(x) = w.x + w_0 \quad (6)$$

To find such a hyperplane, w and w_0 should be estimated in a way that $y_i(w.x_i + w_o) \geq +1$ for $y_i = +1$ (class ω_1) and $y_i(w.x_i + w_o) \leq -1$ for $y_i = -1$ (class ω_2). These two, can combined to provide equation 7:

$$y_i(w.x_i + w_0) - 1 \geq 0 \quad (7)$$

Many hyperplanes could be fitted to separate the two classes but there is only one optimal hyperplane that is expected to generalize better than other hyperplanes.

The goal is to search for the hyperplane that leaves the maximum margin between classes. To be able to find the optimal hyperplane, the support vectors must be defined. The support vectors lie on two hyperplanes which are parallel to the optimal and are given by:

$$w.x_i + w_0 = +1 \quad (8)$$

If a simple rescale of the hyperplane parameters w and w0 takes place, the margin can be expressed as $\frac{2}{\|w\|}$ .The optimal hyperplane can be found by solving the following optimization problem:

$$Minimize: \frac{1}{2}\|w\|^2 \quad (9)$$

Subject to: $y_i(w.x_i + w_0) - 1 \geq 0 \; i = 0,1,..,N$

Using a Lagrangian formulation, the above problem can be translated to:

$$Maximize: \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i.j=1}^{N} \lambda_i \lambda_j y_i y_j (x_i.x_j) \quad (10)$$

$$subject\ to: \sum_{i=1}^{N} \lambda_i y_i = 0\ and\ \lambda_i \geq 0, i = 1,2,..,N \quad (11)$$

Where λ_i are the Lagrange multipliers

Under this formulation, the optimal hyperplane discriminant function becomes:

$$f(x) = \sum_{i \in s} \lambda_i y_i (x_i x) + w_0 \quad (12)$$

Where is a subset of training samples that correspond to nonzero Lagrange multipliers. These training samples are called support vectors. In most cases, classes are not linearly separable, and the constrain of equation 7 cannot be satisfied. In order to handle such cases, a cost function can be formulated to combine

Maximization of margin and minimization of error criteria, using a set of variables called slack variables ξ. This cost function is defined as:

$$Minimize: J(w, w_0, \xi) = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{N} \xi_i \quad (13)$$

$$subject\ to: y_i(w.x + w_0) \geq 1 - \xi_i \quad (14)$$

To generalize the above method to non-linear discriminate functions, the Support Vector Machine maps the input vector x into a high-dimensional feature space and then constructs the optimal separating hyperplane in that space.

One would consider that mapping into a high dimensional feature space would add extra complexity to the problem. But, according to the Mercer's

theorem, the inner product of the vectors in the mapping space, can be expressed as a function of the inner products of the corresponding vectors in the original space.

The inner product operation has an equivalent representation: Maximize:

$$\sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i,j=1}^{N} \lambda_i \lambda_j y_i y_j K(x_i . x_j) \quad (15)$$

Subject

$$\sum_{i=1}^{N} \lambda_i y_i = 0 \; and \; \lambda_i \geq 0, i = 1,2, \dots, N$$

The resolution classifier becomes:

$$f(x) = \sum_{i \in S} \lambda_i y_i K(x_i x) + w_0 \quad (16)$$

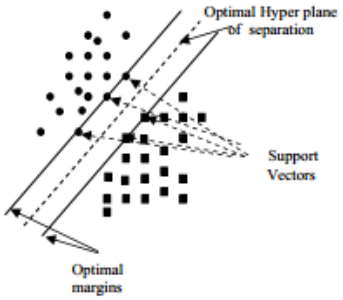


Figure 5: Linear separable classes

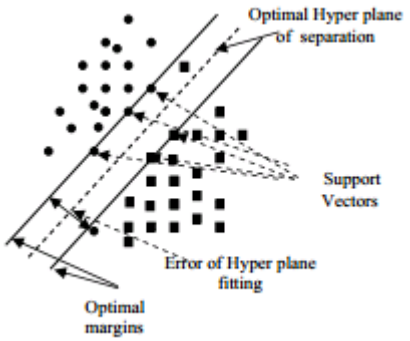


Figure 6: Non linear separable classes

CHAPTER 4

PROPOSED METHOD

Ten mammogram images, among five are benign and five are malignant are taken to describe the accuracy of the images. Feature extraction is performed for texture analysis using the algorithms: Gabor filter, Steerable Pyramid, Grey Level Co-occurrence Matrix (GLCM) . Using feature extraction classification is done using Support Vector Machine (SVM) method with the combination of different texture descriptors.

The combinations are

- GLCM, Gabor filter
- Gabor filter ,Steerable Pyramid
- GLCM, Gabor filter, Steerable Pyramid

These combinations are to determine the accuracy.

CHAPTER 5

FUTURE WORK

Scale Invariant Feature Transform (SIFT) is another method used for feature extraction of mammogram. The future scope of this project is to extract the features of mammogram using scale invariant feature transform and to check whether it is advantageous over these three feature extraction methods.

Features of SIFT

- Bins in feature space are searched in order of their closest distance from the query location (priority queue)
- Only the first x bins are tested
- Returns the closest neighbor with high probability
- Drastic increase in speed

Advantages of SIFT

- robust against image rotation
- scaling
- substantial range of affine distortion
- addition of noise
- change in illumination

CONCLUSION

Breast cancer has become one of the major complication in medical world and the reason behind th ehigh mortality rate of breast cancer is lack of awareness and late diagnosis of malignant masses. The above explained method of feature extraction helps in easy and early detection of suspicious masses. The major advantages of this method is that it is a purely computational method , the delay of the result is highly reduced and the accuracy of the result is comparatively high.

CHAPTER 6
APPENDIX

6.1MATLAB SOFTWARE

MATLAB is a fourth generation programming language and a multi-paradigm numerical computing environment used by millions of engineers and scientists worldwide.It lets you to explore and visualize your ideas and collaborate across disciplines including signalprocessing, imageprocessing, communications, control systems and computational finance.It is used mainly for developing applications.

The special feature of this matlabis,we can interface with the programs written in other languages such as C,C++,Java,Fortran and Python.Matlab allows matrix manipulations,plotting of functions and implementation of algorithms.MATLAB is widely used in academic and research institutions as well as in industrial enterprises.

6.2MATLAB in Image processing:

Image Processing Toolbox provides a comprehensive set of reference-standard algorithms, functions, and apps for image processing, analysis, visualization, and algorithm development. You can perform image analysis, image segmentation, image enhancement, noise reduction, geometric transformations, and image registration. Many toolbox functions support multicore processors, GPUs, and C-code generation. Image Processing Toolbox supports a diverse set of image types, including high dynamic range, gigapixel resolution, embedded ICC profile, and tomographic. Visualization functions and apps let you explore images and

videos, examine a region of pixels, adjust color and contrast, create contours or histograms, and manipulate regions of interest (ROIs). The toolbox supports workflows for processing, displaying, and navigating large images.

Key Feature

- Image analysis, including segmentation, morphology, statistics, and measurement
- Image enhancement, filtering, and deblurring
- Geometric transformations and intensity-based image registration methods
- Image transforms, including FFT, DCT, Radon, and fan-beam projection
- Large image workflows, including block processing, tiling, and multiresolution display
- Visualization apps, including Image Viewer and Video Viewer
- Multicore- and GPU-enabled functions, and C-code generation support

6.3Standard and Specialized File Formats

MATLAB supports standard data and image formats including:

- AVI
- JPEG
- JPEG-2000
- FITS
- HDF
- HDF-EOS
- M4V
- MP4
- PNG

- TIFF
- ASCII
- Binary files
- Microsoft Excel

It also supports the multiband image formats BIP and BIL, as used by LANDSAT. Low-level I/O and memory mapping functions enable you to develop custom routines for working with any data format.

Image Processing Toolbox supports a number of specialized image file formats. For medical images, it supports DICOM files, including associated metadata, as well as the Analyze 7.5 and Interfile formats. The toolbox can also read geospatial images in NITF files and high dynamic range images in HDR files.

6.4Image Enhancement:

Image enhancement techniques in Image Processing Toolbox enable you to increase the signal-to-noise ratio and accentuate image features by modifying the colors or intensities of an image.

The toolbox includes specialized filtering routines and a generalized multidimensional filtering function that handles integer image types, offers multiple boundary-padding options, and performs convolution and correlation.

Using predefined filters and functions you can:

- Filter with morphological operators
- Deblur and sharpen

- Remove noise with linear, median, or adaptive filtering
- Perform histogram equalization
- Remap the dynamic range
- Adjust the gamma value
- Adjust contrast

6.5Image Analysis:

Image analysis is the process of extracting meaningful information from images such as finding shapes, counting objects, identifying colors, or measuring object properties. Image Processing Toolbox provides a comprehensive suite of reference-standard algorithms and visualization functions for image analysis tasks such as statistical analysis, feature extraction, and property measurement.

6.6Image Transforms:

Image transforms play a critical role in many image processing tasks, including image enhancement, analysis, restoration, and compression. Image Processing Toolbox provides several image transforms, including Hough, Radon, FFT, DCT, and fan-beam projections. You can reconstruct images from parallel-beam and fan-beam projection data (common in tomography applications).

Advantages of MATLAB in Image processing:

- A very large database of built-in algorithms immediately without recompilation and it is a good base for developing algorithms.
- MATLAB has the ability to call external libraries.

- It can process both the images and videos.
- It has the ability to autogenerate C code using MATLAB Coder.
- Can read both common and domain-specific image formats.

CHAPTER 7

IMAGE OUTPUT FOR MATLAB CODE

Gabor filter orientation for v=2

Gabor filter orientation for v=4

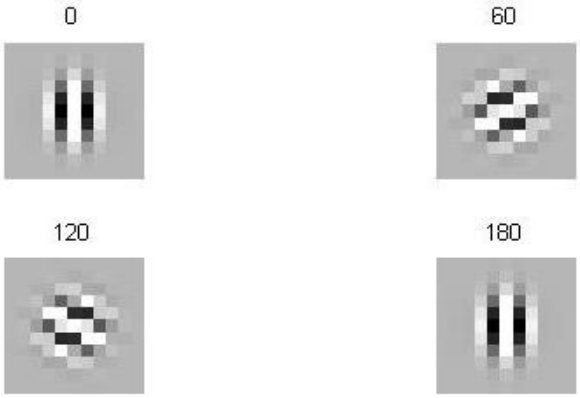
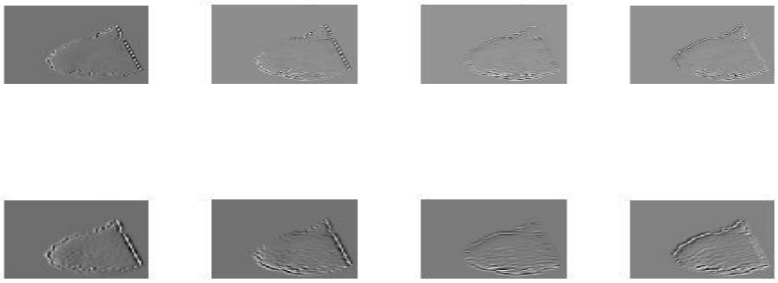
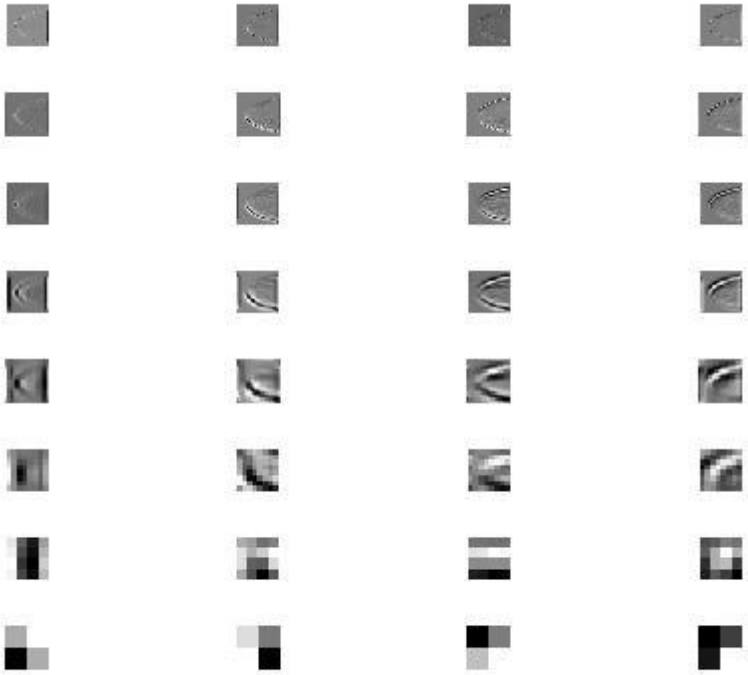


Image through lowpass and highpass filters



Steerable pyramid mean and variance extraction



CHAPTER 8

References

[1] Hossain, S. and Serikawa, S. (2012) Features for Texture Analysis. SICE Annual Conference, 1739-1744.

[2] Roslan, R. and Jamil, N. (2012) Texture Feature Extraction Using 2-D Gabor Filters. International Symposium on Computer Applications and Industrial Electronics, 3-4 December 2012, Kota Kinabalu, 173-178.

[3] Reddy, T. and Kumaravel, N. (2012) A Comparison of Wavelet, Curvelet and Contourlet Based Texture Classification

Algorithms for Characterization of Bone Quality in Dental CT. 2011 International Conference on Environmental, Biomedical and Biotechnology, **16**, 60-65.

[4] Randen, T. and Husøy, J. (1999) Filtering for Texture Classification: A Comparative Study. IEEE Transactions on Pattern Analysis and Machine Intelligence, **21**, 291-310. http://dx.doi.org/10.1109/34.761261

[5] Sharma, M., et al. (1980) Evaluation of Texture Methods for Image Analysis. Proceedings of the 7th Australian and New Zealand Intelligent Information Systems Conference.

[6] Kumar, V., et al. (2007) An Innovative Technique of Texture Classification and Comparison Based on Long Linear Patterns. Journal of Computer Science, 633-638.

A. Barley, C. Town 76

[7] Sumana, I., et al. (2012) Comparison of Curvelet and Wavelet Texture Features for Content Based Image Retrieval. IEEE International Conference on Multimedia and Expo, 290-295.

[8] Zhang, J., et al. (2012) Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study. International Journal of Computer Vision, 1739-1744.

[9] Janney, P. and Geers, G. (2010) Texture Classification Using Invariant Features of Local Textures. IET Image Processing, **4**, 158-171. http://dx.doi.org/10.1049/iet-ipr.2008.0229

[10] Do, T., Aikala, A. and Saarela, O. (2012) Framework for Texture Classification and Retrieval Using Scale Invariant Feature Transform. Ninth International Joint Conference on Computer Science and Software Engineering, 30 May-1 June 2012, Bangkok, 289-293.

[11] Yang, Y. and Newsam, S. (2008) Comparing SIFT Descriptors and Gabor Texture Features for Classification of Remote Sensed Imagery. IEEE International Conference on Image Processing, San Diego, 12-15 October 2008, 1852-1855.

[12] Mustafa, M., Taib, M.N., Murat, Z.H. and Hamid, N.H.A. (2010) GLCM Texture Classification for EEG Spectrogram Image. IEEE EMBS Conference on Biomedical Engineering & Sciences, Kuala Lumpur, 30 November-2 December 2010, 373-376.

[13] Do, M. and Vetterli, M. (2002) Rotation Invariant Texture Characterization and Retrieval Using Steerable Wavelet- Domain Hidden Markov Models. IEEE Transactions on Multimedia, **4**, 517-527.
http://dx.doi.org/10.1109/TMM.2002.802019

[14] Burt, P. and Adelson, E. (1983) The Laplacian Pyramid as a Compact Image Code. IEEE Transactions on Communications,
31, 532-540. http://dx.doi.org/10.1109/TCOM.1983.1095851

[15] Greenspan, H., Belongie, S., Goodman, R., Perona, P., Rakshit, S. and Anderson, C.H. (1994) Overcomplete SteerabPyramid Filters and Rotation Invariance. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,
Seattle, 21-23 June 1994, 222-228. http://dx.doi.org/10.1109/CVPR.1994.323833

[16] Lowe, D. (2004) Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision,
60, 91-110. http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94