

## Clasificación de lesiones tipo distorsión arquitectural en imágenes de mamografías

CARDENAS-Cruz David†\*, MARTÍNEZ-Miriam, MONTERO-José Antonio, GAZGA-José Francisco.

*Instituto Tecnológico de Acapulco*

Recibido: Agosto, 22, 2017; Aceptado febrero 9, 2018

### Resumen

Con el avance de las tecnologías, los Sistemas CADs (Diagnóstico Asistidos por Computadoras) han cobrado gran auge, debido a que utilizan la potencia de las herramientas informáticas actuales para integrar diferentes procesos con el objetivo de reducir tiempo y costo. Este trabajo se enfoca a la detección automática de cáncer de mama a través de mamografías digitales, enfocándose a una lesión que por su naturaleza es difícil de detectar (distorsión arquitectural), se expone la metodología utilizada, análisis de técnicas para las diferentes etapas, así como los avances que se han logrado. Se muestran resultados del diseño de la interfaz, para la manipulación de las imágenes y de las tres primeras etapas. Durante la fase de preprocesamiento se aplica el algoritmo Seam Carving en cual puede modificar el tamaño de la imagen sin alterar su calidad. Al igual, en la misma etapa, se seleccionó un filtro de mediana para suavizado de bordes. Para la reducción de ruidos se emplea el filtro de Gabor el cual proporciona resultados satisfactorios, el aislamiento de la lesión.

**Palabras Clave:** Mamografía, Seam Carving, Gabor, Distorsión arquitectural.

### Abstract

With the advancement of technologies, Computer Assisted Diagnostics (CAD) systems have gained a great boom because they use the power of current computer tools to integrate different processes in order to reduce time and cost. This work focuses on the automatic detection of breast cancer through digital mammography, focusing on an injury that by its nature is difficult to detect (architectural distortion), it exposes the methodology used, analysis of techniques for the different stages, as well as the progress that has been made. It shows results of the interface design, for the manipulation of the images and the first three stages. During the preprocessing phase the Seam Carving algorithm is applied in which you can modify the image size without altering its quality. At the same stage, a median filter was selected for edge smoothing. To reduce noise, the Gabor filter is used which provides satisfactory results, the insulation of the lesion.

**Keywords:** Mammography, Seam Carving, Gabor, Architectural Distortion.

**Citación:** CARDENAS-Cruz David†\*, MARTÍNEZ-Miriam, MONTERO-José Antonio, GAZGA-José Francisco. Clasificación de Lesiones tipo distorsión arquitectural en imágenes de mamografías. Foro de Estudios sobre Guerrero. 2019, mayo 2018 - abril 2019 Vol. 6 No. 1 878 - 894

\*Correspondencia al Autor:

† Investigador contribuyendo como primer autor.

### Introducción

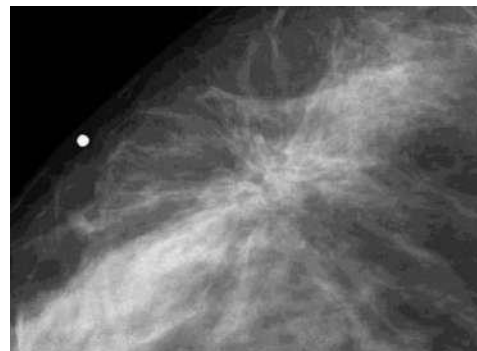
De acuerdo con datos proporcionados por el Instituto Nacional de Estadística y Geografía en su artículo titulado “Estadísticas a propósito del día mundial de la lucha contra el cáncer de mama”, el cáncer de mama es la principal causa de mortalidad entre todos los cánceres para las mujeres. Asimismo, indica que una de cada ocho mujeres está propensa a desarrollar esta enfermedad durante su vida. (INEGI, 2015).

Actualmente el uso de la mamografía de rayos-X es el método más utilizado en las instituciones de salud para detectar esta enfermedad en su etapa temprana. Si el tratamiento es oportuno, el número de muertes por esta causa puede ser reducido al menos en un 15% al 20%, cifra proporcionada por el Instituto Nacional del Cáncer en el 2017. Sin embargo, debido a varios factores, entre ellos la experiencia del radiólogo y la calidad en la imagen observada, hacen que la sensibilidad en el diagnóstico no sea perfecta. En este sentido y con el fin de apoyar el trabajo de los radiólogos se han diseñado sistemas de diagnóstico asistidos por computadora, CAD, los cuales tienen como objetivo, mejorar la calidad de la imagen y mostrar regiones de interés en un mamograma.

Existen básicamente tres tipos de lesiones que se pueden encontrar en la mama: microcalcificaciones, masas y distorsión arquitectural. Aunque la distorsión arquitectural es la lesión que se presenta en menor proporción, es una de las más difíciles de detectar debido a sus características.

La distorsión de la arquitectura glandular: es una alteración del tejido glandular con la

producción de radiaciones finas, espiculadas (forma de estrella) que no se encuentran asociadas con una masa. Como se puede observar en la figura 1.



*Figura 1. La distorsión se describe como una serie de líneas que se dirigen a un punto distinto del pezón, sin nódulo central*

La distorsión en la arquitectura de la mama se identifica por la presencia de tejidos que convergen hacia un mismo punto. Estas lesiones son las más difíciles de localizar, más aún si se dan en tejido denso y graso. La mama contiene varias estructuras lineales en trozos, tales como los conductos de los ligamentos y los vasos sanguíneos, los cuales causan una textura orientada en las mamografías. Con la presencia de la distorsión de la arquitectura se espera que cambie la textura normal de orientación de la mama.

Este trabajo se enfoca principalmente a la etapa de clasificación, sin embargo, se hace indispensable el planteamiento de la metodología general, así como la descripción de sus etapas.

## Objetivos

Diseñar e implementar de un módulo para identificar la lesión de distorsión arquitectural en imágenes de mamografías.

### Objetivos específicos

- Construir una base de imágenes de mastografías de lesiones de distorsión arquitectural obtenidas de una institución de salud.
- Analizar imágenes de mamografías y detectar las lesiones de distorsión arquitectural.
- Clasificar las imágenes de mastografías.
- Evaluar de manera cuantitativa los algoritmos utilizados considerando métodos estadísticos.

### Planteamiento del problema a resolver

Una mamografía de escrutinio busca visualizar lesiones no-palpables (es decir, menores de 0.5 cm si se trata de nódulos), calcificaciones (nunca palpables por su reducido tamaño), asimetrías en la densidad mamaria, y/o distorsión de la arquitectura de la glándula. (Navarro, 2006)

Por lo que puede haber errores de interpretación, dependiendo de la maestría del radiólogo que al interpretar las imágenes tiene que manejar incertidumbre, dando una variabilidad significativa en los diagnósticos, resultando hasta en un quinto de mujeres con cáncer de mama que se les diagnóstica erróneamente, es decir, cuando un radiólogo ve el cáncer, pero lo clasifica como benigno (falso-negativo). Además, el número de resultados falsos-positivos de las mamografías es absolutamente alto, tan solo un poco más que la mitad de las mujeres que experimentan una biopsia tienen realmente cáncer de mama (Sickles, 2002).

Actualmente, la proyección de la mamografía de Rayos-X, es el método más ampliamente utilizado por la población femenina para su diagnóstico. Sin embargo, en México el número de radiólogos especializados para el análisis de mamografías es bastante reducido, aunado a esto, la mayoría se concentra en las grandes ciudades. Por otro lado, un radiólogo tarda en promedio 1.5 horas en analizar un caso (4 imágenes de mamografías: dos vistas por cada mama) y ofrecer un diagnóstico, por lo que el número de casos analizados por día es bastante limitado. Otro aspecto que se debe tomar en consideración en el proceso de ofrecer un diagnóstico es la fatiga visual y el cansancio físico del radiólogo después de haber analizado varios casos. Esta situación puede causar un diagnóstico erróneo (falsos positivos o falsos negativos) lo cual también es grave.

Para apoyar en esta tarea surgen los sistemas de ayuda al diagnóstico (CAD: Computer Aided Diagnosis) que realizan la clasificación automática de las lesiones de mama. Los sistemas de CAD ayudan a radiólogos con la interpretación de las lesiones, resultando en que se refieren a menos mujeres para exámenes adicionales cuando tienen realmente lesiones benignas (Samulski, 2006). Desde hace algunos años, estos sistemas han sido desarrollados, por grupos de investigadores en diferentes partes del mundo, llegando la mayoría de estos a la etapa de prototipo, mientras, otros pocos, se encuentra en fase de comercialización y su uso se encuentra restringido principalmente a los países industrializados.

La distorsión arquitectural es la lesión que se presenta en menor proporción, es una de las más difícil de detectar debido a sus características (Radiology, 1998).

En esta lesión, la arquitectura normal de la mama es distorsionada y no se observa una masa definida visible. Asimismo, se pueden observar espículos radiando desde un punto, así como una distorsión en los bordes del tejido graso (parénquima).

La distorsión en la arquitectura de la mama se identifica por la presencia de tejidos que convergen hacia un mismo punto. Estas lesiones son las más difíciles de localizar, más aún si se dan en tejido denso y graso. La mama contiene varias estructuras lineales en trozos, tales como los conductos de los ligamentos y los vasos sanguíneos, los cuales causan una textura orientada en las mamografías. Con la presencia de la distorsión de la arquitectura se espera que cambie la textura normal de la orientación de la mama.

La distorsión se define en mamografía como la existencia de líneas que confluyen en un punto excéntrico al pezón, sin la existencia de un nódulo, en cuyo caso se trataría de una característica del borde.

La distorsión de la arquitectura es un reto para el radiólogo. Puede ser difícil de detectar y difícil de confirmar en mamografía. El estudio histológico de las mismas es, además, complejo para el patólogo.

### Justificación

Como se ha mencionado anteriormente el cáncer de mama la principal causa de mortalidad entre todos los cánceres para las mujeres esto se ve en el artículo: "Estadísticas a propósito del día mundial de la lucha contra el cáncer de mama. (INEGI, 2015). En la actualidad no existe ninguna cura para el cáncer en una fase muy avanzada, pero si se detecta a tiempo, todo cáncer es curable.

Según información de la Secretaría de Salud, el cáncer de mama consta de cinco etapas que van de la cero a la cuatro. Conforme la enfermedad avanza, las posibilidades de supervivencia se reducen. Sin embargo, los pacientes tienen 95%, 88% y 66% de probabilidades de sobrevivir si descubren la enfermedad en las primeras tres etapas. A partir de la cuarta, el porcentaje de supervivencia es de poco más de uno de cada tres, mientras que, en la etapa más avanzada del cáncer, sólo un 7% de los pacientes sobreviven.

En México, el 90% de los casos se detectan en las dos últimas etapas, lo que ha transformado a esta enfermedad en la primera causa de muerte entre las mujeres mexicanas desde el año 2006. Esto significa que cada dos horas muere una mexicana víctima del cáncer de mama.

En el aspecto económico, la inversión que el Gobierno debe hacer para tratar cada caso confirmado de cáncer de mama va de los 40 mil a los 400 mil pesos, según declaró el secretario de Salud, Salomón Chertorivsky.

Según la Dirección General de Información de Salud (DGIS), la cifra de víctimas mortales del cáncer de mama no ha parado de aumentar desde el año 2000. Ese año tres mil 419 mujeres perecieron víctimas del cáncer, mientras que en 2010 la cifra alcanzó las cinco mil defunciones en México.

Es por eso que la detección a tiempo de este mal debe ser un tema de interés de todos los mexicanos.

## Estado del Arte

De acuerdo a la investigación elaborada por el Dr. Jorge E Núñez Mc Leod, en su artículo "Algoritmos genéticos aplicados a la segmentación de imágenes y reconocimiento de formas" (Núñez Mc Leod, 2016) indica el uso de algoritmos genéticos en imágenes mamográficas.

La técnica de los algoritmos genéticos (AG) se basa en los conceptos de selección natural (supervivencia de los mejor adaptados). Es en si un proceso de optimización y por lo tanto cualquier problema que sea planteado como un proceso de minimización o maximización de un funcional puede ser encarado con esta técnica. La mayor o menor ventaja de la aplicación de los AG dependerá de diversos aspectos; pero posiblemente el de mayor importancia se refiere al grado con que los cromosomas representan el problema. No se debe perder de vista que cualquier problema que tenga una resolución matemática formal será resuelto en forma más óptima por esta que por un AG (Núñez Mc Leod, 2016). Con la aplicación de este algoritmo en imágenes de mastografías se puede llegar a una clasificación de lesiones tipo distorsión arquitectural más exacta debido al algoritmo genético. La gran desventaja que presenta este algoritmo es la gran cantidad de imágenes que se requiere en una base de imágenes para poder compararlas. Esto conlleva a un gran costo computacional debido al tratamiento de muchas imágenes además debemos obtener una gran cantidad de mamografías y digitalizarlas dichas mamografías.

Uno de los puntos a tratar en la presente investigación es la reducción de costo computacional en la actualidad existen diversas técnicas que reducen el tiempo de procesamiento de una imagen.

Una de esas técnicas son las redes Bayesianas Gaussianas como lo presenta los autores S. Kwon, H. Lee y S. Lee en su artículo "Mejora de la imagen con filtrado gaussiano en el sistema de imágenes por microondas en el dominio del tiempo para la detección del cáncer de mama" (Traducción del inglés: "Image enhancement with Gaussian filtering in time-domain microwave imaging system breast cancer detection"). Ellos analizaron la efectividad del método de desconvolución para la eliminación del efecto de fluctuación de tiempo de la media de la señal en una imagen reconstruida. Los resultados de la simulación de un modelo de mama simple y un modelo de mama realista demostraron que la imagen de mama reconstruida y RSM (Relaciones señal / media) obtenida utilizando el FBG (Filtrado de paso de banda gaussiano) de Gauss eran considerablemente mejores que las obtenidas usando deconvolución. Por lo tanto, los resultados demuestran que el método propuesto es muy eficaz tanto para un "espectro" de mama simple como para un espectro de pecho realista numérico. (S. Kwon, 2016).

Dennis Gabor proponía creación y utilización de herramientas de procesamiento de imágenes para mejorar la calidad de imágenes adquiridas con el microscopio electrónico. (Marr., 1982). Entre otras cosas, este trabajo sorprende por su modernidad, pues en él se defiende la relevancia informativa de los contornos o siluetas de los objetos de la imagen y se proponen técnicas para su realce (basadas en ecuaciones en derivadas parciales) que han sido realmente estudiadas y comprendidas en los últimos 15 años.

Esta línea de pensamiento tuvo su formulación explícita como programa de investigación en el libro *Visión* de David Marr publicado en 1982 que ha guiado una parte fundamental de la investigación en procesamiento de imágenes guiado una parte fundamental de la investigación en procesamiento de imágenes dedicada a la detección de contornos y ha conducido a otros desarrollos en segmentación de imágenes. (Marr., 1982)

Existen muchas técnicas distintas de segmentación de imágenes dependiendo del tipo de imágenes y de los objetivos deseados. La mayoría de ellas tienen su punto de arranque en la teoría de la detección de contornos de Marr-Hildreth. (Hildreth., 1980)

Otro punto de vista diferente fue introducido por M. Kass, A. Witkin y D. Terzopoulos en que abordaron el problema de la integración de los puntos donde el gradiente de la imagen es alto proponiendo el modelo llamado desde entonces de los contornos activos. (M. Kass, 1988)

Los modelos geométricos de contornos activos están basados en la evolución de curvas o superficies por flujos geométricos, teoría que ha influido poderosamente en la comunidad de procesamiento de imágenes. (L. Álvarez, 1993) En dichos modelos, la curva o superficie se deforma según una velocidad que depende de parámetros geométricos intrínsecos y de la adaptación a los contornos de la imagen. Se trata de un flujo geométrico basado en la formulación por conjuntos de nivel del movimiento de curvas (resp. superficies) según su curvatura (resp. curvatura media) que requiere para su solución la teoría de soluciones de viscosidad (M.G. Crandall, 1992).

La implementación numérica de estos flujos geométricos se apoya en los trabajos de S. Osher y J. Sethian y permite cambios automáticos de la topología de la superficie que se deforma, así como la detección simultánea de varios contornos sin ningún procedimiento de seguimiento especial. (Sethian., 1998)

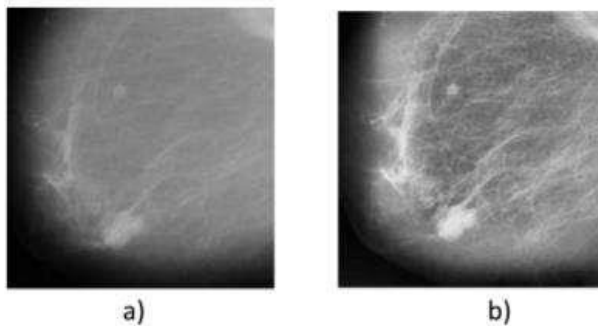
En el artículo *Going Beyond a First Reader: A Machine learning Methodology for Optimizing Cost and Performance in Breast Ultrasound Diagnosis* publicado por los doctores Santosh S. Venkatesh, Benjamin J. Levenback, Laith R. Sultan, Ghizlane Bouzghar, and Chandra M Sehgal en el año 2015 revela una técnica en la cual combina 2 algoritmos: Impulso adaptativo (adaptive boosting) y Clasificador de poda (pruning classifier). (Venkatesh, 2015.). Las combinaciones de estas dos técnicas tuvieron un alto desempeño en diagnóstico para las masas benignas y malignas en imágenes de ultrasonidos mamarios (la evaluación bajo el área bajo la curva de ROC fue de 0.98) con una porción de poda del 20 \%. El impulso adaptativo también mejoró el rendimiento diagnóstico de los observadores humanos individuales y aumentó el acuerdo entre sus análisis.

La combinación del impulso adaptativo con la poda selectiva de algunos métodos independientes es una de las metodologías según este artículo para alcanzar un desempeño alto en cuestión de diagnóstico sin tener que agregar un costo computacional adicional para la detección de masas malignas y benignas en imágenes de ultrasonidos mamarios. (Venkatesh, 2015.)

El método propuesto en el artículo “Combinación de técnicas de procesamiento digital de imágenes e inteligencia artificial para el análisis de imágenes de mamografía” elaborado por Alejandro Díaz Sotolongo y Danays Barbara Costa Alonso en el año 2013, se fundamenta en la correcta combinación de un vector compuesto por características de forma y textura.

También se presenta el proceso de análisis estadístico para la reducción de conjuntos de características y obtención de vectores óptimos en el proceso de clasificación. (Sotolongo, 2013).

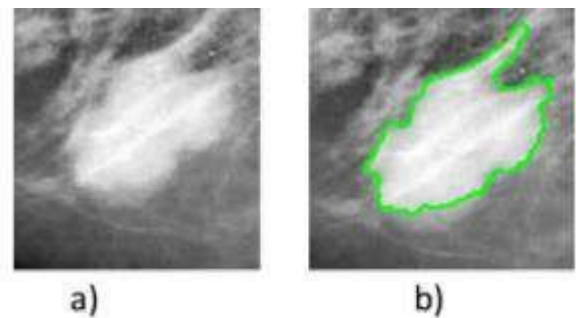
En cuestión de eliminación de ruido se aplicó la Ecuación del Histograma de Contraste Adaptativo, la cual realiza una mejora del contraste de forma uniforme mediante la transformación de los valores de intensidad de los píxeles de las imágenes, consiguiendo de esta forma, mejorar el contorno de los objetos de interés. Ver figura 2.



**Figura 2.** Ecuación del histograma de contraste adaptativo a) Imagen original. b) Imagen procesada.

En el proceso de segmentación del objeto de interés (tumor) se utiliza una técnica conocida como “cortes inteligentes” (live wire or intelligent scissors).

Según Liang, el *live wire* basa su funcionamiento en dos componentes esenciales: una “función de costo” local que asigna el menor costo a las características de mayor interés de los objetos (por ejemplo, los bordes) y un “proceso de expansión” que genera bordes óptimos para los objetos de interés, basados en la mencionada función de costo y un pequeño conjunto de puntos semillas que son suministrados por el usuario.



**Figura 3.** Segmentación con el *live wire* a) Imagen original b) Imagen segmentada.

El proceso de extracción de características se genera a partir del objeto segmentado del cual se extraen un conjunto de treinta y dos características de forma y textura.

Las características de forma brindan información basada en la morfología del objeto, mientras que las de textura describen la topología y se basan en las matrices de coocurrencia de niveles de gris. Se puede observar un ejemplo en la imagen 3.

## Metodología

La metodología que se empleará para la elaboración del clasificador de lesiones tipo distorsión arquitectural en imágenes de mastografías será las siguientes fases:

- Recolección de imágenes de mastografías provenientes de un centro de salud.
- Preprocesamiento de las imágenes de mastografías para la base de imágenes del modulo
- Segmentación de las imágenes
- Extracción y selección de características de las lesiones de distorsión arquitectural
- Clasificación de las regiones de interés (Maligno o benigno)

La metodología mencionada se puede visualizar en la figura 4.

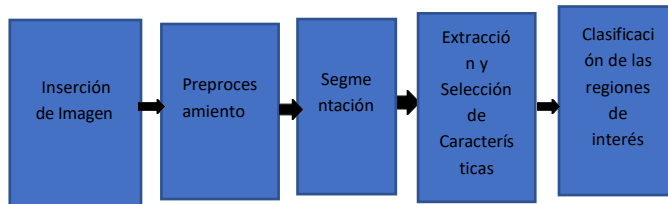


Figura 4. Metodología General

## Inserción de Imágenes de Mamografía

Las imágenes de mastografías recolectadas serán proporcionadas por un centro de salud con los debidos permisos de la institución de salud. Estas imágenes de mastografías se almacenarán como archivos, esto quiere decir que no se necesitara en específico algún sistema gestor de base de datos. Se optó por esta medida para mejorar el desempeño ya que solo se manejaría una sola tabla la cual tendría por campos el nombre de la imagen y la imagen como tal, dando así un desperdicio de recursos.

## Preprocesamiento

Para la preparación de las imágenes para su posterior uso para el clasificador y almacenamiento a la base de imágenes se tiene que aplicar filtros para la eliminación de ruido.

En esta etapa se trabaja con la imagen ingresada. Esta imagen se “prepara” para su procesamiento se puede visualizar esta etapa en la figura 5.

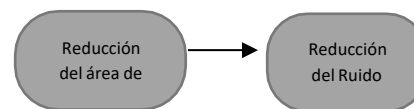


Figura 5. Metodología de preprocesamiento

## Reducción del área de Trabajo

Las mamografías presentan áreas que no corresponden a la glándula mamaria lo que ocasiona costos computacionales innecesarios y afecta el desempeño de los sistemas desarrollados. Por esa razón es necesario identificar y eliminar esas áreas que no brindan información relevante para el proceso de detección.

De igual forma para la reducción del área de trabajo se utilizará el algoritmo *Seam Carving* (tallado de costura). Este algoritmo reducirá el tamaño de la imagen, pero sin alterar su calidad de la misma.

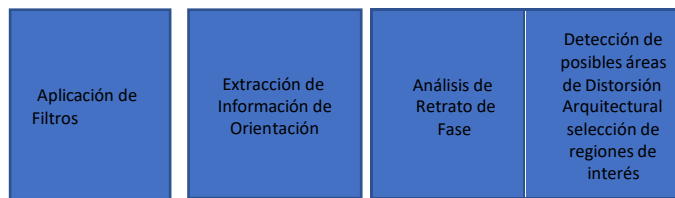
El algoritmo utiliza el termino de costuras. Una costura es una ruta óptima de 8 píxeles conectados en una única imagen de arriba a abajo, o de izquierda a derecha, donde la óptima es definida por una función de energía de la imagen.

Al tallar o insertar repetidamente las costuras en una dirección, podemos cambiar la relación de aspecto de una imagen. Mediante la aplicación de estos operadores en ambas direcciones podemos redirigir la imagen a un nuevo tamaño.



La selección y el orden de las costuras protegen el contenido de la imagen, tal como se define por la función de energía. La talla de costura también se puede utilizar para mejorar el contenido de la imagen y la eliminación de objetos. (Shamir, 2007)

El algoritmo es muy eficaz en comparación cortar o escalar la imagen. La comparación lo podemos ver en la imagen 10.



**Figura 6.** Metodología de segmentación

### Reducción del Ruido

Una vez se realiza la reducción del área de trabajo, y se eliminan las zonas sin información, se procede a realizar la reducción del ruido sobre la imagen resultante. Para este proceso se utiliza el filtro Mediana en la cual se detallará en la sección de resultados.

### Segmentación

Una vez que ya tenemos la imagen preparada y “limpia” podemos trabajar con las siguientes etapas que se visualizan en la figura 6.

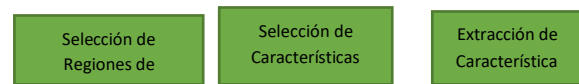
Para la reducción de ruidos se aplicará el filtro de Gabor el cual consiste en un filtro lineal cuya respuesta de impulso es una función sinusoidal multiplicada por una función gaussiana.

Las funciones de Gabor son funciones que operan en el conjunto de los números complejos, y que la parte real es la función de Gabor simétrica (simetría par).

La parte imaginaria es la función de Gabor asimétrica (simetría impar) (Daugman, 1988)

### Extracción y selección de Características

Para la extracción de características se ocupará la metodología vista en la figura 7.

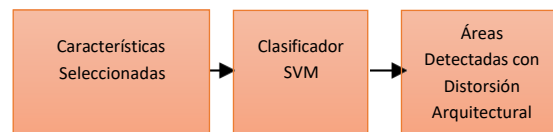


**Figura 7.** Metodología para extracción y selección de características

### Clasificación de las regiones de interés

Una vez que la imagen se encuentre segmentada y la lesión arquitectural se detecta con la eliminación de ruidos, se procederá a clasificar dicha imagen, con el propósito de detectar cáncer. Para esto ocuparemos el algoritmo de *Maquinas de Soporte Vectorial (MVS)*.

La metodología para clasificación de las regiones de interés se puede observar en la figura 8.



**Figura 8.** Metodología para clasificación

La teoría de las Máquinas de Vectores de Soporte (SVM por su nombre en inglés Support Vector Machine) fue desarrollada por Vapnik basado en la idea de minimización del riesgo estructural (SRM).

La MVS mapean los puntos de entrada a un espacio de características de una dimensión mayor, para luego encontrar el hiperplano que los separe y maximice el margen entre las clases.

La formulación matemática de las Máquinas de Vectores Soporte varía dependiendo de la naturaleza de los datos; es decir, existe una formulación para los casos lineales y, por otro lado, una formulación para casos no lineales.

Las MVS han sido desarrolladas como una técnica robusta para clasificación y regresión aplicado a grandes conjuntos de datos complejos con ruido; es decir, con variables inherentes al modelo que para otras técnicas aumentan la posibilidad de error en los resultados pues resultan difíciles de cuantificar y observar.

Para el módulo de clasificación de lesiones tipo distorsión arquitectural en imágenes de mastografías utilizaremos un separador no lineal de MVS.

Hay casos donde los datos no pueden ser separados linealmente a través de un hiperplano óptimo en el espacio de entrada. En muchas situaciones, los datos, a través de una transformación no lineal del espacio de entradas, pueden ser separados linealmente, pero en un espacio de características y se pueden aplicar los mismos razonamientos que para las MVS lineal con margen máximo.

La transformación de los datos de un espacio inicial a otro de mayor dimensión se logra mediante el uso de la función *kernel*.

Una función núcleo o kernel es un producto interno en el espacio de características, que tiene su equivalente en el espacio de entrada. De manera gráfica se puede observar en las imágenes 9.1 y 9.2 como la función kernel permite realizar la separación y el traslado de los datos al espacio de características.

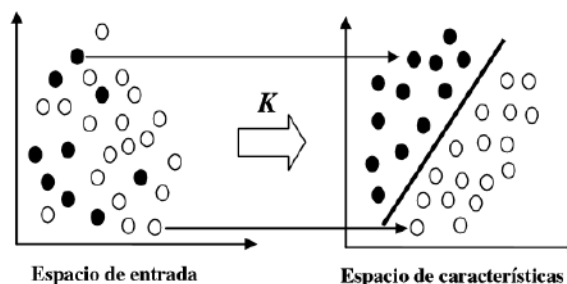


Figura 9.1 MVS no linealmente separable inducida por una función kernel.

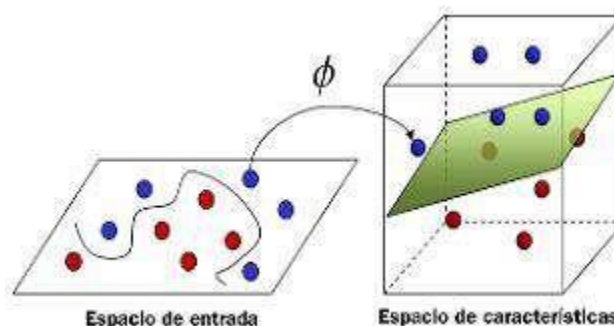


Figura 9.2. Incrementación de los kernel.

La máquina de soporte vectoriales empezara arrojar resultados del clasificador de lesiones tipo distorsión arquitectural es importante evaluar el desempeño del módulo; para esto se ocupará un análisis ROC. Este análisis ROC obtendrá la eficacia del clasificador de lesiones tipo distorsión arquitectural en imágenes de mamografías (falsos positivos).

## Resultados

De acuerdo a la metodología mencionada anteriormente, el proyecto se encuentra terminando la segunda fase, el cual es preprocesamiento.

A continuación, se detallan los resultados parciales obtenidos en estas fases. Así como el diseño de la herramienta.

### Fase 1: Adquisición de la imagen.

En esta fase, actualmente estamos haciendo convenios con el Hospital Estatal de Cancerología de Guerrero. No dudo que en un futuro no muy lejano podamos llegar a un acuerdo para que el hospital nos pueda proporcionar las imágenes de mamografías, ya sean estas de manera física o digital.

Para los resultados siguientes se uso una base de datos *The Digital Database for Screening Mammography (DDSM)* de la universidad de Florida del Sur. El DDSM ha sido ampliamente utilizado por la comunidad de investigadores. Se mantiene en la Universidad del Sur de la Florida con el propósito de mantenerlo accesible en la web.

### Fase 2: Preprocesamiento Reducción del área de trabajo

En esta etapa reduciremos la imagen para no procesar datos innecesarios como podría ser, nombre de la paciente, fecha o datos generales de la mamografía.

Para la eliminación de los datos generales podríamos ocupar el método convencional de recortar la imagen. Esto sería un método eficaz la única problemática que habría después del corte sería la estandarización de las imágenes, ya que, en procedimientos posteriores es necesario que las imágenes tengan el mismo tamaño. Escalar la imagen no será una buena opción debido a la gran pérdida de la calidad de la misma.

Para estandarizar un tamaño en las imágenes de mamografías ocupe un algoritmo llamado Seam Carving, el cual puedo ampliar o reducir las imágenes de mamografías sin alterar su calidad de la misma.

En la imagen 10 podremos ver los resultados de este procedimiento, en esta

situación probamos con la reducción del tamaño.

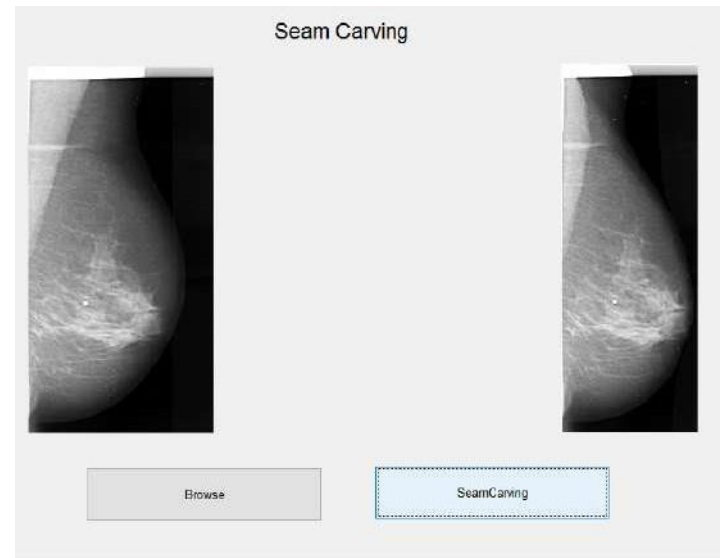


Figura 10. Aplicación del algoritmo Seam Carving

Como se puede observar en la imagen anterior, no solamente redujo la “parte oscura” de la imagen si no también parte de la representación de la mama como se puede observar en la línea blanco y negro que atraviesa la imagen en la figura 11. Esta línea representa los píxeles, *menos visibles*, eliminados. Esto ayudara en gran manera posteriormente puesto que puede reducir el costo computacional.

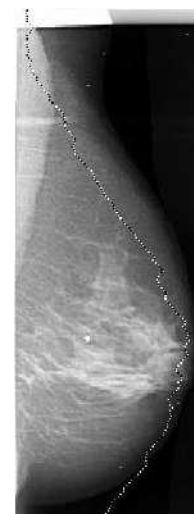


Figura 1. Eliminación de Píxeles

## Reducción de Ruido

En esta etapa se seleccionó un filtro para suavizar los bordes de la imagen. Podemos ver en la imagen 12 como están los bordes de la imagen original, el cual sinceramente solo nos interesan los bordes de las lesiones.



Figura 12. Imagen Original y sus bordes

En la imagen anterior se puede visualizar todos los bordes con la cual la imagen original cuenta, la cual es mucho y no muy significativos.

La aplicación de filtros para suavizar los bordes es muy importante para el proceso futuro de segmentación.

Los filtros que se analizaron fueron: Filtro Gaussiano, Filtro Box y Filtro de Mediana (agregando ruido de sal y pimienta). Los cuales detallaremos los resultados obtenidos.

### Filtro Gaussiano:

Simulan una distribución gaussiana multivariable. El valor máximo aparece en el pixel central y disminuye hacia los extremos tanto más rápido cuanto menor sea el parámetro de desviación típica. El resultado será un conjunto de valores entre 0 y 1. La aplicación de este filtro se puede visualizar en la imagen 12.1.

Para este caso se utilizó la matriz:

$$\begin{bmatrix} 0 & 1 & 2 & 1 & 0 \\ 1 & 3 & 5 & 3 & 1 \\ 2 & 5 & 9 & 5 & 2 \\ 1 & 3 & 5 & 3 & 1 \\ 0 & 1 & 2 & 1 & 0 \end{bmatrix}$$


Figura 12.1. Filtro Gaussiano

Como podemos ver en la imagen anterior, el filtro gaussiano si suavizo un poco los bordes, pero, no lo suficiente aún tenemos bastante bordes innecesarios.

### Filtro Box:

El filtrado *box* implica reemplazar cada píxel de una imagen por el promedio en una caja. Cuando se extiende de varias maneras simples, se convierte en una herramienta de propósito general eficiente para el procesamiento de imágenes. (McDonnell, 1981)



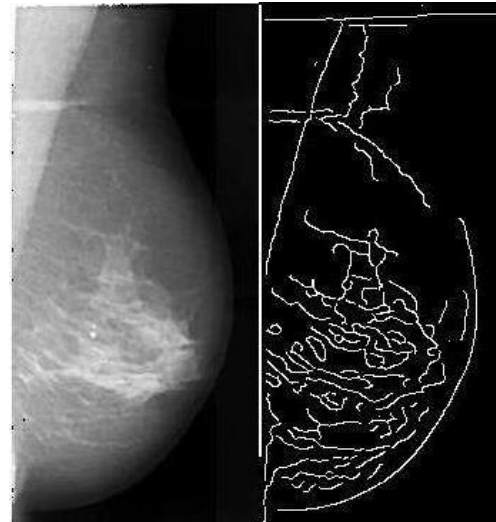
*Figura 12.2. Filtro Box*

En la figura 12.2 se puede observar como el filtro box suavizo un poco más los bordes en comparación del filtro gaussiano, pero el área la cual nos interesa solo es la que presenta alguna lesión. Por ende, este filtro aun proporciona varios bordes innecesarios que pueden afectar la rapidez de procesamiento en etapas posteriores.

#### **Filtro Mediana:**

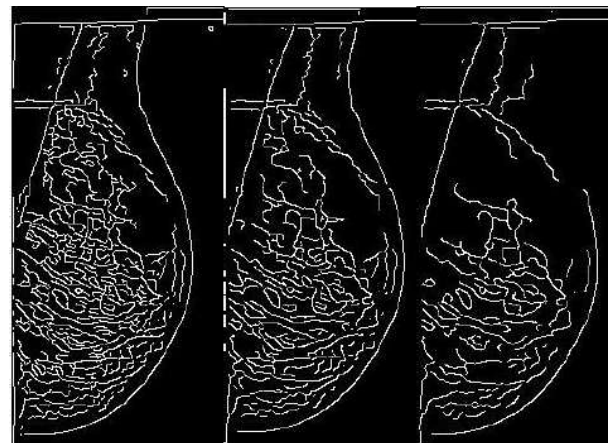
Los filtros mediana tiene la ventaja de que el valor final del pixel es un valor real presente en la imagen y no un promedio, de este modo se reduce el efecto borroso que tienen las imágenes que han sufrido un filtro.

A este proceso se le añade un poco de ruido, sal y pimienta, el cual ayuda a suavizar los bordes aún más.



*Figura 12.3. filtro mediana*

En la figura 12.3 se muestra como el filtro mediana suaviza los bordes de la imagen de mamografía. En comparación con los demás filtros podemos observar que este tipo de filtrado suaviza aún más los bordes de la lesión, omitiendo los demás bordes innecesarios. Además, este filtro no embórrese demasiado la mamografía en comparación con los otros filtros como se puede ver en la imagen 12.4. Es por eso que para este proyecto utilizaremos este tipo de filtrado.



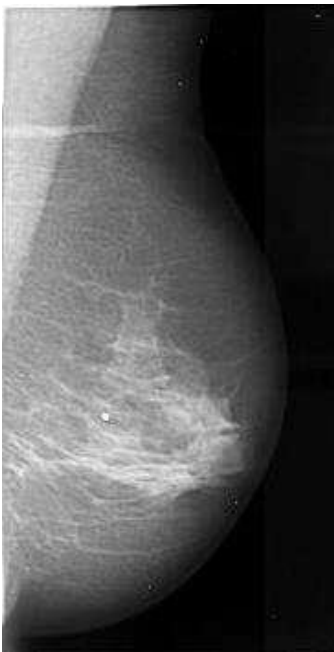
*Figura 12.4. Comparación entre filtros Gaussiano, box y mediana*

### Fase: Segmentación

En esta fase “depuraremos” información innecesaria y mostraremos solamente las lesiones de la mastografía. Esto ayudará en procesos posteriores como lo son la extracción de características.

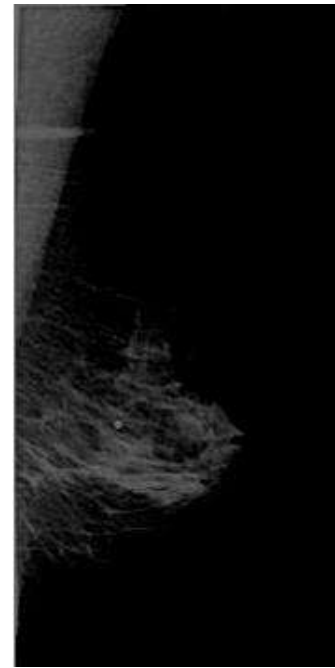
Como mencionado en la metodología ocupare el filtro de Gabor.

Se mostrará la aplicación del filtro de Gabor. El proceso comienza con la obtención de la imagen de la mamografía en su representación en escala de grises mostrada en la figura 13.1.



*Figura 13.1. Imagen con escala de grises*

Se construye el filtro de Gabor con un tamaño de ventana de 370X180 pixeles,  $\theta$  de 100°,  $\sigma$  de 50,  $\lambda$  de 20. Que son los valores propuestos para el proyecto en esta etapa. Se hace una convolución entre el filtro de Gabor y la imagen, la cual nos da como resultado la imagen mostrada en la figura 13.2.



*Figura 13.2. Imagen con filtros de Gabor*

En la figura 13.2 podemos visualizar como se *depuró* los elementos ajenos a la lesión de la mama. Al ver el resultado que arrojó el filtro de Gabor se consultó con un radiólogo para determinar si es suficiente información para detectar la lesión, dando como respuesta la confirmación.

### Conclusión

Bajo los resultados obtenidos en la fase 1 se debe considerar que en la adquisición, la imagen puede variar, debido a tamaño y calidad de la mastografía, por lo cual se trabaja con una BD pública para validar resultados. Actualmente se ha establecido el contacto con un centro de salud, pero debido a cuestiones de seguridad e integridad para los pacientes, se encuentra en vías de un convenio entre instituciones, por lo cual se consideran algunos cambios en el tratamiento de las mamografías.

Teniendo esto en cuenta en la fase 2 se optó por la utilización del algoritmo Seam Carving (Shamir, 2007) en cual se aplica sin importar cuál sea el tamaño de la imagen dado que este algoritmo puede modificar el tamaño de la imagen sin alterar su calidad. Esto además de estandarizar un tamaño de imágenes podrá reducir costo computacional.

En esta misma fase, se analizaron diferentes filtros para suavizar los bordes de la imagen. A pesar de que el filtro de la mediana aparenta ser mejor que los demás, no podemos tener la certeza que se utilizará dicho filtro debido a que no poseemos las imágenes reales, aunque es muy probable que no difieran mucho a la imagen obtenida por la B.D. pública.

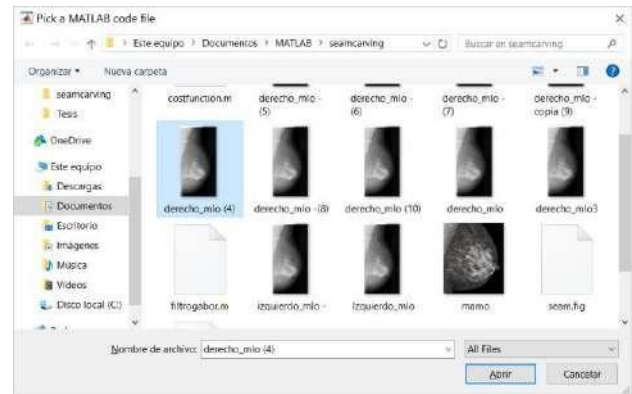
En las primeras etapas de la fase 3 la cual se aplicó el filtro de Gabor pudiendo visualizar de una manera aceptable la lesión de mama de la imagen. Cabe de resaltar que los valores de las variables de la ecuación pueden cambiar, dependiendo la imagen del centro de salud.

En conclusión, la aplicación del algoritmo Seam Carving y el filtro de Gabor arrojaron resultados bastantes satisfactorios.

Con respecto a la Interfaz, se encuentra concluida y se pueden integrar las imágenes para su manipulación. Ver figuras 14.1-14.5.



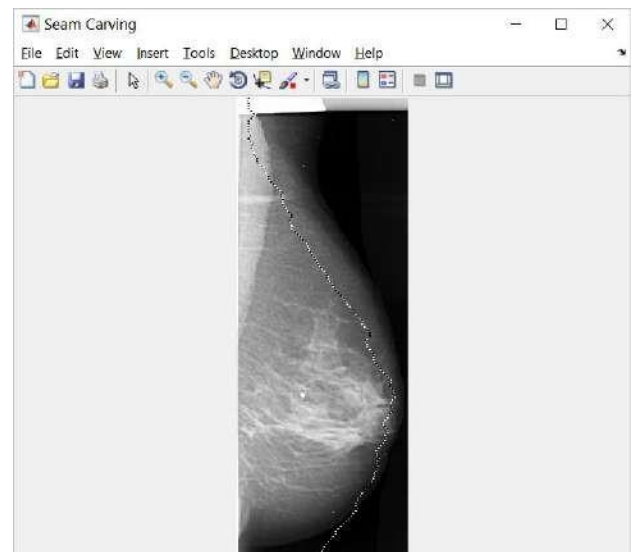
**Figura 14.1.** Ventana Principal de la interfaz



**Figura 14.2.** Selección de Imagen



**Figura 14.3.** Imagen original, sin filtros



**Figura 14.4.** Aplicación del Algoritmo Seam Carving

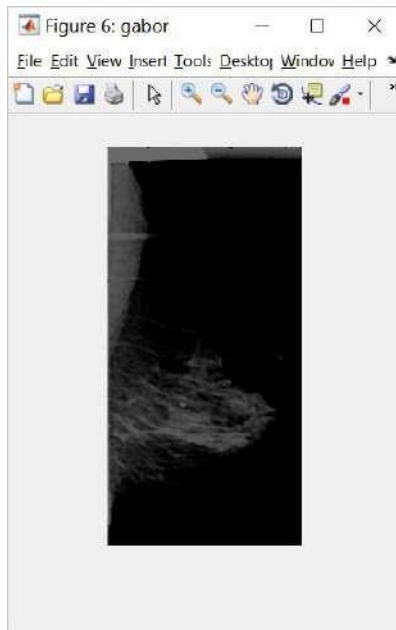


Figura 14.5. Aplicación de Filtro de Gabor

## Referencias

- Daugman, J. G. (1988). Complete discrete 2-D Gabor Transforms by networks for image analysis and compression. *IEE Transactions on Acoustics, Speech and Signal Processing*, 1169-1179.
- Hildreth., D. M. (1980). Theory of edge detection. . *Proc. Royal Soc. London, B*, 207 , 187-217.
- INEGI. (2015). ESTADÍSTICAS A PROPÓSITO DEL... DÍA MUNDIAL DE LA LUCHA CONTRA EL CÁNCER DE MAMA. *Instituto Nacional de Estadística y Geografía*, 1-14.
- L. Álvarez, F. G. (1993). Axioms and fundamental equations of image processing. *Arch. Rat. Mech. and Anal.* 16, IX, 200-257.
- M. Kass, A. W. (1988). Active contour models. *International Journal of Computer Vision* 1, 321-331.
- M.G. Crandall, H. I. (1992). User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc.*, 1-67.
- Marr., D. (1982). Vision. W.H. *Freeman and Company*.
- McDonnell, M. J. (1981). Box-filtering techniques. *Computer Graphics and Image Processing*, 65-70.
- Navarro, M. E. (2006). Detección del Cáncer de Mama: Estado de la Mamografía en México . *Cancerología 1*, 147-162.
- Núñez Mc Leod, J. E. (2016). Algoritmos genéticos aplicados a la segmentación de Imágenes y reconocimiento de formas. *Instituto CEDIAC, acultad de Ingeniería, Universidad Nacional de Cuyo*, 1.
- Radiology, A. C. (1998). *Breast imaging reporting and data system (BI-RADS)*, 3rd ed. Reston, VA : American College of Radiology.
- S. Kwon, H. L. (2016). Image enhancement with Gaussian filtering in time-domain microwave imaging system for breast cancer detection. *Department of Electrical Engineering, Ewha Womans University*, 342-344.
- Samulski. (2006). *Classification of Breast Lesions in Digital Mammograms*. Netherlands: University Medical Center Nijmegen: Master Tesis.
- Sethian., S. O. (1998). Fronts propagating with curvature dependent speed: algorithms based on the Hamilton-Jacobi formulation. *J. Comp. Physics* , 12-49.



- Shamir, S. A. (2007). Seam carving for content-aware image resizing. *ACM Trans. Graph*, 10.
- Sickles, E. A. (2002). Performance parameters for screening and diagnostic mamography: specialist and general radiologists. *Radiology* 1, 861–869.
- Sotolongo, A. D. (2013). COMBINACION DE TECNICAS DE PROCESAMIENTO DIGITAL DE IMAGENES E INTELIGENCIA ARTIFICIAL PARA EL ANALISIS DE IMAGENES DE MAMOGRAFIA. *3C Tecnología*, 3.
- Venkatesh, S. S. (2015,). Going beyond a First Reader: A Machine Learning Methodology for Optimizing Cost and Performance in Breast Ultrasound Diagnosis. *Ultrasound in medicine & biology*.