

IMPLEMENTACIÓN DE UNA BASE DE DATOS DE MAMOGRAFÍA CON LESIÓN DE DISTORSIÓN ARQUITECTURAL

Cruz David Cárdenas Clavel¹ Martínez Arroyo Miriam²,
Montero Valverde José Antonio³, Gazga Portillo José Francisco⁴

Resumen La tecnología ha avanzado de manera exponencial los últimos años. El principal motivo por el cual el ser humano ha creado diferentes tecnologías es, facilitar y reducir costos ante una problemática. El cáncer de mama en México según datos de la INEGI es la principal causa de muerte entre las mujeres mexicanas. En este trabajo se muestran los resultados de una primera etapa para un sistema de detección automática de cáncer de mama. Se implementa una base de datos de imágenes de mamografías con lesiones de distorsión arquitectural, en la fase de preprocesamiento, se aplica el algoritmo Seam Carving con el cual se puede modificar el tamaño de la imagen sin alterar su calidad, para el suavizado de bordes se utiliza el filtro de mediana y para reducción de ruido en la fase de segmentación se emplea el filtro de Gabor el cual proporciona el aislamiento de la lesión.

Palabras clave— Mamografía, Seam Carving, Gabor, Distorsión arquitectural

Introducción

La tecnología ha avanzado de manera exponencial los últimos años. Esto se puede observar a simple vista con los diferentes dispositivos que en la actualidad se están utilizando. El principal motivo por el cual el ser humano ha creado diferentes tecnologías es, facilitar y reducir costos ante una problemática. El cáncer de mama en México según la INEGI es la enfermedad que más mortandad tiene en las mujeres mexicanas. Cuando se crean sistemas de apoyo al diagnóstico de cáncer de mama es fundamental contar con diversos bancos de registros mamográficos. Para mejorar la robustez de dichos sistemas a nivel nacional, es importante tener bases de datos que contengan registros mamográficos de pacientes de la región, pues las características físicas y epidemiológicas pueden cambiar de una zona a otra. Existen bases de datos internacionales como la MiniMIAS (The Mammographic Image Analysis Society) (Suckling, 2015) y la DDSM (*The Digital Database for Screening Mammography*) (Wang, 2015), que sirven para la prueba y validación de metodologías de procesamiento de mamografías a nivel mundial. Sin embargo, actualmente no hay muchos proyectos enfocados en la generación de nuevas bases de datos (Moya, 2014), que contengan registros de mamografías adquiridos con sistemas tecnológicos actualizados, además, a nivel local y nacional no existen o son escasas las bases de datos de registros de pacientes. En este trabajo se implementó una base de datos de imágenes de mamografías las cuales tenían la lesión tipo distorsión arquitectural inicialmente se extrajeron de la DDSM, para posteriormente trabajar con un Centro de salud de la localidad. La lesión es una lesión que por su naturaleza es difícil de detectar para muchos expertos del área. 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. 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.

Descripción del Método

La metodología para la creación de la base de datos inicia con la solicitud del consentimiento informado de las pacientes, luego, se describe la adquisición y almacenamiento de los registros mamográficos obtenidos en cada estudio, en la siguiente etapa se realizó un análisis y validación de la información contenida en el reporte médico con el acompañamiento de un especialista y la última etapa corresponde al etiquetado de las imágenes acorde con la información de la etapa anterior y la confirmación del resultado de las biopsias en la metodología (Figura 1) que se empleará para la elaboración del clasificador de lesiones tipo distorsión arquitectural en imágenes de mastografías será las siguientes fases:

- Inserción de la imagen.
- Preprocesamiento de las imágenes de mastografías para la base de imágenes del modulo
- Segmentación de las imágenes

¹ El Ing. Cruz David Cárdenas Clavel es alumno de la MSC en el Instituto Tecnológico de Acapulco (ITA) del TecNM, cdcc89@gmail.com (autor correspondiente)

² La Dra. Miriam Martínez Arroyo es Profesora de la MSC en el ITA del TecNM miriamma_ds@hotmail.com

³ El Dr. José Antonio Montero Valverde es Profesor de la MSC en el ITA del TecNM jamontero1@infintutmmail.com

⁴ El M.C. Gazga Portillo José Francisco es Profesor de la MSC en el ITA del TecNM jfgazga@hotmail.com

- Extracción y selección de características de las lesiones de distorsión arquitectural
- Clasificación de las regiones de interés (Maligno o benigno)

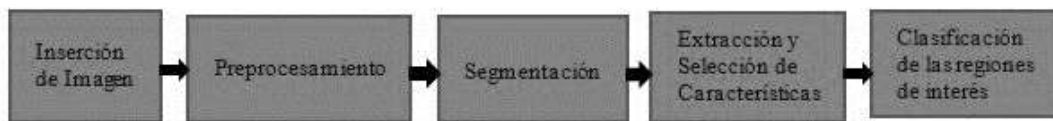


Figura 1: Metodología General

Inserción de Imágenes de Mamografía

Las imágenes de mastografías recolectadas inicialmente de la BD de DDSM. Estas imágenes de mastografías se almacenan como archivos, esto quiere decir que no es necesario usar algún sistema gestor de base de datos. Considerando, primero la importancia de construir nuevas bases de datos y segundo la relevancia de los sistemas de detección para el diagnóstico de cáncer en una etapa temprana; el problema a resolver consiste en la creación de una base de datos para validar el diagnóstico médico mediante técnicas de pre-procesamiento y la extracción de características que permitan la detección.

Preprocesamiento

Para los métodos de pre-procesamiento se han utilizado diversas técnicas que pueden afectar las propiedades de la imagen, lo que se refleja como una pérdida de detalles especialmente en la frontera entre los objetos de interés y el fondo. Existen también metodologías para el realce de contraste y reducción de ruido que mejoran elementos en las imágenes y son de bajo costo computacional, que sin embargo éstas pueden presentar pérdida de información fundamental en el análisis de imágenes de mamografía, ya que pequeños hallazgos que en imágenes de baja resolución puede estar representadas mediante el tamaño de un píxel. Debido a esto, es importante utilizar técnicas que permitan analizar la distribución espacial de un píxel y los píxeles de su entorno, con el fin de incrementar la precisión en la y la severidad de la lesión (benigna o maligna), con lo cual disminuirían los índices de mortalidad por cáncer de mama y los episodios de estrés innecesarios en las pacientes. Esta preparación de las imágenes tras aplicar filtros para la eliminación de ruido es usada para su procesamiento.

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. Para la reducción del área de trabajo se utilizamos el algoritmo *Seam Carving* (*tallado de costura*). Este algoritmo reduce el tamaño de la imagen, sin alterar la calidad de la misma. El algoritmo utiliza el término 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).

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. Para la reducción de ruidos aplicamos 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 utiliza la metodología mostrada la Figura 2.

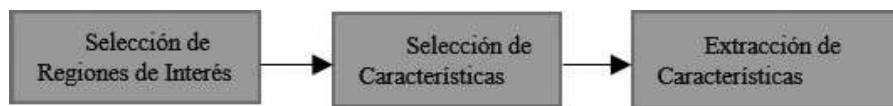


Figura 2: Metodología para extracción y selección de características

Clasificación de las regiones de interés

Una vez que la imagen se encuentra segmentada y área de interés se delimita, se clasifica el tipo de sesión. Para esto usaremos el algoritmo de Maquinas de Soporte Vectorial (MVS). La metodología para clasificación de las regiones de interés se puede observar en la Figura 3.

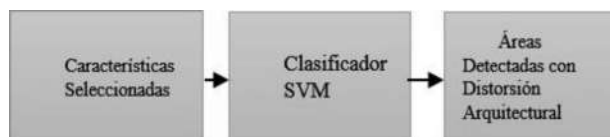


Figura 3: Metodología para clasificación

La teoría de las Máquinas de Soporte Vectorial (SVM por su nombre en inglés Support Vector Machine) fue desarrollada por Vapnik basado en el concepto 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 Soporte Vectorial 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*.

Resultados

De acuerdo a la metodología mencionada, 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. La figura 4 muestra el diseño de la interfaz gráfica implementado en MATLAB en donde se desarrolla el proyecto, en la figura 4(a) se muestra la pantalla principal, la figura 4(b) muestra la selección de las imágenes y la fig.4(c) muestra la opción para la aplicación los filtros en las imágenes.

Fase 1: Adquisición de la imagen.

En esta fase, actualmente estamos haciendo convenios con el Hospital Estatal de Cancerología de Guerrero. No dudando 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 usa 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.



(a) (b) (c)
Figura 4: interfaz gráfica, 4(a) Pantalla principal, 4(b) selección de imágenes, 4(c) Aplicación de Filtros.

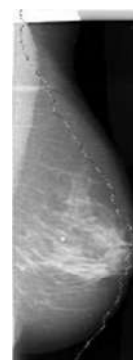
Fase 2: Preprocesamiento

Reducción del área de trabajo

En esta etapa se reduce 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 se utiliza el algoritmo *Seam Carving*, el cual puede ampliar o reducir las imágenes de mamografías sin alterar la calidad de la misma. En la imagen 4(a) se observa la aplicación el algoritmo *Seam Carving* podremos ver los resultados de este procedimiento con la reducción del tamaño, donde no solamente se 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 4(b). Esta línea representa los píxeles, *menos visibles*, eliminados. Esto ayudara en gran manera posteriormente puesto que puede reducir el costo computacional.



a) Aplicación del algoritmo Seam Carving



b) Eliminación de Píxeles

Figura 4 Algoritmo Seam Carving

Reducción de Ruido

En esta etapa se selecciona el filtro para *suavizar* los bordes de la imagen. Podemos observar en la figura 5(a) como se encuentran los bordes de la imagen original, de la cual solo interesan los bordes de las lesiones a detectar.

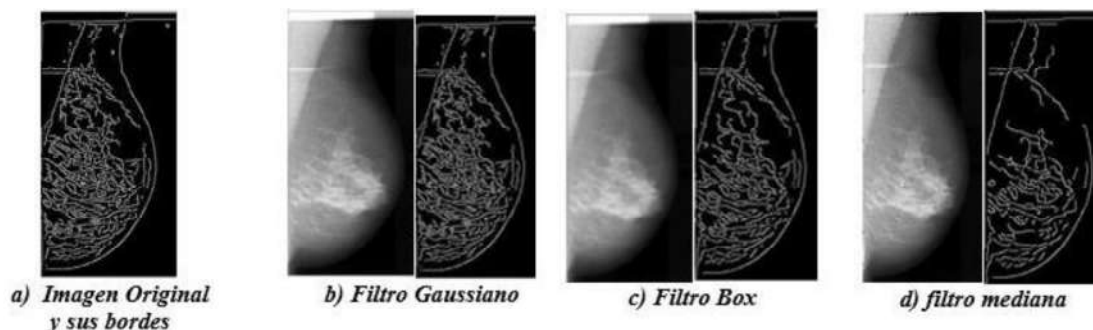


Figura 5 Filtros para bordes

En la imagen anterior se pueden visualizar todos los bordes con los la imagen original cuenta, la cual no es muy significativo. 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 (5b), Filtro Box (5c) y Filtro de Mediana (d).

Filtro Gaussiano: simula 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 5(b).

Filtro Box: El filtrado box implica reemplazar cada pixel 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).

En la figura 5(c) 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. En la figura 5(d) se muestra como el filtro mediana suaviza los bordes de la imagen de mamografía. En comparación con los otros filtros podemos observar que este tipo de filtrado suaviza aún más los bordes de la lesión, omitiendo los bordes innecesarios. Además, este filtro no difumina demasiado la mamografía en comparación con los otros filtros, lo cual se puede observar en la Figura 6. Es por eso que para este proyecto utilizaremos este tipo de filtrado.

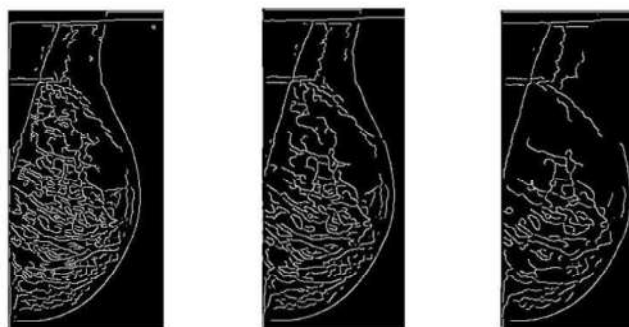
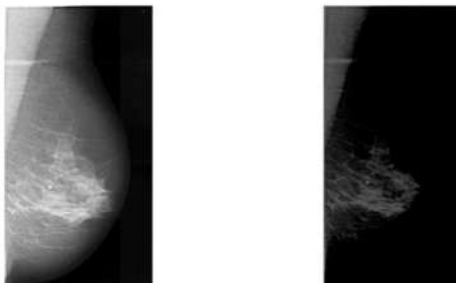


Figura 6 Comparación entre filtros Gaussiano, box y mediana

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 se usa el filtro de *Gabor*. A continuación se mostrará la aplicación de este filtro. El proceso comienza con la obtención de la imagen de la mamografía en su representación en escala de grises mostrada en la figura 7a.



a) Imagen con escala de grises b) Imagen con filtros de Gabor

Figura 7 Filtro de Gabor

Se construye el filtro de Gabor con un tamaño de ventana de 370X180 píxeles, θ de 100°, σ de 50, λ 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 7(b), donde podemos visualizar como se *depuraron* los elementos ajenos a la lesión de la mama. Al ver el resultado que arroja 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

En la actualidad, en México y en varias partes del mundo, el cáncer de mama es una de las enfermedades causantes de los más altos índices de muerte por cáncer, principalmente por ser diagnosticados en etapas avanzadas, es por eso que diversas investigaciones y trabajos enfocados a este padecimiento se han realizados en diferentes países del mundo, en busca de una solución para su diagnóstico temprano, haciendo uso de las tecnologías que existen y del conocimiento de expertos en esta área de la medicina con el objetivo de obtener resultados más rápidos y eficientes a la hora de emitir un diagnóstico. En México las herramientas para este fin que se han realizado son escasas debido a la complejidad del problema y la falta de aporte económico para investigación y desarrollo de este tipo de proyectos. En los países desarrollados existen sistemas de computadoras basados en técnicas de inteligencia artificial enfocados al diagnóstico médico, su aportación ha sido realmente notable, sin embargo en México y la mayoría de los países de la región de Latinoamérica se carece de este tipo de sistemas.

En este proyecto se desarrolló la construcción de una base de datos con registros mamográficos. Por lo tanto se generó una base de datos con imágenes de mamografías, donde se aplicó el pre-procesamiento de las imágenes para trabajos futuros. El algoritmo *Seam Carving* fue utilizado debido a su conveniencia de aplicarse sin importar cuál sea el tamaño de la imagen, ya que este algoritmo puede modificar el tamaño de la imagen sin alterar la calidad de la misma. Esto significa que además de estandarizar el tamaño de las imágenes, podrá reducir costo computacional. En el análisis de filtros para suavizar los bordes de la imagen, se determinó que el filtro de la mediana aparenta obtener mejores resultados, aunque no podemos tener la certeza que se utilizará dicho filtro en la BD futura, debido a diferentes factores como son la calidad de las imágenes y el tamaño, aunque es muy probable que no difieran mucho a la imagen obtenida por la BD pública. La BD creada cuenta ya con las imágenes con el área de interés delimitada (aplicando 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 es por esto que se creó la interfaz para el manejo de las imágenes lo cual facilitará las pruebas futuras. En conclusión, la aplicación del algoritmo *Seam Carving* y el filtro de Gabor arrojaron resultados bastante satisfactorios.

Como trabajos futuros y recomendaciones se plantea el continuo mejoramiento de la base de datos, además del registro de la información y el desarrollo de una plataforma para el libre acceso a la comunidad científica. Además, se plantea el desarrollo de otras metodologías para las siguientes etapas.

Referencias bibliográficas.

- 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.
- Deitel, H. M. (2008). *Cómo programar en C++* (Sexta ed.). Pearson.
- McDonnell, M. J. (1981). Box-filtering techniques. *Computer Graphics and Image Processing*, 65-70.
- Moya, J. S. (2014). Construcción de una base de datos de imágenes.
- Shamir, S. A. (2007). Seam carving for content-aware image resizing. *ACM Trans. Graph.*, 10.

Suckling, J. (2015). Mammographic Image Analysis Society Database. *Mias*.
Wang, Z. (2015). A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. 6. 54.