



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO
DIVISIÓN ACADÉMICA DE INFORMÁTICA Y SISTEMAS



CREACIÓN DE UN BANCO DE DATOS DE MAMOGRAMAS PREPROCESADOS

Trabajo recepcional bajo la modalidad de tesis
que para obtener el grado de

Licenciado en Sistemas Computacionales

Presenta

Omar Trinidad Gutiérrez Méndez

Directores

**Dra. Juana Canul Reich
M. en C. Guillermo de los Santos Torres**

Cunduacán, Tabasco

Agosto de 2013



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO
DIVISIÓN ACADÉMICA DE INFORMÁTICA Y SISTEMAS



CREACIÓN DE UN BANCO DE DATOS DE MAMOGRAMAS PREPROCESADOS

Trabajo recepcional bajo la modalidad de tesis
que para obtener el grado de

Licenciado en Sistemas Computacionales

Presenta

Omar Trinidad Gutiérrez Méndez

Directores

**Dra. Juana Canul Reich
M. en C. Guillermo de los Santos Torres**

Revisores

**M. en C. José Luis
Lic. Martha Patricia Silva Payró
M. en C. Cristina López Ramírez
Julian Assange
Ph. D. Nikola Tesla**

Agradecimientos

A Guido van Rossum por inventar Python.

Dedicatoria

how can I write dedicatoria in English?

Índice general

Capítulo I. Generalidades

1.1	Antecedentes	1
1.2	Problema	2
1.2.1	Planteamiento del problema	2
1.2.2	Delimitation of the investigation	2
1.2.3	Research questions	3
1.3	Objectives	3
1.3.1	General objective	3
1.3.2	Specific objectives	3
1.4	Justification	3

Capítulo II. Theoretical framework

2.1	Frame of reference	4
2.1.1	Bith depths	4
2.2	Conceptual framework	4
2.2.1	Mamogramas	5

Capítulo III. Resultados

Capítulo IV. Conclusiones

4.1	Trabajos futuros	8
-----	------------------	---

Índice de figuras

II.1 Comparison between 12 and 16 bits images	6
---	---

Capítulo I. Generalidades

1.1. Antecedentes

El cáncer de mama es un grave problema de salud pública, es una de las principales causas de decesos en la población femenina de Tabasco y México. Una forma de aminorar estas muertes es la detección temprana de este padecimiento, a su vez, el método más comúnmente empleado en su detección es el estudio de mamografías, tarea a cargo de radiólogos; hay un margen de error en la opinión de los radiólogos.

Es posible incrementar el éxito de los diagnósticos mejorando la apariencia (calidad) de la imagen, *introducir el preprocesamiento y escribir sobre los sistemas CAD-

-El estado del arte del preprocesamiento de mamografías está avanzado-, sin embargo, -es posible seguir contribuyendo a esta área del conocimiento-.

Dos bancos de datos similares al que se menciona en este trabajo son el Digital Database for Screening Mammography (DDSM) de la University of South Florida (USF) [Heath et al., 2000] y el banco de datos MIAS MiniMammographic Database (mini-MIAS) [suc,]. Ambos son ampliamente usados por la comunidad de investigación sobre procesamiento y análisis de imágenes. El objetivo de este trabajo es crear un banco de datos similar a los que se mencionan arriba.

Both of them are widely used by the mammographic image analysis research community. Our main goal is to create a database similar to the above. Our project was supported by “Dr. Juan Graham Casassus” hospital who gave us a batch of raw mammograms in DICOM file format [3]. especificar, el problema DEL hospital –diagnóstico– segunda opinión, decir que no se ha implementado algo así, pero que existe

.... info on what to look at for image treatment

La introducción de [15] da una explicación buena de por qué son buenos los sistemas CAD.

1.2. Problema

1.2.1. Planteamiento del problema

Se dispone de un banco de imágenes mamográficas sin procesar, el reto es mejorar su contraste, disminuir la cantidad de ruido.

It was necessary to discover which filters are the best choice in order to obtain optimized images ready for afterward phases such as segmentation, processing or analysis. Namely, it was necessary to reduce the noise and other artifacts in the image. Furthermore, we need to create an electronic file for each image. These data are obtained from each DICOM file and with the help of specialized doctors. The most important data is the medical mammogram diagnosis. Mammograms are classified according to the American College of Radiology (ACR) in BI-RADS. Mammography Atlas, edition 4th [4].

The assessment categories are: a. Mammographic assessment incomplete: 1. Category 0: Need additional imaging evaluation and/or prior mammograms for comparison. b. Mammographic assessment complete: 1. Category 1: Negative. 2. Category 2: Benign finding(s). 3. Category 3: Probably benign finding. 4. Category 4: Suspicious abnormality. 5. Category 5: Highly suggestive of malignancy. 6. Category 6: Proven malignancy.

1.2.2. Delimitation of the investigation

Scope

decir que hay un estado inicial y uno final la extensión del trabajo incluye — filtros y tipos de filtros aplicados, imágenes listas para etapas posteriores. hablar de base de datos más como un dataset que otra cosa — estructura: imagen cruda, procesada y ficha

a database of preprocessed mammograms, namely, mammograms without noise, reduction of artifacts and leveling of image quality.

Limitations

1.2.3. Research questions

1. What is the best combination of filters?

1.3. Objectives

1.3.1. General objective

Crear un banco de datos de dominio público con mamografías preprocesadas.

1.3.2. Specific objectives

Aplicar los algoritmos adecuados para preprocesar los mamogramas. Crear una ficha electrónica para cada caso. ¿Crear un método?

1.4. Justification

Our project is useful for the scientific community dedicated to processing and image analysis, mainly those who are dedicated to the study of medical images. As mention before our project could be a first step in creating a CAD system. In the long term, the creation of this database would benefit the Mexican radiologists. It will improve the state of the art medical image processing in the country.

Capítulo II. Theoretical framework

2.1. Frame of reference

The pre-processing step serves to improve the processing with other filters like “segmentation”,

best techniques for enhancement of X-ray image may not be best for enhancement for microscopic images. [6]

The enhancement methods can broadly be divided into the following two categories: Spatial Domain Methods (SDM) and Frequency Domain Methods (FDM). Spatial domain methods which are operate directly on pixels. Frequency domain which operates on the Fourier transform of an image. [6]

Many studies show that the histogram equalization is an effective method for improving the quality of medical images. [6, 8].

2.1.1. Bith depths

12-bit image (4096 levels of gray) displayed with 16 allocated bits appears dark. This happens because the maximal possible amplitude of 12-bit image for each pixel is 4095 and the maximal amplitude that can be displayed is 65535 (16 allocated bits). Therefore, 4096 levels need to be linearly scaled to 65536 levels to achieve good image displaying [8].

2.2. Conceptual framework

Digital image is a binary representation ... Pixel Histograms are frequency distributions, and histograms of images describe the frequency of the intensity values that occur in an image [9]. Image histogram offers a graphical representation of the tonal distribution of values in a digital image [10]. Color model Grayscale Image processing Medical image processing DICOM ***

Different kind of medical images Las consideraciones técnicas (físicas) de cómo se obtienen los mamogramas están fuera del alcance de este trabajo. In screening mammography, as practiced in USA, (also in Mexico), two x-ray images of each breast, in the medio lateral oblique and craniocaudal views, are acquired. MRI CT X-ray: Mammography is radiographic examination that is designed for detecting breast pathology, particularly breast cancer. [11] *** Mammo-graphic features characteristic of breast cancer are masses, particularly ones with irregular or “spiculated” margins; clusters of microcalcifications (tiny deposits of calcium); and architec-tural distortions of breast structures. *** Noise: Nuclear images are generally the most noisy. Noise is also significant in MRI, CT, and ultrasound imaging. In comparison to these, radio-graphy produces images with the least noise. The noise can cover and reduce the visibility of certain features within the image. The loss of visibility is especially significant for low-contrast objects. ***

Noise in digital mammographies (Practical digital mammography by Beverly Hashimoto)
De acuerdo a Hachimoto en las imágenes mamográficas tenemos cuatro tipos de ruido: *quan-tum *fixed electronic *señales secundarias *quanta secundario indirecto

2.2.1. Mamogramas

Un mamograma típicamente tiene de 12 a 16 bits de profundidad, con una resolución de 4000x5000 pixeles [13], [14].

2.3. Technological framework

* Recordar escribir que en las primeras pruebas sólo se mejoró la visualización del tejido graso y no del tejido mamario.

* Recordar escribir sobre (general purpose computers)

* Unsharp - masking to reduce noise.

Libros que ayudarán con los conceptos teóricos:

Formación física de las imágenes Algoritmos



(a) 12 bits



(b) 16 bits

Figura II.1: On the left, the original image in 12 bits visualized in Matlab, on the right, the same image in 16 bits. The images were visualized with an specific colormap in order to enhance his color.

Capítulo III. Resultados

La evaluación de resultados se realizó utilizando la opinión de los médicos y también utilizando los diagramas como herramienta evaluadora.

Cada caso preprocesado se almacenó en un servidor y está disponible en la siguiente dirección: www.casi.dais.mx.

El código fuente que se generó está disponible en línea bajo un controlador de versiones en el repositorio: github.com/omartrinidad/mammograms.

Capítulo IV. Conclusiones

4.1. Trabajos futuros

Es posible hacer uso de algoritmos más efectivos para la eliminación del ruido y la mejora del contraste.

Bibliografía

[suc,] The mini-mias database of mammograms.

[Heath et al., 2000] Heath, M., Bowyer, K., Kopans, D., Moore, R., and Kegelmeyer, P. (2000). The digital database for screening mammography. In *Proceedings of the 5th international workshop on digital mammography*, pages 212–218.