

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
FACULTAD DE CIENCIAS E INGENIERÍA



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL PERÚ

PROPUESTA DE TÉCNICA PARA LA ESTIMACIÓN DE DENSIDAD
MAMARIA EN MAMOGRAFÍAS DIGITALES UTILIZANDO
TÉCNICAS DE SEGMENTACIÓN Y ESTIMADORES
GLOBALES

Tesis para optar el Título de Ingeniero Electrónico, que presenta el bachiller:

César Whesly Segura Del Río

ASESOR: Ph.D. Benjamín Castañeda Aphan

Lima, setiembre del 2008

RESUMEN

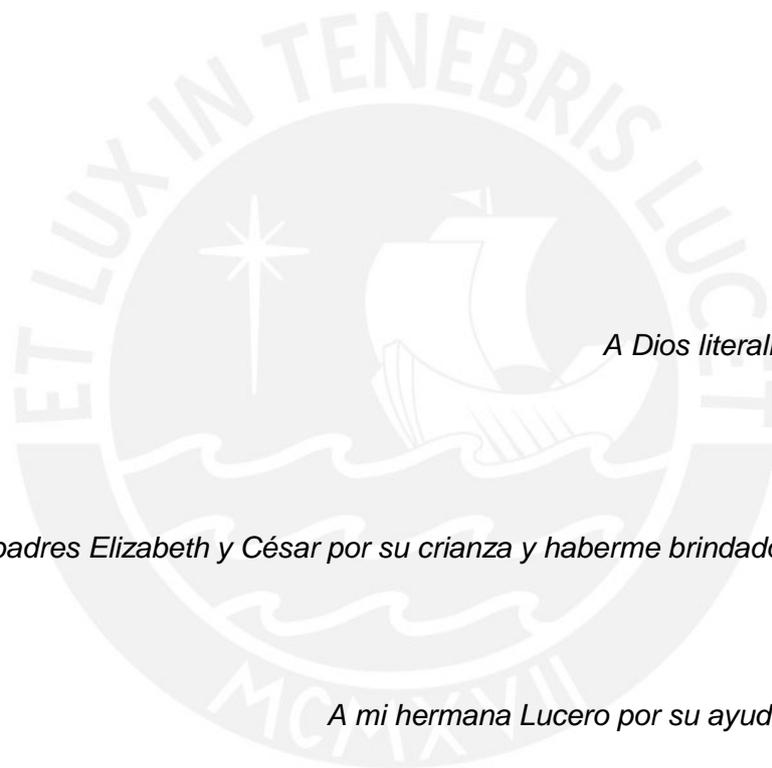
Anualmente un millón y medio de nuevos casos de cáncer de mama son diagnosticados alrededor del mundo. Por tal motivo, se han establecido estrategias de despistaje temprano de cáncer de mama que utilicen a la mamografía como principal herramienta. Adicionalmente, se han evaluado diversas variables clínicas que han sido capaces de estimar el riesgo a desarrollar esta neoplasia. Entre estas variables se encuentra la densidad mamaria o tejido denso catalogado como principal indicador del cáncer de mama. La densidad mamaria o tejido denso de la mama incrementa de 3.5 a 5 veces el riesgo de adquirir cáncer de mama.

La mamografía es una imagen formada a partir del principio de rayos X, describe a los tejidos de la mama y permite la evaluación del estado clínico de los pacientes. Se utilizan dos acercamientos para su estudio, estos consisten en el análisis cualitativo y cuantitativo. En el primero se utiliza el criterio de la observación del médico para dar su juicio de cada caso. En el segundo se procesan digitalmente mamografías y se les aplican técnicas de segmentación, luego se utilizan métodos cuantificables como estimadores globales. La segmentación consiste en subdividir una mamografía en los tejidos que la constituyen mientras que los estimadores globales son algoritmos que se aplican a la imagen completa y brindan indicadores cuantitativos de esta.

El objetivo de la presente tesis es realizar una comparación entre el tejido denso obtenido mediante los técnicas de segmentación de Bden, bimodal, C-medias difusas y Otsu con el establecido por un patrón de oro (segmentación manual de experto). A partir de las comparaciones, se propondrá un algoritmo de segmentación validado en base a estimadores globales. Para alcanzar dicho objetivo, se cumplirán los siguientes objetivos específicos: implementación de técnicas de segmentación de Bden, bimodal, C-medias difusas y Otsu para calcular el tejido denso; implementación de estimadores globales intersección sobre unión, índice de similitud estructural (SSIM) y norma L2; comparación cuantitativa del tejido denso entre mamografías segmentadas y mamografías del patrón de oro; implementación de propuesta de técnica de segmentación y cálculo de tejido denso de la mama.

Se ha encontrado que la propuesta de técnica de segmentación implementada para el cálculo de tejido denso es superior a las demás técnicas en base a estimadores globales. La técnica propuesta también demuestra un mayor grado de correlación con el tejido denso calculado por los médicos que las segmentaciones de Bden, bimodal, C-medias difusas y Otsu.

Se ha encontrado que en base a los estimadores globales de intersección sobre unión y SSIM la técnica de segmentación propuesta presenta mejores resultados que las técnicas implementadas de Bden, bimodal, C-medias difusas y Otsu. El coeficiente de Pearson es el mejor indicador del grado de correlación que existe entre los porcentajes de tejido denso calculado por los médicos y la técnica propuesta que con el calculado por las demás técnicas.



A Dios literalmente por todo

A mis padres Elizabeth y César por su crianza y haberme brindado educación en estos 22 años

A mi hermana Lucero por su ayuda incondicional

A mi asesor, Ph. D. Benjamín Castañeda Aphan por involucrarme dentro del mundo de la investigación, motivar mi interés en publicaciones de artículos científicos y permitirme aprender de todo su conocimiento y experiencia

A mi co-asesora, Ing. Elizabeth Huaroc por ser una excelente guía, brindarme sus enseñanzas, compartir sus experiencias y permitirme mejorar mi trabajo semana a semana

A todos los miembros del Laboratorio de Imágenes Médicas (LIM) por contestar mis dudas, compartir nuevos conocimiento y convertirse en mis amigos

Agradezco al Fondo de Financiamiento de la Pontificia Universidad Católica del Perú por haberme permitido el desarrollo del proyecto titulado “Herramienta para evaluar el riesgo de cáncer de mama en la población peruana a partir de la estimación de la densidad mamaria en mamografías digitales”. Se me brindaron la base de datos mamográfica y recursos para llevarlo a cabo



INDICE	
INTRODUCCIÓN	1
CAPÍTULO 1	
1. Importancia, descripción y formulación del problema	2
1.1. Importancia del examen radiológico de la mama	2
1.2. Descripción y formulación del problema	3
1.3. Hipótesis	6
1.4. Justificación del trabajo	7
CAPÍTULO 2	
2. Técnicas de segmentación y estimadores globales en mamografías digitales	8
2.1. Estado del Arte	9
2.2. Técnicas de segmentación en mamografías digitales	10
2.2.1. Segmentación Unimodal	9
2.2.2. C-medias Difusas	11
2.2.3. Segmentación Bimodal	13
2.2.4. Segmentación de Otsu	15
2.2.5. Segmentación de Bden	16
2.3. Estimadores globales para la estimación de la densidad mamaria	19
2.3.1. Índice de Similitud Estructural (SSIM)	19
2.3.2. Función Norma - L2	20
2.3.3. Método de Intersección sobre Unión	20
CAPÍTULO 3	
3. Implementación de técnicas de segmentación y estimadores globales en mamografías digitales	21
3.1. Objetivos	21
3.2. Técnicas de segmentación de tejido denso	22
3.2.1. Segmentación unimodal	22
3.2.2. C-medias difusas	24
3.2.3. C-medias difusas a partir de MATLAB	27
3.2.4. Segmentación de Otsu	29
3.2.5. Segmentación bimodal	32
3.2.6. Segmentación de Bden	34
3.2.7. Propuesta de técnica para segmentar tejido denso	38
3.3. Estimadores globales	42
3.3.1. Intersección sobre unión	42
3.3.2. Índice de similitud estructural – SSIM	43
3.3.3. Norma L2	43
CAPÍTULO 4	
4. Pruebas y resultados	44
4.1. Resultados de estimadores globales por técnica	44
4.2. Selección de la mejor técnica en base a los estimadores globales	45
4.3. Resultados del grado de correlación entre técnicas	46
4.4. Discusión	50
CONCLUSIONES	52
RECOMENDACIONES	53
BIBLIOGRAFÍA	

Índice de figuras

Figura 1. Componentes principales del mamógrafo	4
Figura 2. Imagen adaptada de la estructura de la mama	5
Figura 3. Ligamentos de Cooper de la mama	5
Figura 4. Vista medio-lateral oblicua (MLO) de la mama	6
Figura 5. Vista cráneo-caudal (CC) de mama	6
Figura 6. Relación entre el tejido adiposo y denso	7
Figura 7. Selección de umbral en segmentación unimodal	12
Figura 8. (a) Procedimiento de selección de umbral. (b) Imagen Segmentada por método unimodal	12
Figura 9. Clasificación de mamografía. (a) Mamografía original y (b) Mamografía clasificada en dos niveles	14
Figura 10. Histograma con distribución bimodal	14
Figura 11. (a) Mamografía Original. (b) Segmentación Bimodal de mamografía	15
Figura 12. Histograma 2D que se divide en cuatro partes: la mama, los bordes y el ruido.	16
Figura 13. Umbralización de Otsu. (a) Mamografía original y (b) Mamografía umbralizada.	17
Figura 14. Resonancia magnética de la mama de 640x640 píxeles.....	19
Figura 15. Elemento estructural B de 9x9 píxeles	19
Figura 16. Imagen C es la erosión de A con B	19
Figura 17. Diagrama de bloques para estimado global “Intersección sobre Unión”...20	20
Figura 18. Mamografía sin pre-procesamiento	21
Figura 19. Diagrama de bloques segmentación unimodal	20
Figura 20. Mamografía normalizada de segmentación unimodal.....	23
Figura 21. Histograma normalizado de segmentación unimodal.....	23
Figura 22. Mamografía en complemento a 1 de segmentación unimodal.....	24
Figura 23. Histograma de mamografía en complemento a uno	24
Figura 24. Procedimiento del trazado de rectas en histograma unimodal	24
Figura 25. Mamografía segmentada por método unimodal.....	24

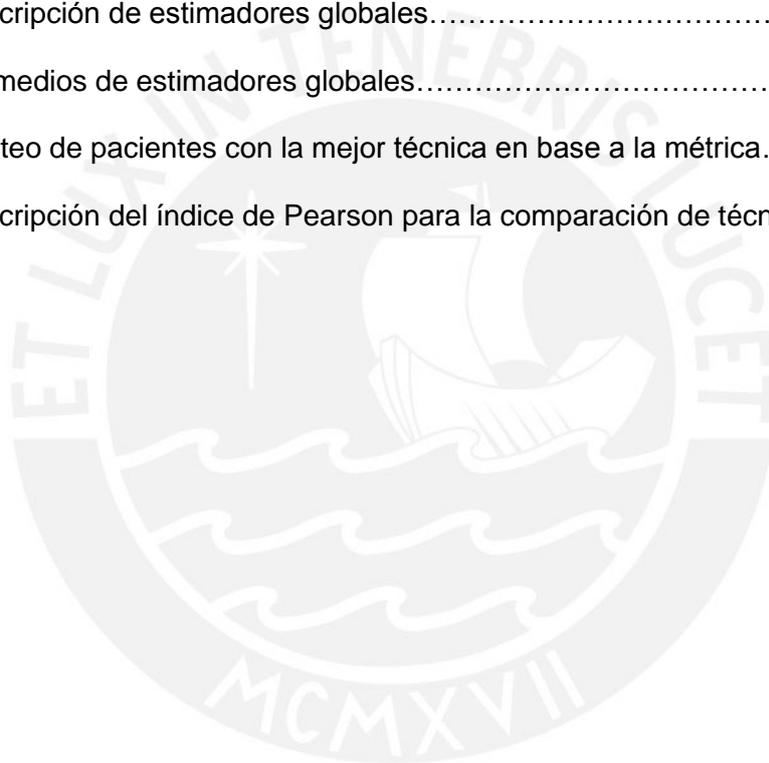
Figura 26. Mamografía en complemento a 1 de técnica C-medias difusas.....	25
Figura 27. Diagrama de bloques C-medias difusas	25
Figura 28. Puntos rojos (centros del grupo 1) y puntos verdes (centros del grupo 2)...	26
Figura 29. Segmentación luego de una primera iteración de C-medias difusas	26
Figura 30. Segmentación final de Medias Difusas de C (iteración n° 6).....	27
Figura 31. Mamografía no pre-procesada para C-medias difusas a partir de MATLAB	28
Figura 32. Mamografía normalizada para C-medias difusas a partir de MATLAB.....	28
Figura 33. Mamografía en complemento a 1 para C-medias difusas a partir de MATLAB	28
Figura 34. Segmentación de mamografía por C-medias difusas a partir de MATLAB..	29
Figura 35. Diagrama de bloques de segmentación de Otsu	30
Figura 36. Mamografía normalizada y en complemento a 1 utilizada en Otsu.....	30
Figura 37. Histograma de mamografía normalizada.....	31
Figura 38. Varianza entre clases de Otsu.....	31
Figura 39. Resultado segmentación de Otsu.....	31
Figura 40. Mamografía Original utilizada en segmentación Bimodal	32
Figura 41. Mamografía normalizada utilizada de Otsu.....	32
Figura 42. Mamografía en complemento a 1 de Otsu	32
Figura 43. Diagrama de bloques de la técnica de segmentación bimodal.....	32
Figura 44. Mamografía segmentada por umbral $T_0 = 0.2$	33
Figura 45 Intensidades de píxeles mayores a " T_0 " se suman en variable "P1".....	34
Figura 46. Resultado segmentación bimodal.....	34
Figura 47. Mamografía normalizada de segmentación de Bden	34
Figura 48. Diagrama de bloques de segmentación de Bden	35
Figura 49. Búsqueda de umbral de mediana en histograma	35
Figura 50. Mamografía segmentada con el umbral de mediana de Figura 47.....	36
Figura 51. Complemento en 1 de Figura 50.....	36
Figura 52. Filtro de media aplicado a Figura 51.....	37

Figura 53. Cálculo de bordes de Sobel de Figura 52.....	37
Figura 54. Resta de mamografía umbralizada con los bordes de la misma	37
Figura 55. Primera erosión de la mamografía proveniente de la resta anterior	37
Figura 56. Erosión final de mamografía restada, Resultado de segmentación de Bden	37
Figura 57. Diagrama de bloques de propuesta de segmentación de tejido denso	38
Figura 58. Pre-procesamiento: Mamografía normalizada y en complemento a 1 propuesta de segmentación	39
Figura 59. Histograma de Figura 58.....	39
Figura 60. Mamografía umbralizada de Figura 58 con marca de agua.....	39
Figura 61. Mamografía sin marca de agua de Figura 60	40
Figura 62. Mamografía con área de la mama rellenada completamente.....	40
Figura 63. Mamografía resultante de pre-procesamiento.....	40
Figura 64. Ecuación por expansión de histograma de Figura 63.....	40
Figura 65. Umbralización de mediana por histograma	41
Figura 66. Mamografía segmentada por método bimodal.....	41
Figura 67. Bordes de la mama de Figura 66.....	41
Figura 68. Resta de los bordes: Figura 66 menos Figura 67.....	41
Figura 69. Resultado de la propuesta de técnica de segmentación de tejido denso ...	42
Figura 70. Imagen de muestra “A”	42
Figura 71. Imagen de muestra “B”	42
Figura 72. Intersección entre A y B	42
Figura 73. Unión entre A y B.....	42
Figura 74. Muestra la distribución de los casos clínicos para cada técnica dependiendo de la métrica que se utilice	46
Figura 75. Porcentaje de densidad manual del médico vs. Bden	47
Figura 76. Porcentaje de densidad del médico vs. C-medias difusas	47
Figura 77. Porcentaje de densidad manual del médico vs. bimodal	48
Figura 78. Porcentaje de densidad manual del médico vs. Otsu.....	48

Figura 79. Porcentaje de densidad del médico vs. propuesta de técnica de segmentación de tejido denso	49
---	----

Índice de tablas

Tabla 1. Representación de la máscara de Sobel por sus componentes.....	18
Tabla 2. Máscara de Sobel en el eje x.....	18
Tabla 3. Máscara de Sobel en el eje y.....	18
Tabla 4. Actualización de valores de los centros de grupo "v".....	27
Tabla 5. Descripción de estimadores globales.....	43
Tabla 6. Promedios de estimadores globales.....	45
Tabla 7. Conteo de pacientes con la mejor técnica en base a la métrica.....	46
Tabla 8. Descripción del índice de Pearson para la comparación de técnicas.....	49



INTRODUCCIÓN

En Perú la cantidad de casos detectados de cáncer de mama en el año 2011 fueron de 1191 nuevos casos, mientras que en los últimos 10 años se han diagnosticado 4800 nuevos casos clínicos en ambos sexos, según datos del INEN. El incremento en el número de casos viene acompañado con la detección de la patología en sus etapas iniciales, intermedias o avanzadas. Para etapas avanzadas solamente un 3.3% de pacientes sobreviven el cáncer de mama. Por ello es importante la detección temprana de esta neoplasia o enfermedad, además de un examen médico que sea eficaz y no tome mucho tiempo. La técnica de la mamografía digital cumple estos requisitos y se hace presente en la lucha contra el cáncer de mama.

El uso de la mamografía se inició en el año 1965 con la construcción del mamógrafo, el primer equipo médico que generaba imágenes radiográficas de la mama. Más tarde con el desarrollo tecnológico en el área de la bioingeniería se construyeron mamógrafos digitales. La mamografía tiene una gran importancia porque permite detectar volúmenes densos en el tejido de la mama de la mujer ya que estos incrementan de 3.5 a 5 veces la posibilidad de desarrollar cáncer de mama. El ratio que existe entre el tejido denso y todo el tejido de la mama es catalogado como la densidad mamaria.

La evaluación de una mamografía digital consiste en realizar un despistaje de la densidad mamaria, existen dos acercamientos para la evaluación de los pacientes: análisis cualitativos y cuantitativos. En el primer caso un radiólogo evalúa la mamografía y brinda un diagnóstico en base al criterio médico. En el segundo caso se utilizan técnicas asistidas por ordenador en las cuales se procesan digitalmente las imágenes tomadas del mamógrafo digital.

El objetivo de la presente tesis es desarrollar una propuesta de técnica computacional que permita el diagnóstico asistido por computadora y que está basada en algoritmos de procesamiento de imágenes que sea capaz de estimar de manera cuantitativa la densidad de la mama en 100 imágenes mamográficas digitales. La base de datos con la que se cuenta es una muestra del estado clínico de los casos de cáncer de mama en la población peruana y es importante para conocer la densidad mamográfica de una muestra de la población peruana.

A los médicos que realicen un diagnóstico de paciente se les asiste con herramientas de "diagnóstico asistido por computador" o CAD (de sus siglas en inglés). Los CAD están conformados por una gama de técnicas para el procesamiento de imágenes entre las cuales se encuentran las mamografías digitales. Estas técnicas se componen de la segmentación, el uso de histogramas y la clasificación. La técnica de segmentación que se propone estima el tejido denso de la mama y además sus resultados son validados mediante estimadores globales. Finalmente con la técnica se calcula a la variable clínica de densidad mamaria.

Capítulo 1

Importancia, descripción y formulación del problema

1.1. Importancia del examen radiológico de la mama

El cáncer de mama es la causa más común de muerte en mujeres en todo el mundo (el quinto caso de muerte en ambos sexos), en el 2008 murieron 460,000 personas debido a esta neoplasia y la incidencia de casos en la actualidad no se ha incrementado desde el 2008 debido principalmente a diagnósticos tempranos y mejores tratamientos [1]. En el año 2011 se detectaron 1191 nuevos casos de cáncer de mama en mujeres en el Instituto Nacional de Enfermedades Neoplásicas [2], mientras que en los últimos 10 años se han diagnosticado más de 4800 nuevos casos clínicos tanto en hombres y mujeres [3]. Las mamografías digitales se utilizan como primera herramienta para evaluar el riesgo de cáncer de mama. Las mamografías digitales son exámenes de diagnóstico temprano del cáncer de mama y son más accesibles económicamente que imágenes de resonancia magnética. Son versátiles porque el tiempo que toma un examen de mamografía digital es menor al de una resonancia magnética. Finalmente brindan al médico una imagen con el que llega a conocer el panorama general del paciente porque puede realizar un análisis preliminar de la densidad mamaria.

La Organización Panamericana de la Salud (OPS, por sus siglas en inglés) ha reportado para América Latina y El Caribe una disminución en la mortalidad para los países de Argentina, Brasil, Canadá, Chile, E.E.U.U, México, Nicaragua, Paraguay y Venezuela; sin embargo, en El Caribe se reportan la mayor cantidad de muertes de todo el bloque de países.

Se han reportado métodos cuantitativos y cualitativos para realizar la evaluación de mamografías digitales. Dentro de los métodos cualitativos está el que planteó John Wolfe y divide en cuatro patrones, denominado niveles de Wolfe, para caracterizar la mama. El patrón N representa al tejido adiposo que es explicado más adelante en el capítulo 1, el patrón P1 y P2 es un tejido más endurecido que el patrón N, y el patrón DY le pertenece a las áreas muy endurecidas (tejido denso de la mama). El tejido más endurecido indica el riesgo más alto de tener cáncer de mama [37, 38].

Métodos cuantitativos iniciales para evaluar el tejido denso de la mama en pacientes consistían en técnicas como la planimetría, método que utilizaba al planímetro cuya función era medir el área de tejido denso en mamografías. Se realiza trazando los bordes de tejido denso y luego sumando estas áreas limitadas, se prosigue con el cálculo del área de toda la mama y se finaliza hallando el ratio que existe entre el área de tejido denso y de toda la mama [35,36]. En un estudio se evaluó a 266 casos y se midió el área del tejido más endurecido en la mama, es decir, el tejido denso fue utilizado como principal indicador del riesgo de adquirir cáncer de mama. Los resultados fueron susceptibles a errores en el cálculo del tejido denso [4]. Sin embargo, técnicas de procesamiento de imágenes modernas como la segmentación, clasificación y el análisis de histogramas son utilizados

en mamografías digitales para realizar un diagnóstico asistido por ordenador que apoya al médico en la evaluación y diagnóstico de sus casos.

Otras formas adicionales a la formación de imágenes por rayos-X de la mama son la resonancia magnética y ecografía. Estas son técnicas no invasivas porque no dejan residuos de radiación en el cuerpo a diferencia de la radiografía.

1.2. Descripción y formulación del problema

El mamógrafo digital es la herramienta utilizada para la formación de imágenes radiográficas de la mama. Existen dos sistemas: el primero se llama mamografía digital indirecta o CR y utiliza una pantalla de fósforo que se carga eléctricamente luego del disparo de rayos X, luego se escanea y se forma una imagen de la mama; el segundo se llama mamografía digital directa DR porque los rayos X atraviesan la mama y generan la imagen digital de manera directa. En la Figura 1 se observa al mamógrafo en su versión analógica y digital, cada una de las partes que lo constituyen juegan un papel importante en el proceso de formación de imágenes médicas ya que la alteración en una de sus etapas influenciará a los demás elementos y evitará el principal objetivo del aparato que es obtener imágenes detalladas de alta resolución y que permitan observar la áreas de tejido denso de la mama [5].

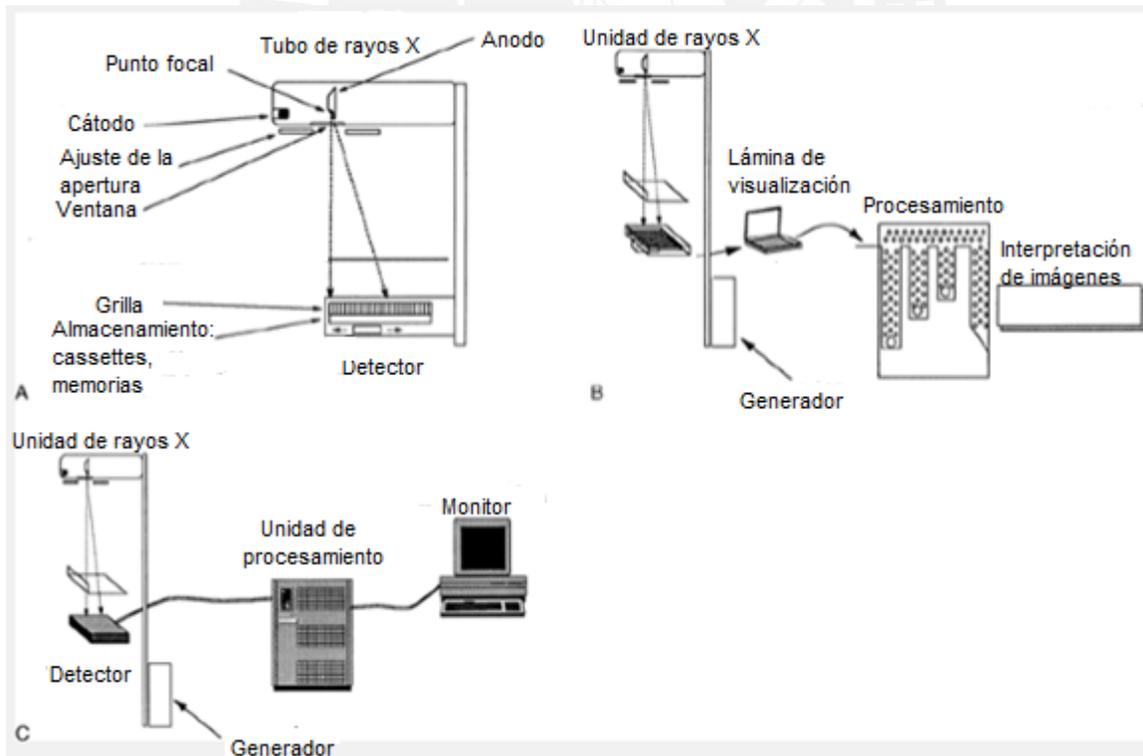


Figura 1. Componentes principales de mamógrafos adaptado de [6]. (A) Estructura de un mamógrafo estándar. (B) Estructura de un mamógrafo analógico. (C) Estructura de un mamógrafo digital.

Las mamas están conformadas por tres principales clases de tejido: adiposo o graso, conectivo y glandular, además de vasos capilares. Debajo de la piel o tejido conectivo se localiza al tejido adiposo que se extiende hasta la zona pre-pectoral. En la figura 2 se muestran los tejidos adiposo y glandular de la mama. Se observan los lóbulos glandulares (las unidades básicas del tejido glandular), estas se conectan a través de tejido conectivo intra-lobular. En su conjunto forman el tejido glandular, también se muestra al tejido adiposo que ocupa el mayor porcentaje de área de la mama y la piel que lo limita o tejido conectivo.

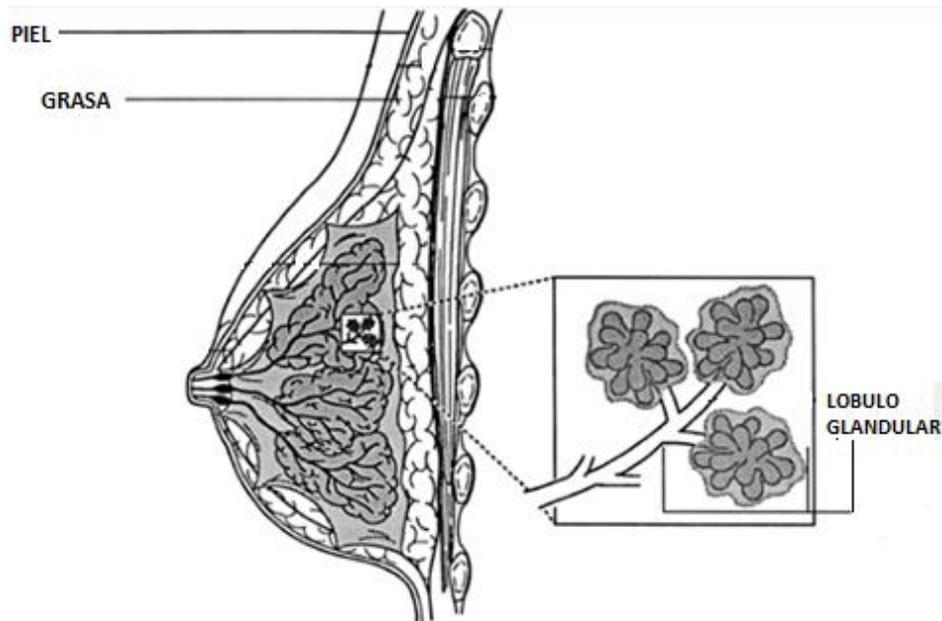


Figura 2. Imagen adaptada de la estructura de la mama. [7]

En la Figura 3 se muestran los “Ligamentos de Cooper” que funcionan como sostenimiento para todos los tejidos de la mama. A estos se les llama “parénquima” en la terminología médica.

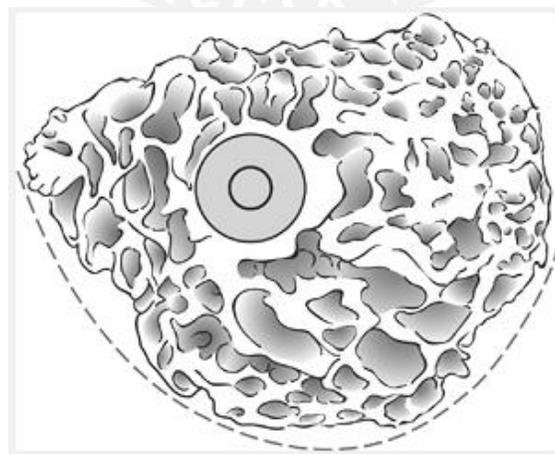


Figura 3. Ligamentos de Cooper de la mama. [8]

El mamógrafo digital genera dos vistas: medio-lateral oblicua y cráneo-caudal (MLO y CC por sus siglas en inglés, respectivamente). La vista medio-lateral oblicua se obtiene posicionando a la fuente de rayos X de manera lateral a la mama (figura 4), mientras que la imagen cráneo-caudal es una toma de la cara superior que parte de la cabeza y recorre con dirección vertical hasta llegar a la mama (figura 5). De manera práctica la imagen CC solo se utiliza para acceder a tejidos que se encuentran más cercanos al pecho y oscurecen lesiones de una proyección MLO [9].

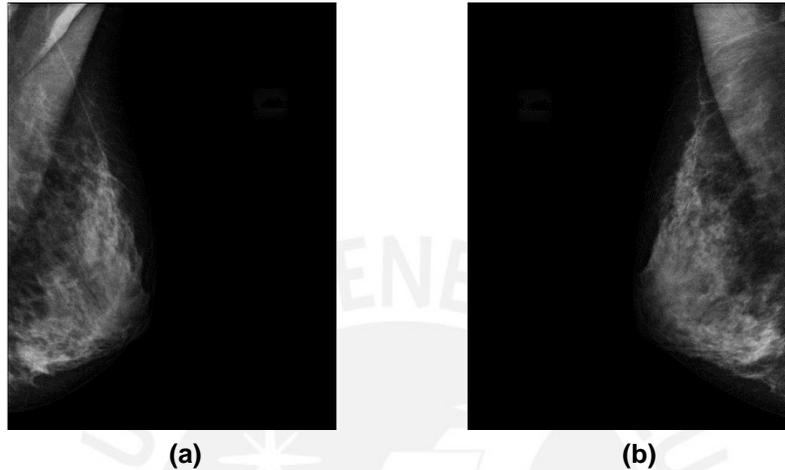


Figura 4. (a) Vista Medio-Lateral Oblicua izquierda y derecha (b) [10].

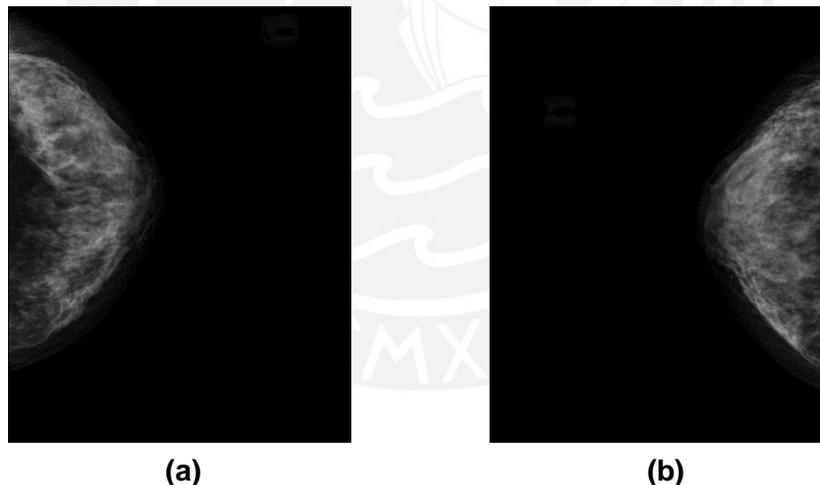


Figura 5. (a) Vista Cráneo-Caudal izquierda y derecha (b) [10].

Acerca de la observación cualitativa de mamografías; médicos radiólogos se encargan de evaluar y diagnosticar el estado clínico del paciente; sin embargo, la observación de las mamografías es susceptible a equivocaciones. Los errores de diagnóstico pueden ser de dos tipos: falsos positivos para aquellos que no tienen cáncer pero se les evaluó como si lo tuvieran y falsos negativos para personas que tienen cáncer y se les evaluó como si no lo tuvieran. De manera cuantitativa se utilizan herramientas computacionales como “diagnóstico asistido por computador” o CAD brindan resultados que dependen del acercamiento que se le dio al algoritmo para obtener indicadores de la densidad mamaria, entre las principales dificultades están las siguientes: los resultados que se obtienen mediante técnicas cuantitativas del procesamiento de mamografías digitales varían de

acuerdo a la aplicación que se busque (obtención de bordes, formas de la textura o zonas de rápido cambios); los tejidos adiposo, piel y glandular se solapan, esto hace difícil enfocarse en uno solo de ellos. Para el objetivo de esta tesis el tejido más importante es el glandular ya que al endurecerse forma tejido denso que pueden llegar a formar microcalcificaciones (volúmenes densos pequeños en la mama) [11].

En la figura 6 se muestran dos mamografías de la misma persona tomada en fechas diferentes. En la primera mamografía la persona tenía un mayor índice de grasa corporal. En la segunda, la persona disminuyó este índice de grasa corporal. Aquí se muestra la dificultad que tienen los médicos para diferenciar el tejido denso del adiposo en la mama.

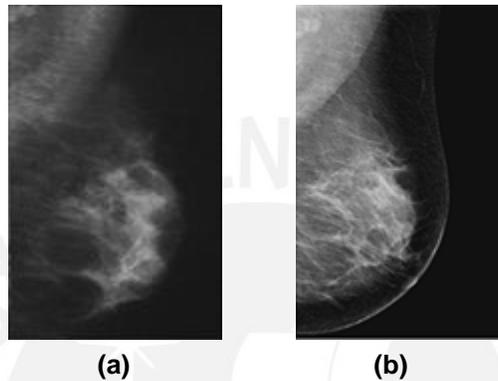


Figura 6. Relación entre tejido adiposo y denso. (a) Mamografía MLO de paciente. (b) Mamografía MLO de la misma paciente luego de haber perdido peso. [13]

En el caso del estudio de la densidad mamaria en la población peruana las dificultades cuantitativas son la selección de técnica o algoritmo para detectar el tejido denso y que este resultado sea exacto para que no se confundan los tejidos. Requiere seguir un procedimiento análogo al de los radiólogos expertos en la observación de mamografías. Se sigue la secuencia de identificar la zonas más brillantes y discriminar de si estas son tejido denso para luego graficar los bordes de las áreas. Replicar la tarea de los radiólogos es la dificultad y esta se hace utilizando a una base de datos de 100 mamografías digitales.

1.3. Hipótesis

Las técnicas de segmentación son utilizadas para procesar las mamografías y obtener una imagen binaria. Los píxeles de color blanco muestran al tejido de interés para esta tesis como lo es el tejido denso, mientras que los píxeles de color negro muestran el fondo de la imagen (píxeles oscuros alrededor de la mama) y tejidos de la mama que no son de interés. A partir de la implementación de las segmentaciones unimodal, bimodal, Bden, C-means difusas, Otsu y la técnica propuesta es posible calcular el tejido denso que se asemeje al patrón de oro de los médicos radiólogos, esto resultados serán probados de manera cuantitativa.

La propuesta de segmentación de tejido denso de la mama se desarrolla tomando las mejores características de cada una de las técnicas previamente implementadas. Por consiguiente tiene mejores resultados que todas las demás en el cálculo de tejido denso.

Se prueba su superioridad al compararlo con el patrón de oro de los radiólogos utilizando a estimadores globales y grado de correlación entre imágenes.

Finalmente, los resultados del trabajo provendrán de la comparación de los estimadores globales y densidad mamaria ordenados en tablas en las que se indicará la superioridad de la técnica propuesta

1.4. Justificación del trabajo

La densidad mamaria es catalogada como principal indicador para la detección del cáncer de mama. Si una mama es muy densa esto incrementa la posibilidad de 3.5 a 5 veces de haber adquirido la neoplasia. Por consiguiente es importante estimar la densidad mamaria en mamografías ya que cuando el paciente conoce esta variable clínica puede realizarse exámenes médicos de mayor exactitud como exámenes de resonancia magnética.

Si una mama es muy densa se le puede identificar las zonas más endurecidas. En estas zonas es muy probable encontrar tumores pequeños a los que se les denomina microcalcificaciones. Las microcalcificaciones son difíciles de detectar cuando se encuentran más cerca a la zona pectoral y se pueden llegar a confundirse con otros tejidos de la mama. Es por ello que se plantea el desarrollo de una propuesta de técnica que estime el tejido denso de la mama de manera exacta en base a implementaciones previas de las técnicas de C-medias difusas, bimodal, Bden y Otsu. La eficacia de esta herramienta es medida de manera cuantitativa para validar su eficacia.

Si bien el tejido denso no forma parte de los tejidos de la mama (glandular, adiposo y conectivo), se origina en estos. Dentro del tejido glandular se encuentran los lóbulos glandulares, es en estos lóbulos donde se origina el tejido denso y se incrementa dándole a la mama la denominación de densa [11]. El siguiente factor a tomar en cuenta debe ser la distribución de estos tejidos, pueden estar esparcidos por la zona de la mama o en caso contrario, muy condensados. Esta es otra razón por la que se realiza el planteamiento del desarrollo de una técnica de segmentación que pueda ser validada por médicos radiólogos expertos.

Capítulo 2

Técnicas de segmentación y estimadores globales en mamografías digitales

En este capítulo se desarrolla la teoría de las técnicas de segmentación y estimadores globales que se implementan en el capítulo tres, se describen la secuencia de pasos, fórmulas y consideraciones realizadas durante el desarrollo de las técnicas de segmentación de C-medias difusas, bimodal, Bden, Otsu y técnica propuesta.

2.1. Estado del arte en técnicas de estimación de la densidad mamaria

En la actualidad se utilizan técnicas CAD para estimar el tejido denso de la mama cada una con diferentes niveles de exactitud para realizar la evaluación de pacientes con el riesgo de adquirir el cáncer de mama y además detectar tumores en las mamas.

La característica principal que diferencia a cada una de las técnicas CAD es la exactitud de los resultados para una correcta estimación de la densidad mamaria a partir de mamografías. La extracción de características tales como el nivel de contraste e intensidad de mamografías forman parte de la base de datos de técnicas asistidas por ordenador y consisten en etapas de pre-procesamiento de la imagen. El método de detalles por contraste ha sido mejorado utilizando algoritmos vectoriales de soporte (SVM, por sus siglas en inglés) que permiten un mejor tratamiento sobre los procesos no lineales que se aplican a las imágenes [14]. Computadoras modernas en hospitales se encargan de realizar procesos de segmentación simultánea en el cual se aplican las siguientes técnicas: “K-medias”, “C-medias” también conocidas como algoritmos de clasificación (Clustering Algorithm, de su nombre original) [22, 23, 24] y “Redes Neuronales” mediante modelos gaussianos; la combinación de técnicas resulta en una mejora en el grado de exactitud del procedimiento específico sobre procedimientos generales. Utilizar una combinación de estas técnicas es aplicable a biopsias del tejido mamario (en las cuales se forman imágenes microscópicas digitales que son procesadas), aquí las áreas de la citología, medicina y electrónica están envueltas en este procedimiento que resulta en características morfológicas, topológicas y de tipos de textura de las mamas clasificadas con una exactitud del 96 al 100% [15]. Métodos de “Maximización de la Expectación” segmentan tejido denso de las mamas [16]. Finalmente para identificar la formación de microcalcificaciones en mamografías se utiliza la segmentación acompañado de técnicas que utilizan “curvelets” y “wavelets” [17].

2.2. Técnicas de segmentación y estimadores globales en mamografías digitales

Segmentación significa división, es decir solo utilizar aquella parte de la imagen que es de interés para el estudio. Al segmentar se encuentran a los tejidos de interés separados de los que no lo son, para el caso de una mamografías es el tejido denso. Luego de segmentar

imágenes de la mama solo se muestran 2 intensidades de color: blanco para aquellas que son el tejido y negro para todo lo restante.

Actualmente se utilizan técnicas de clasificación y segmentación en mamografías digitales. La segmentación utiliza dichas técnicas de clasificación y segmentación. Estas se utilizan para estimar el tejido denso de la mama y cuentan con una gama de métodos que se utilizan para obtener características tales como la textura de tejidos, división en dos o más clases, cálculo de bordes y aclaramiento de zonas oscuras que durante evaluaciones médicas no son visibles. Las técnicas que se explican a continuación son los algoritmos de segmentación unimodal, C-medias difusas (FCM, por sus siglas en inglés), Otsu, bimodal y segmentación de Bden.

2.2.1. Segmentación Unimodal

El primer algoritmo es la segmentación unimodal planteado por Paul L. Rosin [18]. El método consiste de los siguientes pasos:

Primero se grafica el histograma de la mamografía. Para explicar al histograma se debe representar a las mamografías por matrices de “ m ” filas y “ n ” columnas. Cada uno de estos elementos son denominados píxeles y tienen una determinada intensidad de gris. El histograma es una gráfica de la frecuencia de los píxeles versus la intensidad de gris de cada píxel.

Dentro de este histograma se identifica a la máxima frecuencia (mayor cantidad de repeticiones de un pixel de cierta intensidad), a la mínima frecuencia (el pixel con menor número de repeticiones). Luego se trazó una recta que une al valor máximo y mínimo del histograma. A continuación muestra a la distribución de un histograma unimodal ideal (Figura 7) y el proceso de selección del umbral.

a. Método de segmentación unimodal

El método de segmentación unimodal requiere de un histograma unimodal, esto quiere decir que el histograma tenga una sola concavidad. Se traza una recta que pasa por el valor máximo y mínimo del histograma de la mamografía. La (Figura 7) muestra este procedimiento de manera ideal mientras que en la (Figura 8) es de manera práctica.

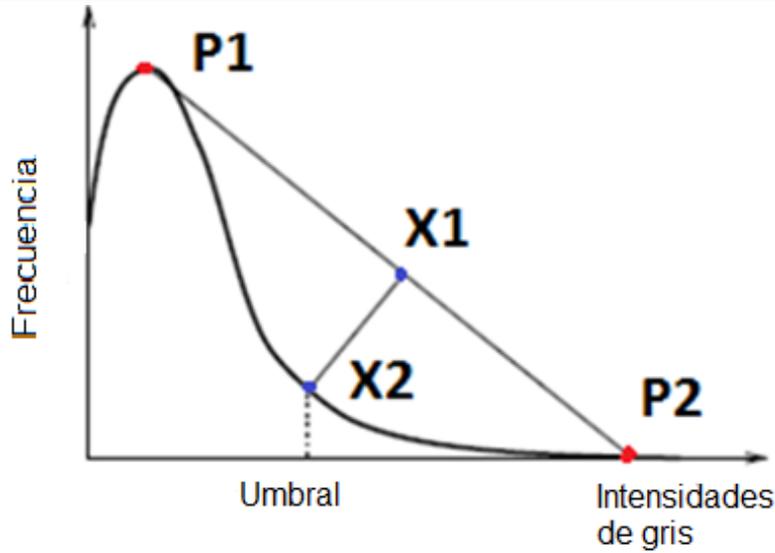


Figura 7. Segmentación unimodal, selección del umbral a partir de la maximización de la distancia de una mamografía en escala de grises. Imagen adaptada de [19].

La recta $\overline{P_1P_2}$ (Figura 7) está compuesto por un vector de 256 puntos y se define de la siguiente manera:

Paso 1. Con P_1 y P_2 se calcula la dirección \vec{v} de la recta

$$\overline{P_1} = (argimax, Hmax) \quad (1)$$

$$\overline{P_2} = (argimin, Hmin) \quad (2)$$

P_1 : punto inicial

P_2 : punto final

$Hmax$: frecuencia máxima de píxel

$Hmin$: frecuencia mínima de píxel

$argimax$: intensidad máxima de píxel

$argimin$: intensidad mínima de píxel

$$\vec{v} = \overline{P_1} - \overline{P_2} \quad , \text{ dirección de la recta}$$

Paso 2. Con \vec{v} y $\overline{P_1}$ se calcula la recta $\vec{l}(t)$ que pasa por $\overline{P_1P_2}$

$$\vec{l}(t) = \overline{P_1} + t \cdot \vec{v} \quad , \forall t \in T \quad (3)$$

El vector T cuenta con 256 posiciones:

$$T = \{0, \frac{1}{256}, \frac{2}{256}, \frac{3}{256}, \dots, 1 - \frac{1}{256}\} \quad (4)$$

Para calcular el umbral se traza múltiples rectas perpendiculares a la recta $\overline{P_1P_2}$, su vector dirección es perpendicular a \vec{v} (ecuación 6).

Definición de vectores ortogonales:

$$\vec{v} = (a, b) \quad (5)$$

$$\vec{v}_{perpendicular} = (-b, a) \quad (6)$$

Paso 3. Se traza la recta perpendicular a $\vec{l}(t)$ utilizando como dirección a $\vec{v}_{perpendicular}$. Esta recta es perpendicular a $\overline{P_1P_2}$ y tiene como puntos de paso a "X".

Donde:
$$X = \{0, \frac{1}{256}, \frac{2}{256}, \frac{3}{256}, \dots, 1 - \frac{1}{256}\}$$
 (7)

$$\vec{l}(t)_{perpendicular} = \vec{X} + t \cdot \vec{v}_{perpendicular}, \forall t \in T$$
 (8)

Paso 4. Se calcula la intersección de $\vec{l}(t)_{perpendicular}$ con el histograma y $\vec{l}(t)$ para cada valor de "X". La recta $\vec{l}(t)_{perpendicular}$ es perpendicular a la recta $\overline{P_1P_2}$. Se calcula la distancia entre las intersecciones señaladas.

Paso 5. La máxima distancia que se encuentre será $\overline{X_1X_2}$ (Figura 7). El punto "X₂" es el que brinda el umbral para la segmentación ya que está compuesto por el par ordenado de intensidad de gris del umbral y su frecuencia.

La mamografía se ha segmentado utilizando el procedimiento (Figura 8a). El resultado de la segmentación unimodal es todo el tejido de la mama (Figura 8b).

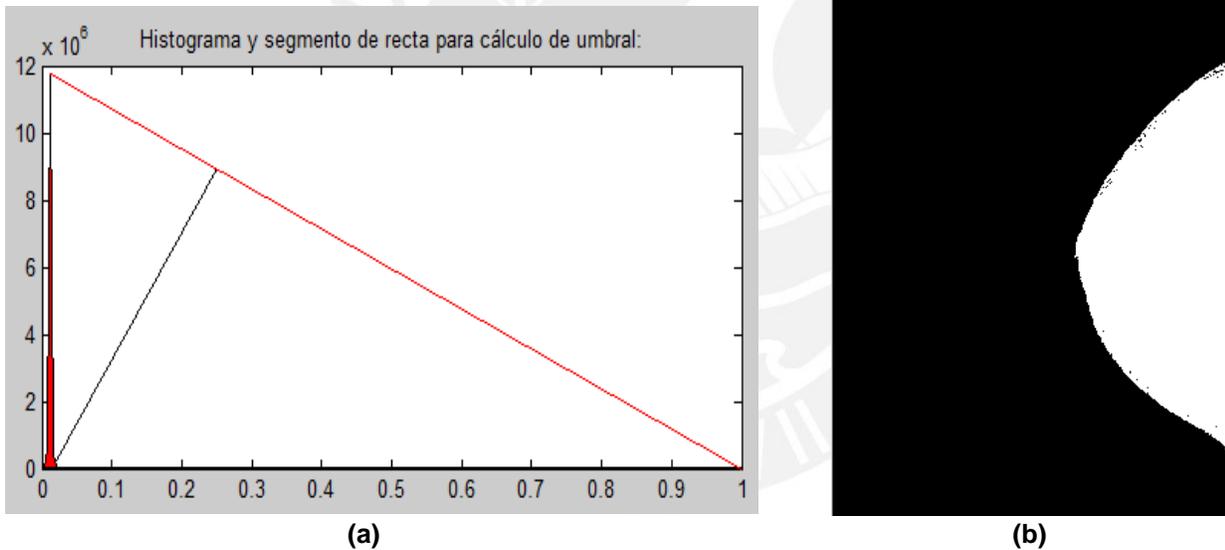


Figura 8. (a) procedimiento realizado para la obtención del umbral [20]. (b) Imagen Segmentada por método unimodal [20] de figura (8).

2.2.2. C-medias difusas

El algoritmo C-medias difusas fue planteado por Dunn en 1973 [21] y luego mejorado por Bezdek [22]. Está basado en la agrupación de los datos en dos o más clases, también conocidos como "clusters". La separación de los datos representa estructuras de tejido adiposo y denso. El algoritmo ha sufrido mejoras entre las que se encuentran las siguientes: optimización por técnica combinatoria, técnica genética basada en computación evolutiva.

a) Método de cálculo para C-medias difusas [23]

Dada la imagen "X" con una distribución de píxeles definidos por la función:

$$X(i, k) \quad (9)$$

Donde : X : mamografía

(i, k): posición de los píxeles y representa la coordenada en el formato (fila, columna) de una matriz

Dada las funciones:

$d(i, k)$: representa la distancia entre píxeles individuales o norma.

$v(i)$: centro de cada grupo

$u(i, k)$: función de familiaridad para cada clase

Datos Iniciales: Se cuenta con los valores para $v(i)$

$$d(i, k) = || X(i, k) - v(i) || \quad (10)$$

En la siguiente ecuación se utiliza la constante "m" conocida como índice difuso y se le asigna el valor de 2 [23].

$$u(i, k) = \frac{1}{\sum_{j=1}^c \left(\frac{d(ik)}{d(jk)}\right)^{\frac{2}{m-1}}} \quad \forall k, i \quad (11)$$

Fórmula 2. Función de familiaridad $u(i, k)$ [23].

$$v(i) = \frac{\sum_{k=1}^n (u(ik))^m x(i, k)}{\sum_{k=1}^n (u(ik))^m} \quad \forall i \quad (12)$$

Fórmula 3. Función que calcula al pixel con intensidad $v(i)$ [23]. Este es centro de uno de las "n" clases o clusters.

a) Procedimiento de C-medias difusas

Paso 1. Adquirir data de la mamografía de (9).

Paso 2. Selecciones valores iniciales para $v(i)$

Paso 3. Aplicar (10) y luego (11).

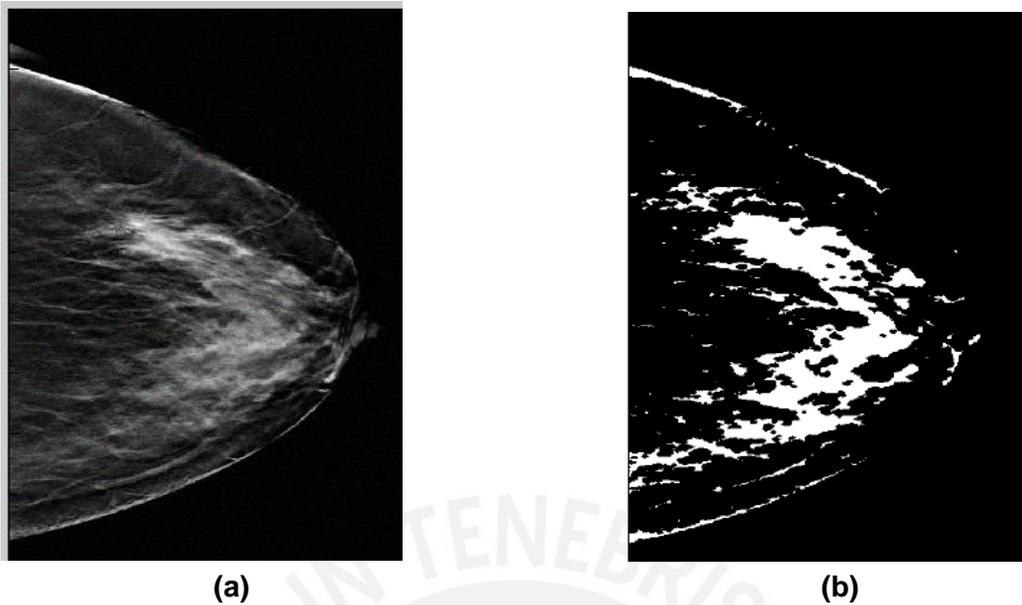
Paso 4. Aplicar (10) para actualizar centros de $v(i)$.

Paso 5. Repetir el proceso hasta que se cumpla lo siguiente:

$$|v_{ik}(t+1) - v_{ik}(t)| < e \quad (13)$$

Donde: "e": error en el orden de las centésimas

"t": número de iteraciones o repeticiones



(a) (b)
Figura 9. Clasificación de mamografía. (a) Mamografía original
(b) Mamografía clasificada en dos niveles. [24]

2.2.3. Segmentación Bimodal

La “técnica bimodal” se utiliza en histogramas con distribuciones bimodales. Un histograma con una distribución bimodal es la suma de dos histogramas unimodales. En la figura 10 se observa a una distribución bimodal que tiene la forma de dos lóbulos convexos. Se señalan las áreas en la Figura 10 que representan a las funciones P1 y P2 en la ecuación 15 y 16.

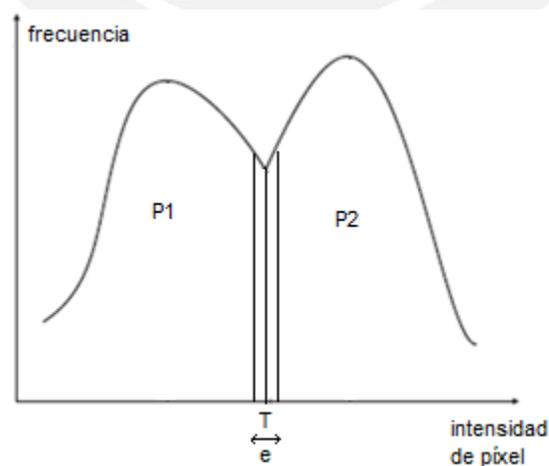


Figura 10. Histograma con distribución bimodal adaptado de [26].

a) Procedimiento de segmentación bimodal [25]:

Dada la imagen "X" con una distribución de píxeles definidos por la función:

$$X(j, k) \quad (14)$$

Donde: X : mamografía

(j, k): posición de los píxeles y representa la coordenada (fila, columna) de una mamografía

Dadas las funciones:

T: umbral inicial para segmentación bimodal

Se definen las funciones de pesos P1 y P2 de la siguiente manera:

$$P1 = \frac{\sum_{j=1}^J \sum_{k=1}^K X(j,k) * (X(j,k) < T)}{\sum (X(j,k) < T)} \quad (15)$$

$$P2 = \frac{\sum_{j=1}^J \sum_{k=1}^K X(j,k) * (X(j,k) \geq T)}{\sum (X(j,k) \geq T)} \quad (16)$$

Se calcula el error "e" y umbral "T" dinámico

$$e = T - (P1 + P2) / 2 \quad (17)$$

$$T = (P1 + P2) / 2 \quad (18)$$

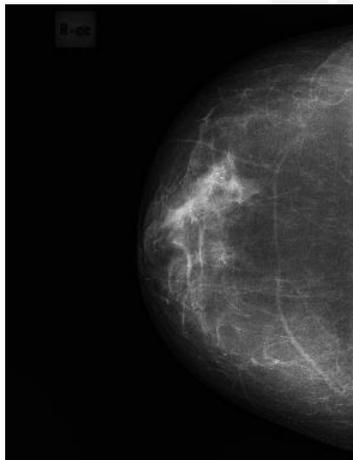
Procedimiento:

Paso 1. Se adquieren los datos de la mamografía (14)

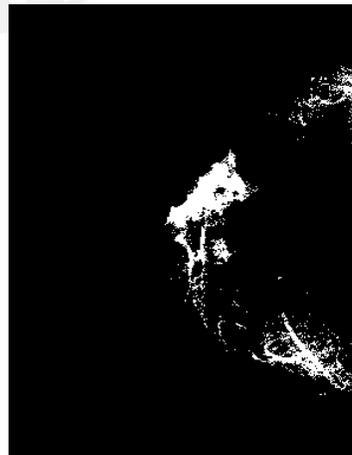
Paso 2. Se utiliza X(j, k) para calcular (15) y (16)

Paso 3. Se calcula (17) y (18)

Paso 4. Se verifica que se cumpla la siguiente condición: |e| < T. De no ser así, se repite desde el Paso 2. En caso se cumpla utilizar "T" para segmentar la mamografía.



(a)



(b)

Figura 11. (a) Mamografía Original. (b) Segmentación Bimodal de mamografía. [10]

2.2.4. Segmentación de Otsu

La segmentación de Otsu puede ser considerada equivalente a la técnica de k-medias (“K-means”) para solo 2 clases de grupos: el primer grupo son los píxeles con mayor intensidad y el segundo lo conforman los píxeles con menor intensidad.

Dada una imagen $f(x,y)$ con L niveles de intensidad (L niveles de gris) se genera una imagen $g(x,y)$ en base a la relación de los píxeles vecinos con el pixel con ubicación (x,y) de $f(x,y)$ de la ecuación 19.

$$g(x,y) = \frac{1}{2k+1} \sum_{\Delta x=-k}^k \sum_{\Delta y=-k}^k f(x + \Delta x, y + \Delta y) \quad (19)$$

Para la ecuación 19 se toma un valor de $k=1$ que resulta de tomar a 4 vecinos de cada píxel y a partir de ellos calcular el valor de $g(x,y)$. Los resultados de las funciones “ $f(x,y)$ ” y “ $g(x,y)$ ” son las intensidades de cada uno de los píxeles de la mamografía y se denominarán como “ i ” y “ j ”, respectivamente. La cantidad de ocurrencias del par (i,j) se denomina la frecuencia f_{ij} . La probabilidad conjunta de (i,j) es la frecuencia f_{ij} dividida entre el tamaño de la mamografía “ M ” de la ecuación 20.

$$P_{ij} = \frac{f_{ij}}{M} \quad (20)$$

Donde: M: tamaño de la imagen

La probabilidad acumulada es la suma de los productos de la probabilidad conjunta por la intensidad del píxel.

$$u_i = \sum_{i=1}^n P_i * i \quad (21)$$

El gráfico de la distribución de probabilidad conjunta $\{p_{ij}\}$ es denominada histograma 2D.

El valor de umbral es el par (s,t) , siendo “s” el umbral para $f(x,y)$ (intensidades de la matriz original) y “t” el umbral para $g(x,y)$ (intensidades de la matriz modificada) [27].

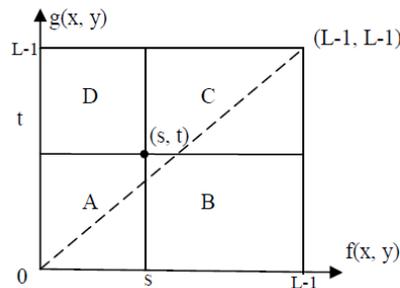


Figura 12. El par umbral (s,t) divide a la gráfica $g(x,y)$ Vs. $f(x,y)$ divide en cuatro áreas al histograma 2D [28].

Donde: **A y C:** representan área de la mama

B: representa bordes de la mama

D: representa ruido

Las áreas A y C representaran a los grupos de píxeles de mayor y menor intensidad, la probabilidad para ambos grupos es $w(k)$:

$$w_k = \sum_{i=0}^{s-1} \sum_{j=0}^{t-1} P_{ij} \quad (22)$$

Se utiliza la notación w_k para representar a los grupos A y C. w_0 para A y w_1 para C. En la ecuación 22 se muestra que Otsu utiliza la varianza entre clases como métrica para calcular el umbral.

$$\sigma^2 = \sum_{i=1}^L \frac{(P_i * u_i(\text{ultimo}) - u_i)^2 P_i}{1 - P_i} \quad (23)$$

A la intensidad de píxel "i" que maximice ese valor se le utiliza como umbral y se le denomina " τ ". La ecuación 24 describe el umbral.

$$\tau = \text{Arg}(\max(\sigma_{(i)}^2)) \quad (24)$$

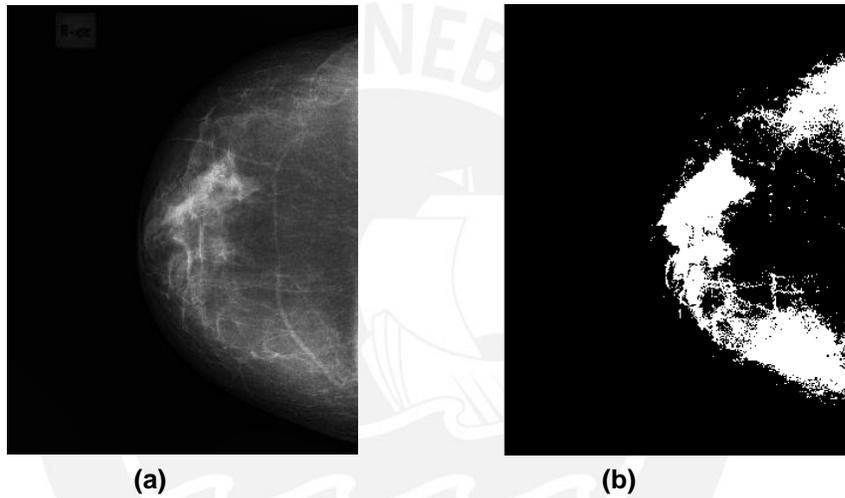


Figura 13. Umbralización de Otsu. (a) Mamografía original.
(b) Mamografía umbralizada. [10]

2.2.5. Segmentación de Bden

Es una técnica que utiliza histogramas, funciones de diferenciación, filtros y erosión y se utiliza para extraer el tejido denso de la mama. El procedimiento es explicado de manera teórica. Primero, se remueve a toda el área fuera de la mama (a todo el contorno). Luego se detectan los bordes de la mama para determinar las ubicaciones en donde tanto el tejido denso como adiposo coinciden; procedimientos adicionales tales como erosión son seguidos para segmentar la imagen tejido denso de toda área dentro de la mama que no sea [29].

La detección de bordes es un conjunto de técnicas que se implementaron desde la década de los 70 para extraer los bordes de una imagen. Macleod utilizó funciones de dos dimensiones y obtuvo bordes sin discontinuidades utilizando como base matemática la superposición de dos funciones Gaussianas [31]. Para la extracción de bordes de las mamografías se utiliza la diferenciación discreta de "Sobel".

Dada:

$$f(x, y) \tag{25}$$

Donde: “x”: filas

“y”: columnas

$$\nabla f(x, y) = \sqrt{\partial_x f(x, y)^2 + \partial_y f(x, y)^2} \tag{26}$$

La ecuación 26 se aproxima para llegar a 27

$$\nabla f(x, y) = \left| \frac{df}{dx} \right| + \left| \frac{df}{dy} \right| \tag{27}$$

La detección de bordes de Sobel es la suma de los cambios en los ejes “x” e “y” (27), es decir está compuesta por 2 máscaras. El resultado del cálculo de la máscara de Sobel se explica en las ecuaciones (28), (29) y (30).

Tabla 1. Representación de la máscara de Sobel por sus componentes

$f(x-1, y-1)$	$f(x-1, y)$	$f(x-1, y+1)$
$f(x, y-1)$	$f(x, y)$	$f(x, y+1)$
$f(x+1, y-1)$	$f(x+1, y)$	$f(x+1, y+1)$

Para los cambios en el eje x se toma la primera componente de (27):

$$\frac{df(x, y)}{dx} = f(x + 1, y) - f(x, y) \tag{28}$$

Derivando nuevamente (28):

$$\frac{df^2(x, y)}{dx^2} = f(x + 1, y) + f(x - 1, y) - 2f(x, y) \tag{29}$$

Para tener la característica de la máscara de Sobel se suman los efectos en los ejes “x” e “y”, según la ecuación (28).

$$\frac{df}{dx} = |f(x - 1, y + 1) + 2f(x, y + 1) + f(x + 1, y + 1)| - |f(x - 1, y - 1) + 2f(x, y - 1) + f(x + 1, y - 1)| \tag{30}$$

$$\frac{df}{dy} = |f(x + 1, y - 1) + 2f(x + 1, y) + f(x + 1, y + 1)| - |f(x - 1, y - 1) + 2f(x - 1, y) + f(x - 1, y + 1)| \tag{31}$$

Tablas 2 y 3. Máscaras de Sobel en el eje “x” (2) y en el “y” (3)

1	0	-1
2	0	-2
1	0	-1

(2)

-1	-2	-1
0	0	0
1	2	1

(3)

El histograma posee una distribución bimodal. Se utiliza una umbralización por mediana. Se elige a la mediana que se encuentra en medio de los picos del histograma bimodal.

a) Erosión

La erosión es una operación morfológica del procesamiento de imágenes digitales. Se utiliza el término morfológico, por ejemplo, porque este se puede referir a los bloques de píxeles blancos de una matriz binaria que describen completamente a un objeto morfológico. El proceso de erosión está definido como aquel grupo de píxeles “Z” ubicados en una posición aleatoria en la imagen de dimensiones “m x n” donde un elemento estructural que puede ser un rectángulo, se traslada a una posición tal que quede totalmente dentro de la imagen y se superponga a ese grupo de píxeles “Z”. Las zonas de la imagen que no cumplan la condición anterior se cancelan en la imagen y se pintan los píxeles de color negro. El operador para definir una operación de erosión de la mamografía “A” con un elemento estructural “B” es la siguiente:

$$A \ominus B = \{Z | B_z \subset A\} \quad (32)$$

Definición I de la operación de erosión [32].

Una segunda forma de definición se explica como aquel elemento B que no comparta ningún elemento en común con los píxeles negros y que se encuentre contenido en A. Los píxeles de color negro de la imagen A son su complemento, se definen como A^c .

$$A \ominus B = \{Z | (B)_z \cap A^c = \emptyset\} \quad (33)$$

Definición II de la operación de erosión [32].

La figura muestra un ejemplo de la erosión, el elemento A es una imagen de resonancia magnética de la mama que ha sido transformada a una imagen binaria, el elemento estructural es B es un rectángulo. El resultado de operación es C porque las zonas de la imagen en las que B se interseca con aquellos puntos que no son píxeles de color blanco se cancelan en la imagen y son ahora píxeles de color negro [40].

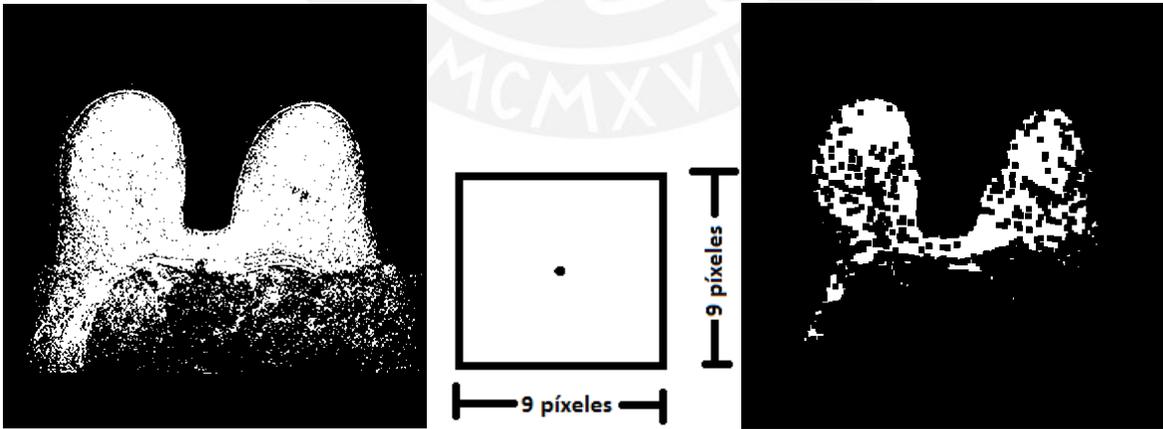


Figura 14. “A” de 640x640 píxeles [39] Figura 15. B de 9x9 Figura 16. C es la erosión de A con B

2.3. Estimadores globales para la estimación de la densidad mamaria

Los estimadores globales permiten comparar dos mamografías y brindan como resultado un valor cuantitativo. Estos permiten cuantificar los resultados que se obtienen al aplicar técnicas de segmentación en mamografías digitales. Estos métodos analíticos son formulas planteadas en función de parámetros estadísticos como desviación estándar, media, mediana, varianza, covarianza, correlación de una o más matrices; transformaciones matriciales matemáticas como inversas y transpuestas. Mientras que los métodos gráficos consisten en utilizar la geometría de la mamas, como su intersección y unión.

Más adelante en los capítulos 3 y 4 se comparan los resultados de haber aplicado técnicas o algoritmos de C-medias difusas, bimodal, Bden, Otsu y técnica propuesta de segmentación que calculan el tejido denso de mamografías digitales originales con el tejido denso calculado por médicos radiólogos expertos en las mismas mamografías. Se toma los resultados de los médicos como el “estándar de oro”. Al tejido denso calculado por los médicos se compara con el calculado por las segmentaciones implementadas en computadora. Una mamografía del estándar de oro se describe como los trazos realizados por el médico experto que indican las partes que son tejido denso y que partes no lo son, además se toma esta como referencia ya que se busca que la implementaciones computaciones sean iguales al #estándar de oro”

Se utilizan tres estimadores globales para obtener datos cuantitativos: índice de similitud estructural (SSIM por sus siglas en inglés), norma L2 (norma de una matriz), estos son métodos analíticos mientras que método de intersección sobre unión, método gráfico.

2.3.1. Índice de Similitud Estructural (SSIM)

SSIM utiliza parámetros estadísticos para relacionar dos imágenes “x” e “y”. Compara las estructuras que se encuentran dentro de toda la imagen.

$$SSIM(x, y) = \frac{4 \cdot u_x u_y \sigma_{xy}}{(u_x^2 + u_y^2 + C1)(\sigma_x^2 + \sigma_y^2 + C2)} \quad (34)$$

Fórmula 1. Descripción de la fórmula SSIM [33].

X : mamografía de referencia

Y : mamografía a comparar

u_x, u_y : promedio de x e y

σ_x, σ_y : desviación estándar x e y

σ_{xy} : correlación cruzada

Los resultados se encuentran en el rango de valores reales desde cero hasta uno, donde un valor igual a 1 significa que la imagen “X” es de igual calidad que la imagen de referencia “Y”.

2.3.2. Función Norma – L2

La segunda métrica es la norma de una matriz, “Norma L2”, es una medida de la cantidad de puntos que en común tienen dos imágenes que se comparan. La métrica Norma L2 se define según la fórmula 5.

$$|A| = \sqrt{\max(\text{eigenvalue}(A^H \cdot A))} \quad (35)$$

Fórmula 2. Descripción de la fórmula norma L2 [34].

A : mamografía original

A^H : mamografía conjugada y transpuesta

Las métricas que obtenemos a la salida son valores que inician en 0 (imágenes iguales y de la misma calidad) hasta valores mayores a 0 (píxeles con intensidades diferentes en cada una de las imágenes).

2.3.3. Método de Intersección sobre Unión

La tercera métrica consiste en superponer la mamografía de referencia y la mamografía que se busca comparar a esta para calcular la intersección de áreas comunes. De la misma manera se superponen las imágenes anteriores para calcular la unión. Como resultado de intersección y unión se obtendrán dos imágenes que se vuelven superponen para calcular el porcentaje de intersección; los resultados indican el área en común que tienen la intersección y unión de manera porcentual, estos valores tienen un rango desde cero (0% de exactitud) hasta uno (100% de exactitud). Gráficamente un resultado de cero indica que la mamografía que se compara no se encuentra de acorde con una imagen de referencia, mientras que un resultado máximo (100%) indica que la mamografía que se compara es exactamente igual a la mamografía de referencia.

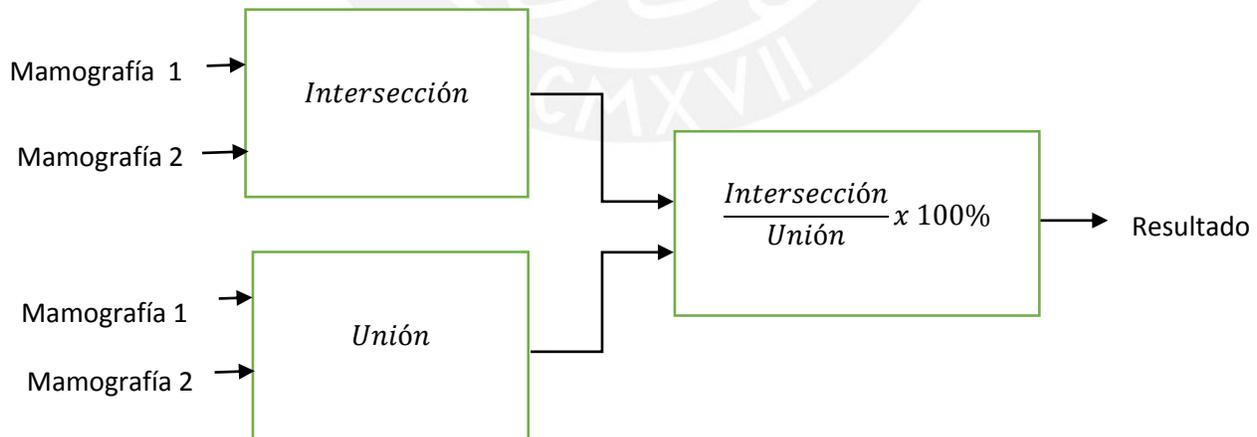


Figura 17. Diagrama de bloques para métrica “Intersección sobre Unión” [20].

Capítulo 3

Implementación de técnicas de segmentación y estimadores globales en mamografías digitales

En las siguientes implementaciones de técnicas de segmentación se tienen como data a imágenes de mamografías digitales de vista cráneo-caudal. Cada uno de los algoritmos utiliza como entrada a las mamografías digitales sin haber recibido ningún tipo de procesamiento o data cruda.

Las imágenes utilizadas están en el formato DICOM (de “Imágenes Médicas y Comunicaciones en Medicina”) utilizado extensivamente para la manipulación, almacenamiento e impresión de imágenes de la anatomía del cuerpo humano. Las imágenes han sido almacenadas en escala de grises de 16 bits y son de resolución 3560x4640 píxeles.

Notación:

$$X(i, j) \quad (36)$$

Donde: **X**: mamografía

(i, j): ubicación en el formato fila y columna

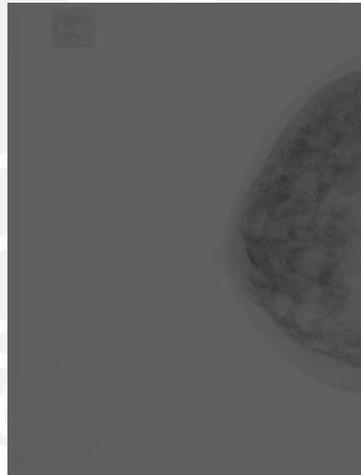


Figura 18. Mamografía sin pre-procesamiento

3.1. Objetivos

El siguiente trabajo tiene como fin implementar una herramienta (software) capaz de realizar procesamiento digital sobre mamografías para identificar el tejido denso de 100 casos clínicos obtenidos en la red de clínicas y hospitales de Oncosalud. Para alcanzar dicho objetivo, se cumplirán los siguientes objetivos específicos:

1. Implementación de técnicas de segmentación existentes para calcular el tejido denso.
2. Implementación de estimadores globales.
3. Comparación cuantitativa entre mamografías segmentadas y mamografías del estándar de oro.

4. Implementación de propuesta de algoritmo de segmentación y estimación de tejido denso de la mama.
5. Evaluación cuantitativa entre propuesta de segmentación y las demás segmentaciones respecto a mamografías del estándar de oro.

3.2. Técnicas de segmentación de tejido denso

3.2.1 Segmentación Unimodal

- a. Diagrama de bloques de segmentación unimodal

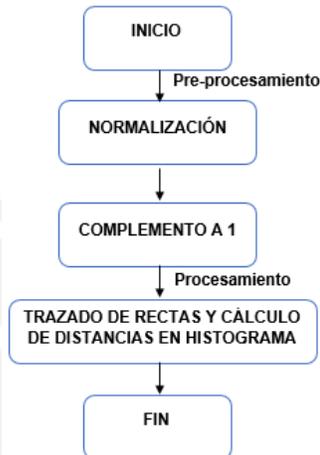


Figura 19. Diagrama de bloques segmentación unimodal

- b. Procedimiento

Paso 1. Normalización, los valores de los píxeles de la mamografía se encuentran en el rango de $[0, 2^{16} - 1]$ ya que son de 16 bits de resolución. Se aplica una transformación a cada uno de los píxeles para que ahora se encuentren en el rango de $[0, 1]$. La Figura 20 muestra a la mamografía normalizada y la Figura 21 a su histograma.

La transformación del pixel $X_{i,j}$ a $Y_{i,j}$ es la siguiente:

$$Y_{i,j} = (X_{i,j} - \text{mínimo}(X_{i,j})) / (\text{máximo}(X_{i,j}) - \text{mínimo}(X_{i,j})) \quad (37)$$

Fórmula 3. Normalización para mamografías

Paso 2. Operación de Complemento a “1”, a cada uno de los píxeles de la mamografía normalizada se le aplica la operación de complemento a 1 para que los tejidos de la mama como adiposo, glandular y piel (“foreground”) se observen de color blanco y aquellas áreas fuera del área de la mama (“background”) se observen de color negro. En la figura 22 se muestra a la mamografía en complemento a 1 que previamente se le ha normalizado y en la figura 23 su histograma.

La operación de complemento a “1” de una mamografía $X_{i,j}$ a $Y_{i,j}$ es la siguiente:

$$Y_{i,j} = 1 - X_{i,j} \quad (38)$$

Fórmula 4. Complemento a 1 para mamografías

Los procesos de “Normalización” y “Complemento a 1” se utilizan para la mayoría de las técnicas de segmentación implementadas en el capítulo 3. Se explicará convenientemente cuando se utilizan para cada segmentación.

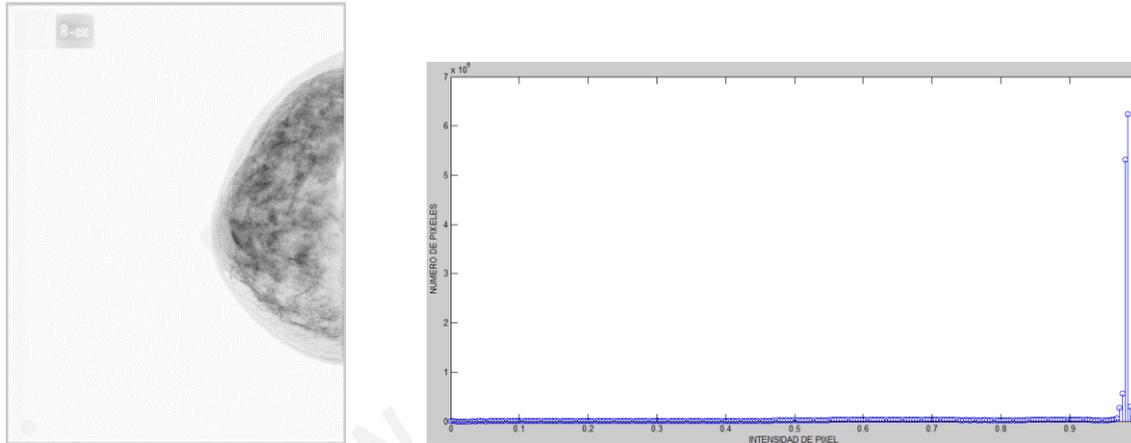


Figura 20 y 21. Mamografía Normalizada y su histograma de Figura 18

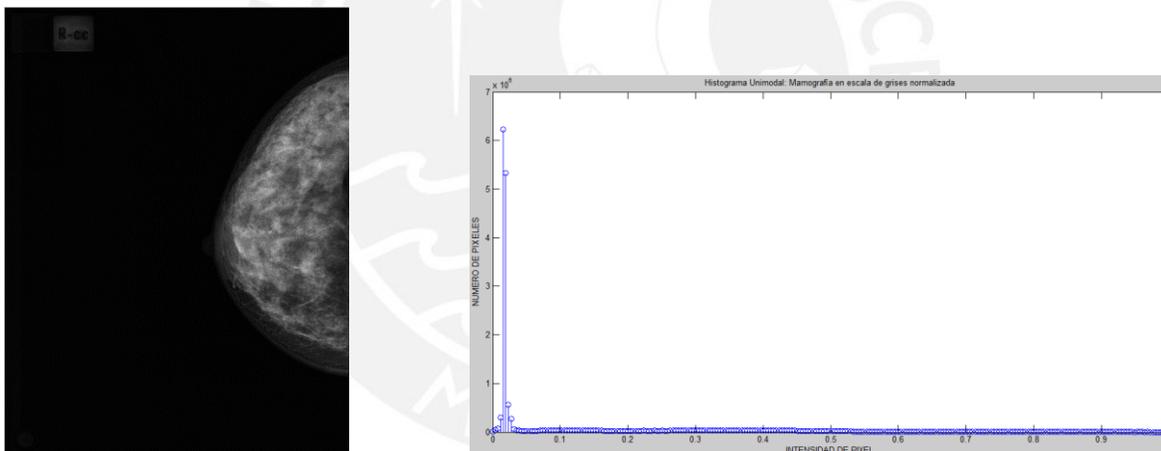


Figura 22 y 23. Mamografía en complemento a 1 y su Histograma, respectivamente.

También se conocen los puntos de paso del histograma, a partir de ellos se traza un segmento que une el máximo y mínimo (Figura 24) y se calcula la distancia de todas las rectas perpendiculares al segmento que crucen el histograma y el segmento de recta. Finalmente todas las distancias se almacenan en un vector, se escoge al mayor valor de este vector y se ubica el correspondiente par de coordenadas $P(i, j)$ (Figura 25). Se utiliza el umbral para calcular el tejido denso.

$$P(i, j) = (\text{umbral}, \text{frecuencia de umbral}) \quad (39)$$

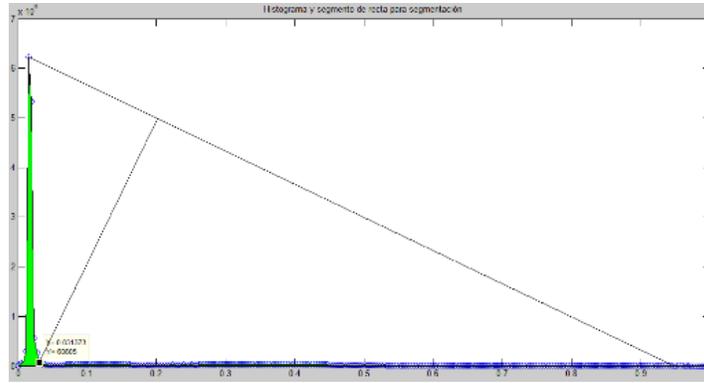


Figura 24. Procedimiento del trazado de rectas en histograma unimodal.

El resultado de la segmentación unimodal se muestra en la Figura 25.



Figura 25. Mamografía segmentada por método unimodal.

3.2.2. C-medias difusas:

Es un método estadístico que separa a la imagen en clases. Se utilizan principalmente dos funciones, estas son las funciones de “distancia entre clases” que se denomina “ v ” y “familiaridad entre píxeles” que se denomina “ u ”. El resultado proviene de la iteración entre estas 2 funciones hasta que el cambio en la función de distancia entre clases, “ v ”, sea menor a un error que se encuentra en el orden de las centésimas.

El pre-procesamiento de la mamografía, “Normalización” y “Complemento a 1”, es igual al explicado en páginas 22 y 23.

“C-medias difusas” segmenta dos áreas: tejido denso y otros tejidos a los que se agrega el contorno de la mamografía. Para cada una de estos grupos se elige un centro, alrededor del cual los píxeles con características afines a cada grupo se juntan. Los centros iniciales para los grupos 1 y 2 de una mamografía se definen por el vector de centros (40).

$$C = [C1; C2] \quad (40)$$

Donde: $C1$: centro del grupo 1

$C2$: centro del grupo 2

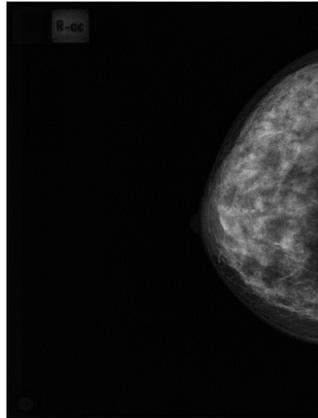


Figura 26. Mamografía en complemento a 1.

a. Diagrama de bloques de C-medias difusas

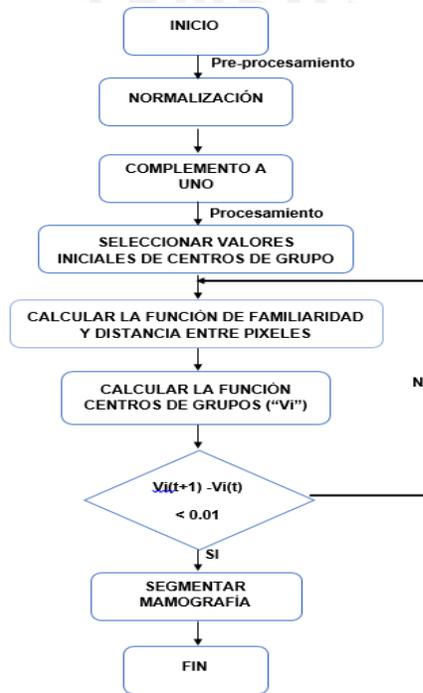


Figura 27. Diagrama de bloques de C-medias difusas

b. Procedimiento

Paso1. Valores iniciales del vector de centros:

$$C = [0.15; 0.85] \quad (41)$$

Paso 2. El centro del grupo de tejido denso tiene píxeles con intensidad 0.85, el centro del grupo del contorno de la mamografía tiene píxeles con intensidad 0.15. A continuación se define una matriz de dimensiones “4640 x 3560x2” que almacena los valores de la función de “u”. Así, “u” es equivalente a 2 matrices cada una de tamaño “4640x3560”. Además esta es la resolución de la mamografía.

Paso 3. La función distancia o “ v ” se calcula a partir de la función “ u ”. La función distancia entre centros de grupo, “ v ”, es un escalar que indica el grado de cercanía de cada uno de los grupos con sus valores anteriores ya que se actualizan en cada iteración. Ya que se ha elegido segmentar la mamografía con dos clases entonces en “ v ” se almacenan 2 distancias, cada una proviene de su respectivo grupo de píxeles.

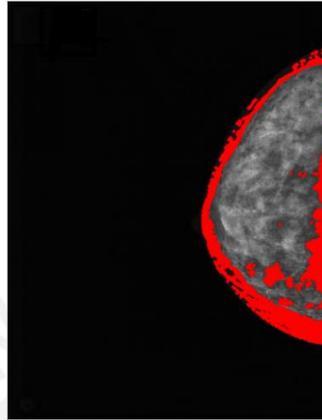


Figura 28. Puntos rojos (candidatos a centro del grupo 1) y puntos verdes (candidatos del grupo 2).

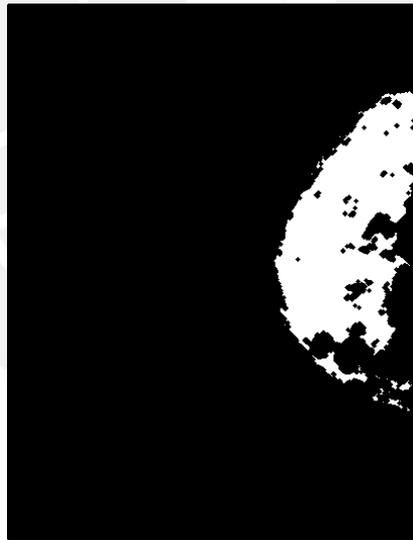


Figura 29. Imagen segmentada en una primera iteración de C-medias difusas

Paso 4. Para una primera iteración de medias difusas de C se obtuvo la actualización del valor calculado v .

En una segunda iteración de la misma mamografía o $i = 1$

$$v_{i=0} = [0.15; 0.85] \quad (42) \quad v_{i=1} = [0.8599; 0.15] \quad (43)$$

Los cambios en los valores de v para $i = 1$

$$\Delta v_1 = [0.7099, -0.7] \quad (44)$$

En una segunda iteración de la misma mamografía o $i = 2$

$$v_{i=1} = [0.8599; 0.15] \quad (45)$$

$$v_{i=2} = [0.0584; 0.5015] \quad (46)$$

La actualización de v para $i=2$

$$\Delta v_2 = [-0.802, 0.5015] \quad (47)$$

Tabla 4. Actualización de valores de los centros de grupo “ v ”

Iteración	v_i	Δv_i (error)
1	[0.15, 0.85]	-
2	[0.8599, 0.15]	[0.7099, -0.7]
3	[0.058403, 0.50151]	[-0.802, 0.5015]
4	[0.058403, 0.501159]	$[\Delta v_{i(1)} < 0.01, \Delta v_{i(2)} < 0.01]$
5	[0.058403, 0.501159]	$[\Delta v_{i(1)} < 0.01, \Delta v_{i(2)} < 0.01]$
6	[0.058403, 0.501159]	$[\Delta v_{i(1)} < 0.01, \Delta v_{i(2)} < 0.01]$

Desde la cuarta hasta la sexta iteración ambos componentes de Δv_i se encuentran en el orden de las centésimas.

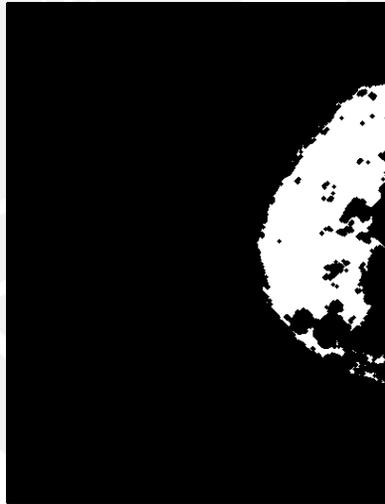


Figura 30. Segmentación final de C- Medias Difusas (iteración n° 6)

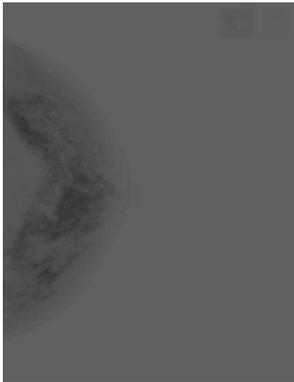
El requerimiento del algoritmo C-medias difusas es que el cambio entre los $C1$ y $C2$ sea muy pequeña, en este caso se encuentra en el orden de 10^{-6} .

3.2.3. C-medias difusas a partir de MATLAB

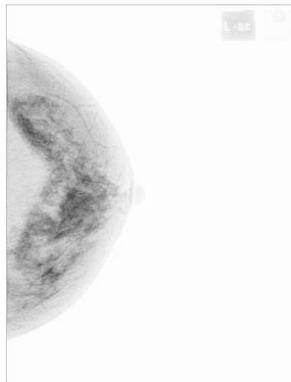
El pre-procesamiento de la mamografía, “Normalización” y “Complemento a 1”, es igual al explicado en las páginas 22 y 23.

El método de “C-medias difusas” que aplica el software MATLAB recibe como parámetros de entrada los datos de la mamografía y el número de clases que se desea tenga la mamografía segmentada, se utiliza cuatro. Utiliza las variables “familiaridad entre píxeles”

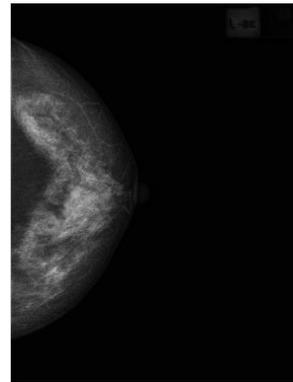
o “ u ”, “función objetiva” que se busca minimizar [23] y los “centros de grupos” o “ v ”, cuatro grupos. Además se pueden agregar hasta 4 opciones:



(31)



(32)



(33)

Figura 31, 32 y 33. Mamografía Original (a), normalizada (b) y mamografía en complemento a 1 (c), respectivamente.

El método de “C-medias difusas” que aplica el software MATLAB recibe como parámetros de entrada los datos de la mamografía y el número de clases que se desea tenga la mamografía segmentada, se utiliza cuatro. Utiliza las variables “familiaridad entre píxeles” o “ u ”, “función objetiva” que se busca minimizar [23] y los “centros de grupos” o “ v ”, cuatro grupos. Además se pueden agregar hasta 4 opciones:

Descripción de los parámetros de la función C-medias difusas de MATLAB:

$$[\text{centros}, \text{familiaridad}, \text{funcion}_{\text{objetiva}}] = \text{fcm}(\text{set}_{\text{datos}}, \text{numero}_{\text{clases}}, \text{opciones}) \quad (48)$$

a.- *Opción 1*: Es el valor que se utiliza como exponente para “ u ”. Se utiliza como dos valor predeterminado de esta opción (Ecuación 9).

b.- *Opción 2*: Es el máximo número de iteraciones que se realiza hasta alcanzar un error en el orden de las centésimas. El valor predeterminado es de 100.

c.- *Opción 3*: Es el error que se obtiene al actualizar los valores de los centros (Δv_i de Ecuaciones 32 y 34). El valor predefinido es de $1e-5$.

d.- *Opción 4*: Muestra información de cada iteración. Número de iteración y error. El valor predefinido es de 1

a. Procedimiento

Paso 1. A la mamografía que se le ha pre-procesado (normalizado y aplicado el complemento a 1). Se segmenta con el algoritmo de “segmentación unimodal”. El resultados es una matriz que solo contiene valores de “0” y “1” (matriz lógica), luego se convierte esta a una de tipo doble de valores reales para utilizarla como parámetro de la función “ fcm ”, además el número de clases se selecciona como 4.

Paso 2. A la matriz de “*u*” de tamaño 4x4040735 se le aplica la transpuesta para que ahora tenga un tamaño de 4040735x4. A continuación se ordenan los valores de los centros de los 4 grupos de manera ascendente y se guarda su número de grupo.

Paso 3. Se ordena la matriz “*u*” de acuerdo a la posición que tienen los centros de tal manera que las posiciones de los centros de grupo se ordenen con los de esta matriz. Se prosigue con el cálculo del valor máximo de la matriz “*u*” por fila, es decir, la matriz de tamaño 4040735x1 contiene los valores máximos de cada grupo a la cual se denomina “matriz etiquetada”.

Paso 4. Se calcula el valor máximo de la “matriz etiquetada” para el grupo 2 y del valor mínimo del grupo 3. Luego se promedian los valores.

Paso 5.- Se realiza la segmentación de la mamografía con el valor Paso 4. El resultado es la Figura 31.



Figura 34. Segmentación de mamografía por “C-medias difusas” de MATLAB

3.2.4 Segmentación de Otsu

Se utiliza la función del software MATLAB que emplea el método de Otsu. El algoritmo calcula un umbral global (en toda la mamografía) a partir de un método estadístico. El objetivo del algoritmo es minimizar la varianza de dos grupos de píxeles blancos y negros.

Descripción y sintaxis de la función *graythresh* de MATLAB:

$$[Nivel, métrica] = graythresh(I) \quad (49)$$

La función “*graythresh*” tiene como parámetros de entrada:

I: mamografía

Como parámetros de salida:

Nivel: El umbral global de la mamografía que convierte la imagen en binaria. Es un valor que se encuentra entre [0; 1]

Métrica: El valor de la efectividad de la umbralización entre [0; 1]. Para imágenes de entrada con pocos niveles de gris se obtiene valores cercanos a cero y cuando se tienen más niveles de gris valores cercanos a 1.

a. Diagrama de bloques de segmentación de Otsu

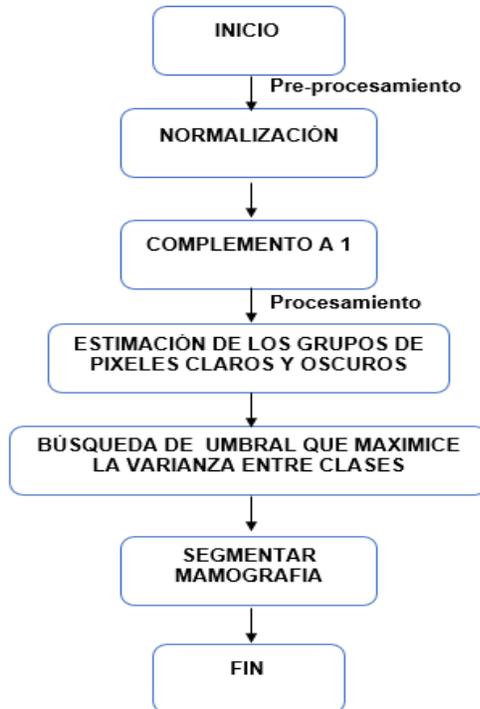


Figura 35. Diagrama de bloques de segmentación de Otsu

b. Procedimiento

Paso 1. El pre-procesamiento de la mamografía, “Normalización” y “Complemento a 1”, es igual al explicado en las páginas 22 y 23.

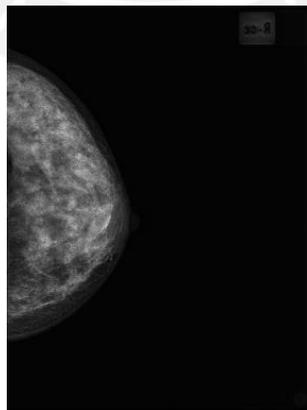


Figura 36. Mamografía normalizada y en complemento a 1.

Paso 2. Se grafica el histograma de la mamografía

$$I(n) \quad (50)$$

Donde:
I: histograma,
n: número de intensidades de gris

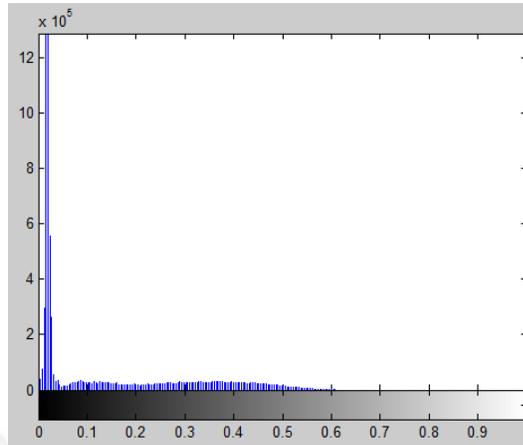


Figura 37. Histograma de mamografía normalizada

Paso 3. Se calcula la distribución de probabilidad del histograma (Ecuación 51). Se tiene una probabilidad entre 0 y 100% para cada intensidad de gris. Luego se calcula la suma acumulada (Ecuación 53).

$$P(i) = \frac{I(n=i)}{N} \tag{51}$$

Donde:
n: intensidades de gris
N: resolución de píxeles

$$w(n) = \sum_{i=1}^n P(i) \tag{52}$$

$$u_i = \sum_{i=1}^n P_i * i \tag{53}$$

Paso 4. A partir de 51, 52 y 53 se calcula la varianza entre clases (Ecuación 54). Otsu utiliza varianza entre clases como métrica para calcular un umbral.

$$\sigma^2 = \sum_{i=1}^L \frac{(P_i * u_i(\text{ultimo}) - u_i)^2 P_i}{1 - P_i} \tag{54}$$

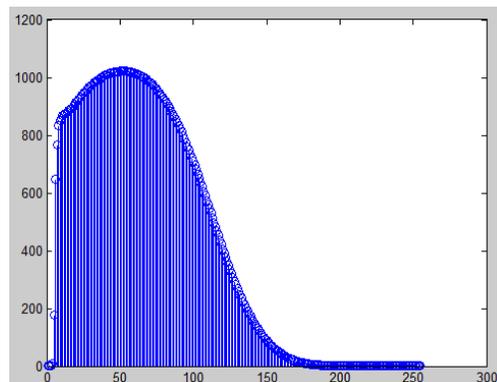


Figura 38. Varianza entre clases de Otsu " σ^2 "

Paso 5. Se calcula el umbral encontrando la intensidad de pixel tal que se maximice la varianza entre clases. El umbral es 0.1686, los píxeles con menores valores a esta intensidad se pintan de negro y los restantes de blanco.



Figura 39. Resultado segmentación de Otsu

3.2.5. Segmentación bimodal

La segmentación bimodal utiliza a la función error y función de umbral. El procedimiento inicial es el mismo que para casos anteriores.

El pre-procesamiento de la mamografía, “Normalización” y “Complemento a 1”, es igual al explicado en las páginas 22 y 23.

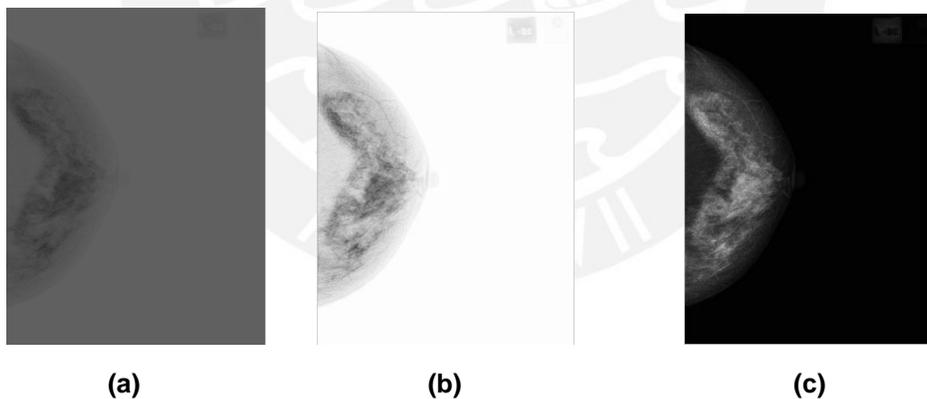


Figura 40, 41 y 42. Mamografía Original (a), normalizada (b) y en complemento a 1 (c).

a. Diagrama de bloques de segmentación Bimodal

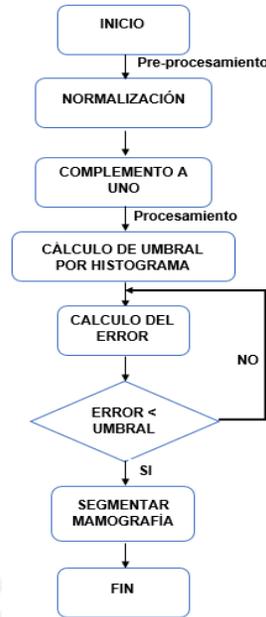


Figura 43. Diagrama de bloques de segmentación bimodal

b. Procedimiento

Paso 1. Se elige un umbral inicial de $T_0 = 0.2$ para la mamografía normalizada para que al segmentar la imagen con este valor se obtenga la mayor cantidad de píxeles de la mamografía (Figura 44).



Figura 44. Mamografía segmentada por umbral $T_0 = 0.2$

Paso 2. Se calcula las funciones de peso " P_1 " y " P_2 " utilizando el umbral inicial " T_0 ". Para " P_1 " se suman las intensidades de los píxeles menores a " T_0 " y se le divide entre el número de píxeles que cumplen esta condición. Para " P_2 " se repite el proceso para valores mayores a " T_0 ".

Paso 3. Se calcula el error " e " y se actualiza el umbral a " T_1 " (Ecuación 55 y 56).

$$e = T - (P_1 + P_2) / 2 \quad (55)$$

$$T = (P1 + P2) / 2 \quad (56)$$

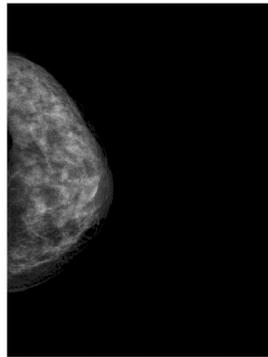


Figura 45. Intensidades de píxeles mayores a " T_0 " se suman en variable "P1"

Paso 4. Se compara el error " e " con " T_0 ". Si " e " es menor entonces se utiliza " T_0 ", si " e " es mayor o igual entonces se utiliza " T_1 ". Se repite el procedimiento hasta que el error sea menor al umbral. Finalmente se segmenta la mamografía digital.

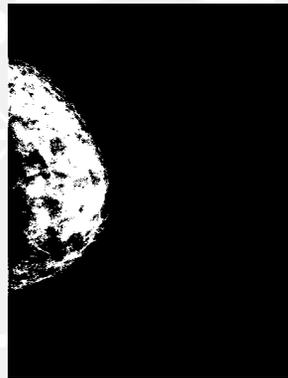


Figura 46. Resultado segmentación Bimodal

3.2.6 Segmentación de Bden

La segmentación de Bden es una técnica mixta porque utiliza métodos gráficos como histogramas para calcular un umbral y luego operaciones matemáticas como erosión y filtrado.

El pre-procesamiento de la mamografía utilizado es el explicado en las páginas 22 y 23.

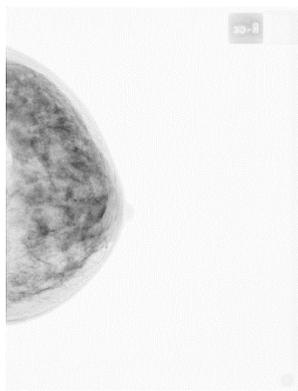


Figura 47. Mamografía normalizada de segmentación de Bden

a. Diagrama de bloques de la segmentación de Bden

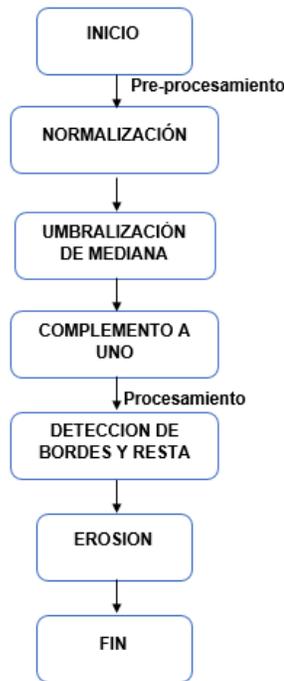


Figura 48. Diagrama de bloques de segmentación de Bden

b. Procedimiento de segmentación de Bden

Paso 1. Se traza el histograma de la mamografía normalizada, para esto se toma en consideración a sólo los píxeles de la mama, obviándose también el recuadro que lleva en la esquina derecha o izquierda dependiendo de la imagen. La razón por la que se utiliza a la mamografía normalizada es para que la distribución de los píxeles comience a crecer para valores de intensidad de pixel mayor a 0 (0.3 para este ejemplo), es decir que la función de distribución se encuentra en la parte final de la imagen.

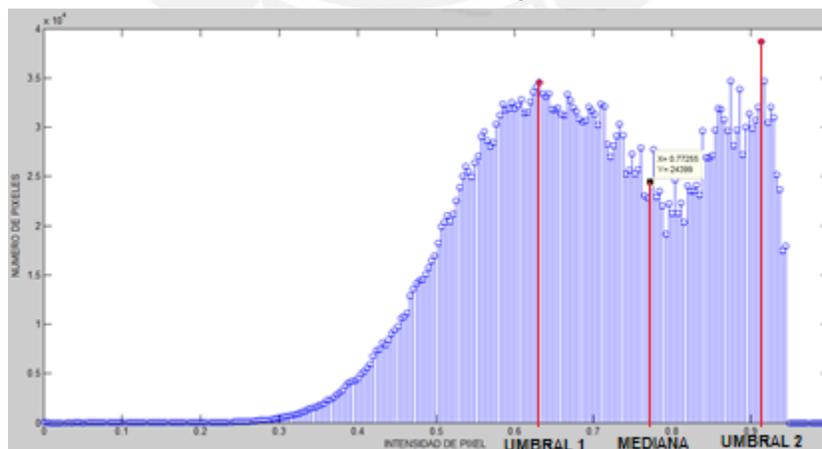


Figura 49. Búsqueda de umbral de mediana en histograma

El paso siguiente es seleccionar los 2 puntos más altos del histograma anterior (cada 1 pertenece a una figura convexa) y a partir de ellos encontrar las coordenadas (intensidad

de píxel, número de píxeles) que será la mediana entre los 2 puntos marcados de color rojo. Para verificar que es la mediana se considera que debe haber igual número de pares (intensidad, cantidad) tanto a la izquierda como a la derecha de la mediana. La mediana se utiliza para umbralizar la mamografía

Paso 2. Se umbraliza la imagen con la intensidad del píxel de la mediana y luego se aplica el complemento en 1 para que la mamografía tenga el mismo formato que todas las mamografías que se procesó.



(a) (b)
Imagen 50 y 51. Mamografía segmentada con el umbral de mediana (a) y su complemento en 1 (b).

Paso 3. El tercer paso es aplicar un filtro de media en la mamografía segmentada, esto para hacer que se noten menos los detalles de la imagen, luego se utilizó la máscara de Sobel para volver a filtrar la imagen que se había promediado.

El filtro de media que se utilizó fue de tamaño 5x5 y su peso es de 0.04; para el filtrado de Sobel que calcula los bordes y límites de la mamografía se utilizó el principio de gradiente (página 16).

Paso 4. Al procedimiento le siguió la resta de la imagen promediada (luego de aplicar el filtro de media) con los bordes y límites que se calcularon con la máscara de Sobel. Finalmente, se aplica la operación morfológica de la erosión un máximo de 10 veces hasta que asociaciones pequeñas de píxeles se descarten de la mamografía. El resultado es la segmentación de Bden final.

Durante el procedimiento de la segmentación de Bden se han utilizado técnicas de histogramas como el de umbralización por mediana y segmentación bimodal que utilizan como base a la intensidad del píxel. Luego se aplicaron técnicas que toman en cuenta la posición espacial de los píxeles tales como la detección de bordes de “Sobel” y la erosión. Esto ha permitido calcular la versión final de la segmentación de Bden.

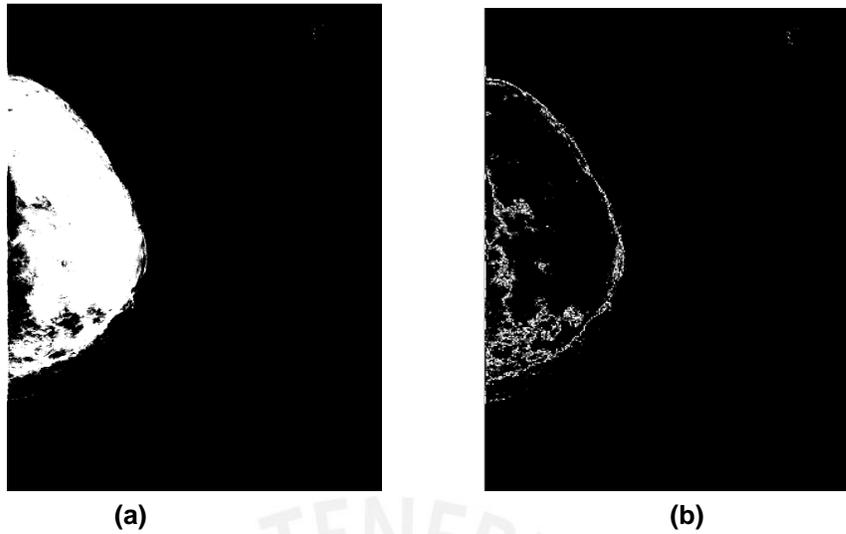


Figura 52 y 53. Filtro de media aplicado a mamografía (a). Cálculo de bordes de Sobel (b).

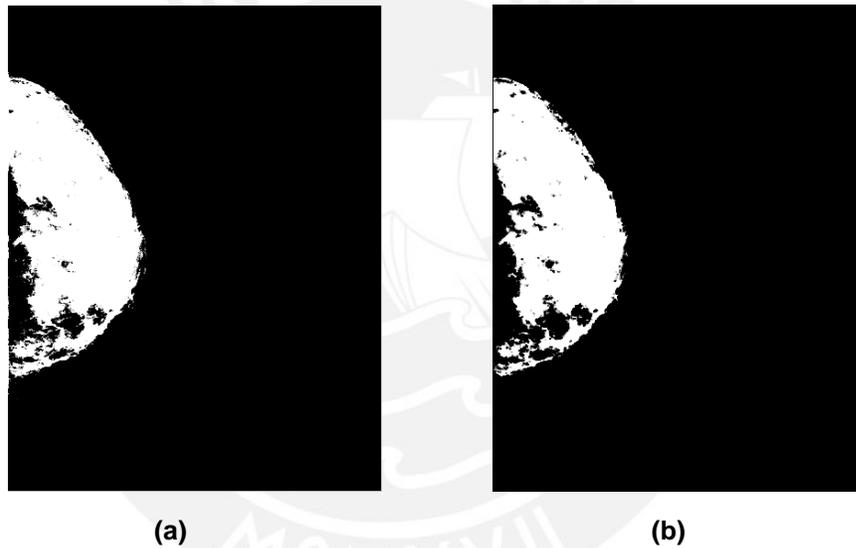


Figura 54 y 55. Resta de mamografía umbralizada con los bordes (a). Primera erosión de la mamografía proveniente de la resta anterior.

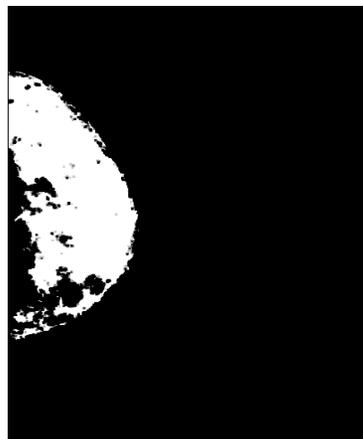


Figura 56. Erosión final de mamografía restada, Resultado de segmentación de Bden .

3.2.7. Propuesta de técnica de segmentación de tejido denso

A partir de la técnicas implementadas se ha visto conveniente utilizar en la propuesta de segmentación de tejido a la técnica de segmentación bimodal porque calcula a las áreas de tejido denso de manera equilibrada distribuyendo los pesos de los píxeles que son más oscuro de los más claros de acuerdo a su intensidad . De la segmentación se Bden se reutiliza la máscara de Sobel para la detección de bordes y se realiza una modificación en la utilización del elemento estructural para la erosión.

a. Diagrama de bloques de la propuesta de segmentación de tejido denso

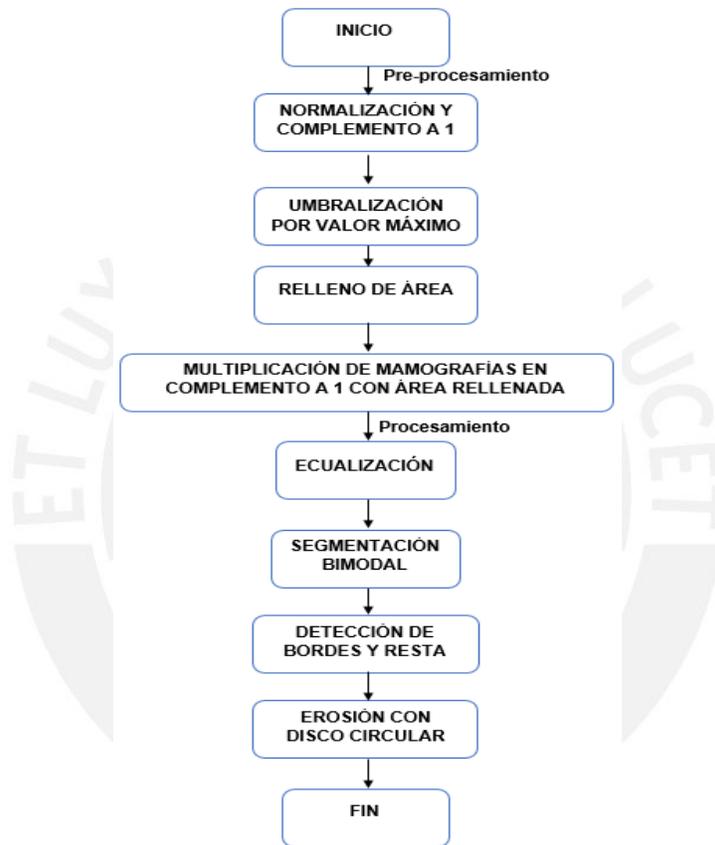


Figura 57. Diagrama de bloques de propuesta de segmentación de tejido denso

b. Procedimiento

Paso 1. Se comienza con una normalización y operación de complemento a 1 de la mamografía original (Figura 18).

El área interna de la mama está compuesta por distintos niveles de intensidad del gris, que se utilizan en el procedimiento de la técnicas ; sin embargo, también fuera de la área de la mama se tienen también píxeles con distintos valores de gris que perjudicarían el procedimiento en caso se mantengan.

Se calcula toda el área interna y bordes de la mama como parte del pre-procesamiento de la propuesta de técnica de segmentación.

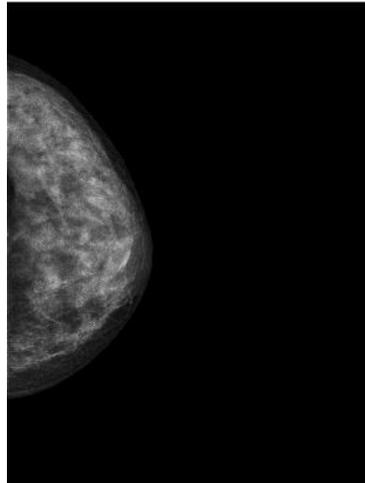
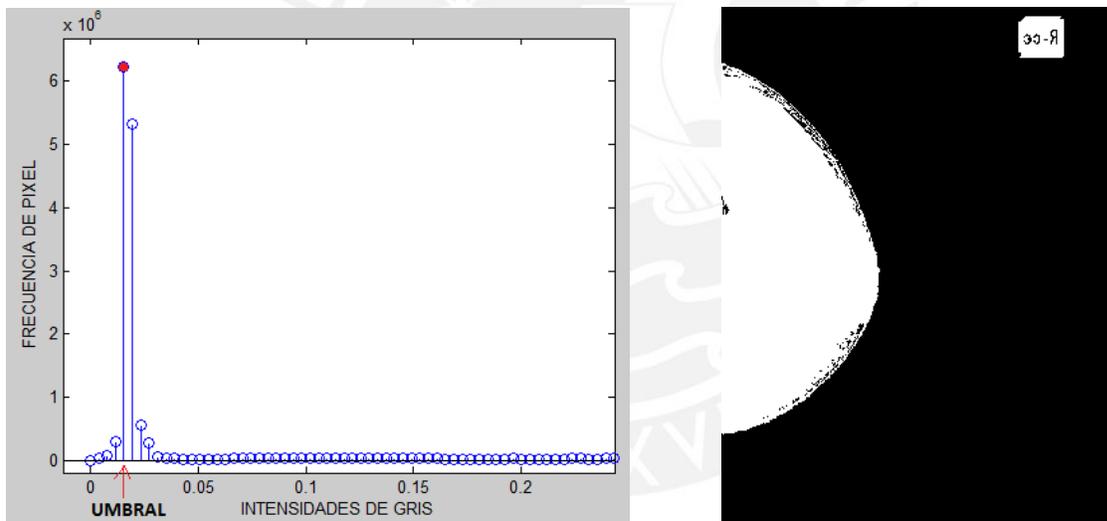


Figura 58. Mamografía normalizada y en complemento a 1

I. Pre-procesamiento

Paso 2. Se umbraliza la mamografía por ajuste del histograma con la intensidad de píxel de mayor frecuencia en su histograma. El histograma de la mamografía es unimodal.



Figuras 59 y 60. Histograma de Figura 58 y mamografía umbralizada

Paso 3. Se etiquetan a cada uno de los píxeles de la imagen de acuerdo a si estos se encuentran conectados, así se tienen 2 áreas. Una mayor en la cual está la mama y otra menor donde está la marca de agua. Se selecciona el área mayor ya que la menor no es de interés y luego se cierra la imagen completando los píxeles internos de color blanco tomando como límites los bordes (esto lo hace la función “bwconvhull” de MATLAB).

Paso 4. Se multiplica elemento a elemento la figura 58 y 62 para que dentro de la mama se mantengan todos los valores de intensidad de gris de los píxeles y fuera de ellos todo sea cero.

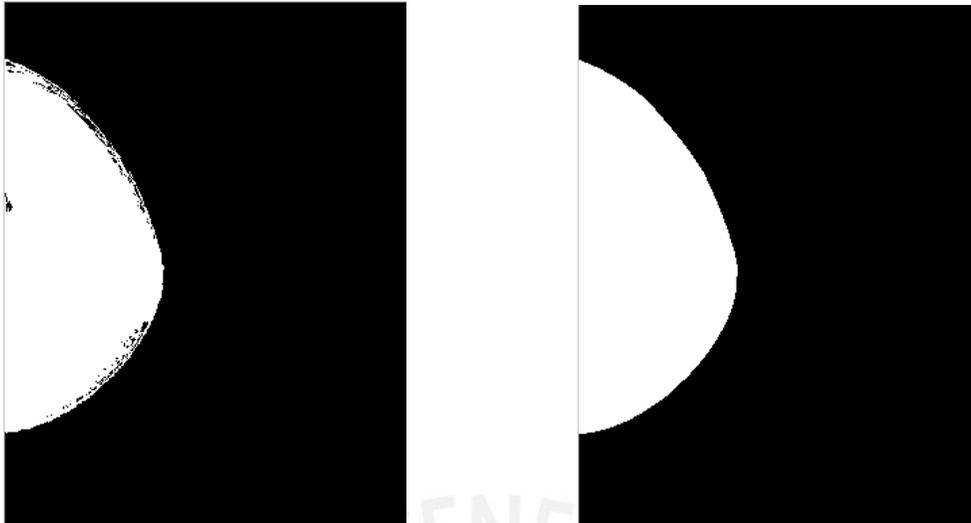


Figura 61 y 62. Mamografía sin la marca de agua y área de la mama rellenada completamente, respectivamente.

II. Procesamiento

Paso 5. Se ecualiza la mamografía para usar todo el rango de valores de la imagen de 16 bits de resolución, esto permite que se puedan diferenciar mejor los cambios de acuerdo al tipo de tejido en la mamografía.



Figura 63 y 64. Mamografía pre-procesada y luego ecualizada, respectivamente.

Paso 6. Se utiliza el método de segmentación bimodal en la mamografía de la figura 46. El valor inicial del umbral es el promedio de los umbrales 1 y 2 que se muestran en la figura 49 (los valores de los umbrales se calculan con los píxeles de mayor frecuencia que forman los picos). El histograma de la mamografía y su segmentación se muestran en la figura 59 y 60.

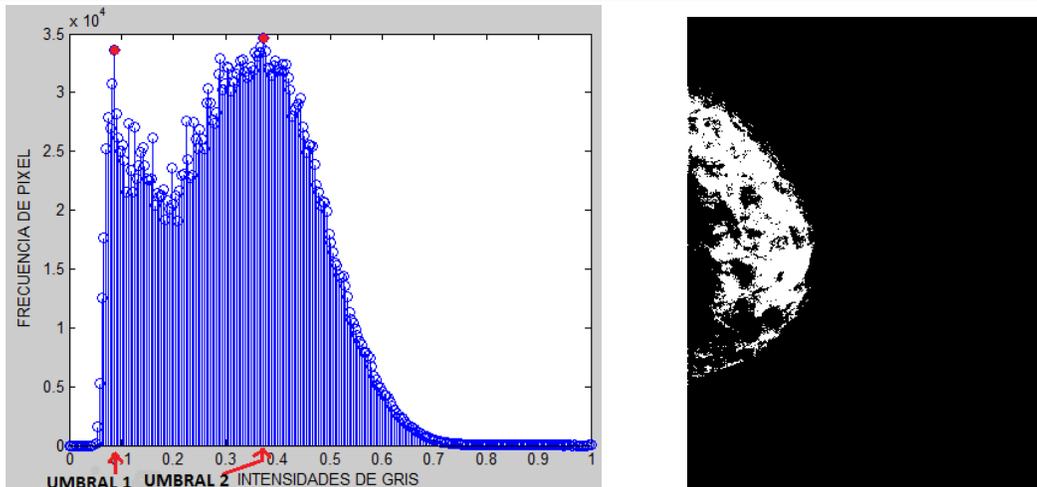


Figura 65, 66. Umbralización de mediana por histograma y mamografía segmentada por método bimodal, respectivamente.

Paso 7. Se detectan los bordes en la Figura 67 con la máscara de Sobel, estos bordes representan en su mayoría puntos aislados que no forman parte de grupos más grandes de píxeles. Son rezagos de la piel en la mama o tejido que no es denso. Se prosiguió con la resta de la mamografía segmentada con los bordes de la misma, entre las Figuras 66 y 67.



Figuras 67 y 68. Bordes de la mama y resta de los bordes, respectivamente.

Paso 8. Se erosiona la imagen para que los elementos como puntos y agujeros dentro del tejido denso se cancelen resulta en que los píxeles se pinten de negro. Se utiliza como elemento estructural a un círculo de radio de 3 píxeles para que las áreas que se consuman tengan de un trayectoria circular y así se mantengan las curvas en los borde del tejido denso.

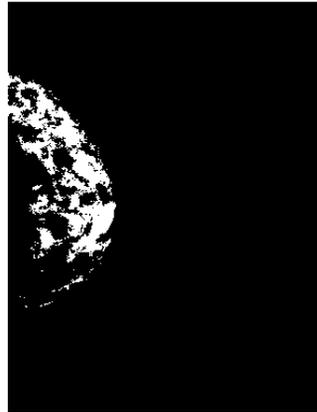


Figura 69. Resultado de la propuesta de técnica de segmentación de tejido denso

3.3. Implementación de estimadores globales

3.3.1. Intersección sobre unión (I/U)

Esta métrica fue elaborada para identificar el grado de intersección entre 2 imágenes. Se reciben como entrada a dos imágenes.



Figura 70 y 71. Imagen "A" e Imagen "B", respectivamente.

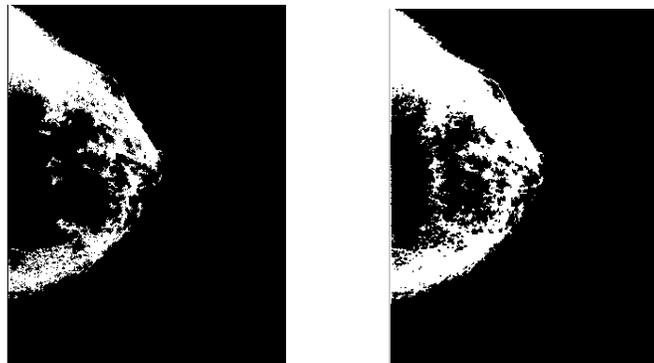


Figura 72 y 73. Intersección entre A y B, y Unión entre A y B, respectivamente

$$I/U = \frac{\text{área del tejido intersección}}{\text{área del tejido unión}} \quad (57)$$

Fórmula 5. Intersección sobre unión

3.3.2. Índice de similitud estructural

La implementación del algoritmo fue desarrollado por su autor, Z. Wang en “Calidad de Adquisición de las Imágenes” [39]. La función de similitud estructural utiliza valores estadísticos para comparar dos imágenes.

3.3.3. Norma L2

El algoritmo utilizado fue la función “norm” del software MATLAB 2013, no se implementó de manera propia. La norma de una matriz se calcula restando la imagen de referencia, en este caso, la mamografía del médico con la mamografía procesada por computadora.

Tabla 5. Descripción de estimadores globales

Estimador Global	Rango de valores posibles	Descripción
1. Intersección/unión	[0; 100] %	Valores cercanos al 100% significan que las imágenes son muy parecidas
2. SSIM	[-1; 1]	Valores cercanos a 1 significan que la calidad de la imagen la hace muy parecida a la original
3. NORMA	[0; ∞[Valores cercanos a 0 indican que las imágenes son iguales

4. Pruebas y Resultados

En este capítulo se comparan los resultados de las segmentaciones de fuzzy c-means, bimodal, Bden, Otsu y la técnica propuesta respecto a la segmentación del experto. Las comparaciones son cuantitativas y utilizan estimadores globales para comparar todos los píxeles de cada una de las imágenes segmentadas. Las métricas utilizadas para estas comparaciones son (i) intersección sobre unión, (ii) SSIM y (iii) norma L2. El objetivo final de la tesis es el cálculo del porcentaje de la densidad mamaria (ecuación 35) y validarlos a través de estos estimadores.

Para medir la efectividad del porcentaje de tejido denso calculado para fuzzy-c means, bimodal, Bden, Otsu y la técnica propuesta se compara con el porcentaje de la segmentación del experto. Se utiliza el coeficiente de Pearson para medir el grado de correlación que existe de todas las técnicas computacionales con respecto a la de los médicos (Ecuación 35).

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (58)$$

Fórmula 6. Índice de correlación de Pearson

X_i : set de datos 1 (Mamografía del médico)

Y_i : set de datos 2 (Mamografía segmentada en software)

\bar{X} : promedio aritmético del set de datos 1

\bar{Y} : promedio aritmético del set de datos 2

Para probar los algoritmos se contó con una base de datos de 100 mamografías digitales, y las respectivas 100 segmentaciones manuales de los expertos para cada caso clínico. Todos los resultados muestran el tejido denso de la mama. El porcentaje de densidad mamaria se define según la (Ecuación 36).

$$\% \text{ DENSIDAD}_{\text{mamaria}} = \frac{\text{área}_{\text{tejido denso}}}{\text{área}_{\text{total}_{\text{mama}}}} \times 100\% \quad (59)$$

Fórmula 7. Porcentaje de tejido denso

4.1. Resultados de estimadores globales por técnica implementada

Para las segmentaciones de Bden, bimodal, Otsu, fuzzy c-means y la técnica propuesta para la tesis se selecciona a la mejor técnica en base a los resultados de los estimadores. Luego se evalúa de manera general qué técnicas son superiores a las demás bajo la elección de una métrica y se explica por qué se utiliza esta para realizar el análisis.

En la Tabla 6 se ha calculado el promedio de los resultados de procesar y comparar las 100 mamografías digitales. Estos demuestran la efectividad de la herramienta desarrollada dependiendo de la técnica de segmentación empleada.

Tabla 6. Promedios de estimadores globales

Técnicas de segmentación	Estimador global		
	I/U	SSIM	norma L2
Bden	52.08%	99.04%	466.5
C-medias difusas	43.99%	99.03%	552.6
Bimodal	45.11%	98.77%	578.1
Otsu	37.20%	98.10%	370.1
Técnica propuesta	66.82%	99.54%	404.3

Para los 100 resultados de una técnica y métrica específica en la Tabla 6 se calculó el promedio. Para el estimador de intersección sobre unión el área en común promedio entre los tejidos del tipo denso es de un 66.82% frente a valores de 52.08% de la segmentación de Bden, 43.99% de C-medias difusas, 45.11% de bimodal y 37.2% de Otsu.

SSIM es un método analítico basado en variables estadísticas. El máximo valor promedio le corresponde a la propuesta de técnica de segmentación con un 99.54% frente a valores de 99.04% de Bden, 99.03% de C-medias difusas, 98.77% de bimodal y 98.1% de Otsu. Si se compara la calidad de la imagen de la segmentación para la técnica propuesta con la del estándar de oro de los médicos ésta resulta ser de muy buena calidad ya que tiene el porcentaje más alto.

La norma L2 forma parte de los métodos analíticos. La segmentación de Otsu con una norma promedio mínima de 370.1 es superior a Bden con 466.5, Fuzzy C-means con 552.6, Bimodal con 578.1, y la técnica propuesta con 404.3. Al utilizar el estimador norma se resta del estándar de oro o segmentación manual del experto la mamografía a comparar (C-medias difusas, Bden, bimodal, Otsu y técnica propuesta) y a ese resultado se le aplica el estimador. Por consiguiente, se puede explicar la norma como la medida de la magnitud de una resta de imágenes.

4.2. Selección de la mejor técnica en base a los estimadores globales

En la Tabla 7 se ha contabilizado los 100 casos clínicos. Por ejemplo, se ha agrupado los resultados de intersección sobre unión (I/U) en 5 columnas correspondientes a la propuesta de segmentación, Bden, C-medias difusas, bimodal y Otsu. Posteriormente se seleccionó el mayor valor de I/U de cada fila y se contabilizó a qué técnica le pertenecía y se repitió para las 100 mamografías. Para SSIM el procedimiento fue el mismo; sin embargo, para la norma L2 se toma al menor valor de cada fila (Tabla 5).

Tabla 7: Conteo de pacientes con la mejor técnica en base a la métrica

Métrica/Técnica	Propuesta de segmentación	Bden	C-means	Bimodal	Otsu	Total
I/U	74	18	4	4	0	100
SSIM	64	23	11	2	0	100
Norma L2	22	17	6	9	46	100

Como resultado se armó una tabla que selecciona cual es la mejor técnica dependiendo de la métrica o estimadores global que se utilice. En la Figura 74 los resultados indican que la propuesta de técnica es mejor cuando se utiliza intersección sobre unión y SSIM. Otsu es mejor cuando se utiliza a la norma L2.

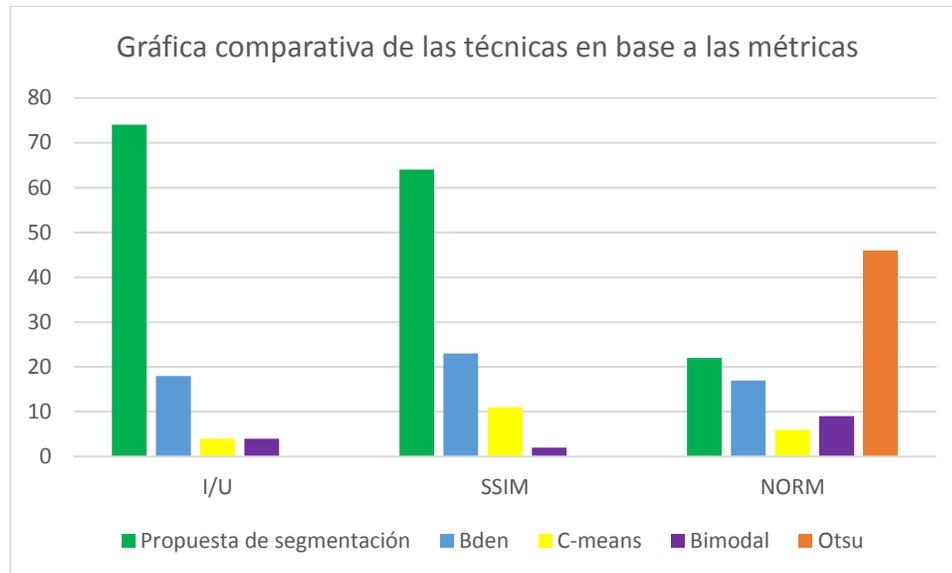


Figura 74. Muestra la distribución de los casos clínicos para cada técnica dependiendo de la métrica que se utilice

4.3. Resultados del grado de correlación entre técnicas

Ya que se conoce el performance de Bden, bimodal, Otsu, fuzzy c-means y la técnica propuesta se realiza lo siguiente: se calcula el porcentaje de tejido denso de la mama. En la gráfica 79 se muestra el porcentaje de densidad mamaria de la técnica implementada versus la del experto. Las gráficas de correlación desde la Figura 75 hasta la número 79 muestran además de la dispersión de puntos a la recta de tendencia. La recta de tendencia tiene una ecuación de pendiente y punto de paso de los cuales depende el coeficiente de correlación de Pearson que se calcula finalmente.

La figura 75 tiene a los puntos muy dispersos y alejados de la recta. La recta tiene una pendiente de 0.099, valor que está muy alejado de 1. A partir de estas observaciones se puede enunciar que el grado de correlación no es bueno ni incluso moderado porque el coeficiente de Pearson de 0.184 al analizar la segmentación de Bden versus la de los médicos.

La figura 76 tiene a los puntos más cercanos a la recta de tendencia, su pendiente es de 0.5695 que es más cercano a la unidad. A partir de estas observaciones se puede enunciar que la técnica de C medias difusas es superior a la de Bden (Figura 76) porque llega tener el grado de correlación más alto con un coeficiente de Pearson de 0.344.

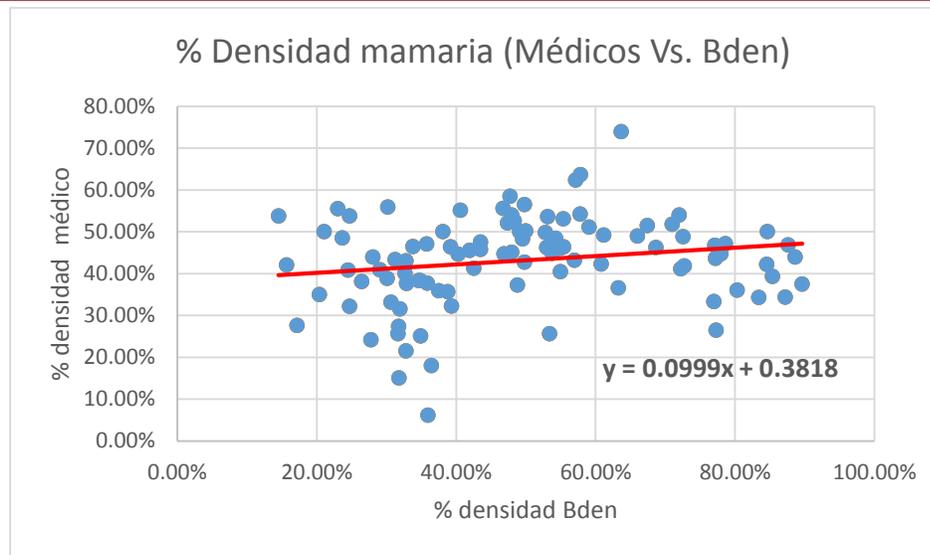


Figura 75. Porcentaje de densidad manual del médico vs. Bden

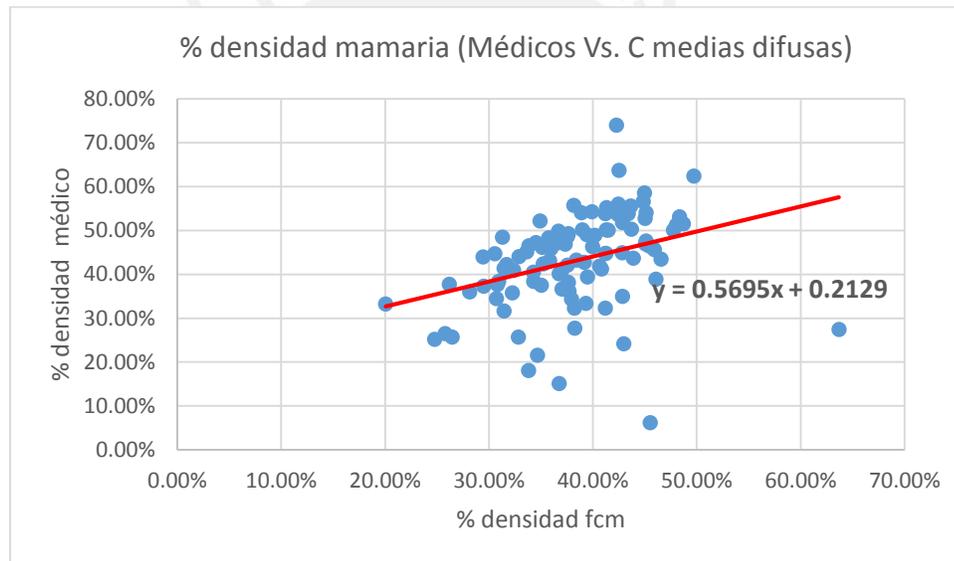


Figura 76. Porcentaje de densidad del médico vs. C-medias difusas

La Figura 77 tiene a la mayor cantidad de puntos cerca a la recta de tendencia, aunque hay valores lejanos. El valor de la pendiente de 0.0132 es muy lejano a unidad por lo que la técnica bimodal no genera buena correlación como resultado (su coeficiente de correlación es menor a 0.45) con los demás datos. Para la segmentación bimodal versus la de los médicos el coeficiente de correlación de Pearson disminuye a 0.018.

La Figura 78 tiene la densidad de la mama de Otsu versus la de los médicos. La mayoría de sus puntos agrupados se encuentran sobre los 60%. La pendiente es -0.1911, lo ideal es que se acerque a la unidad ya sea positiva o negativa. A partir de cálculos se obtiene que el grado de correlación es 0.195 (Pearson muestra una pobre correlación).

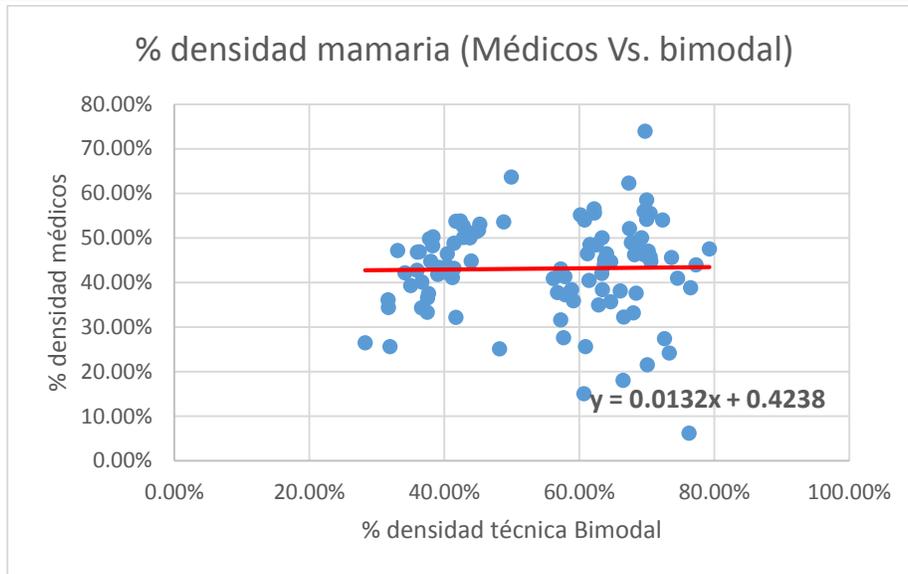


Figura 77. Porcentaje de densidad del médico vs. bimodal

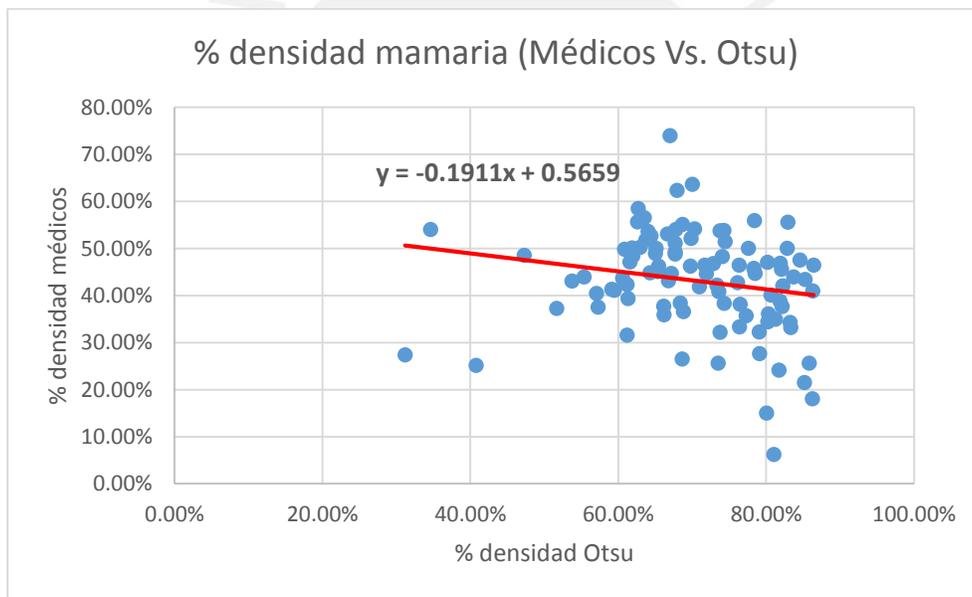


Figura 78. Porcentaje de densidad del médico vs. Otsu

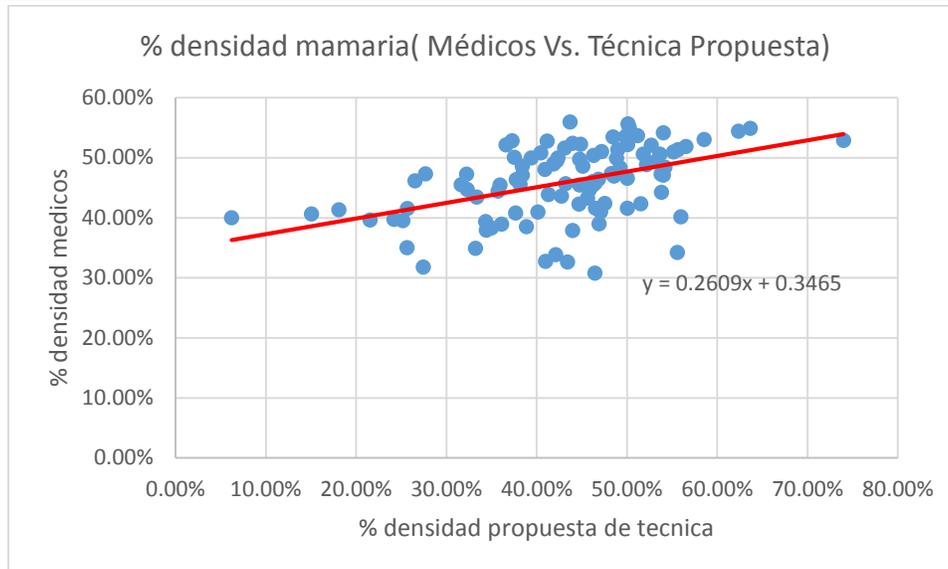


Figura 79. Porcentaje de densidad del médico vs. propuesta de técnica de segmentación de tejido denso

La figura 74 tiene a la mayor parte de sus puntos muy cerca a la recta de tendencia. La recta de tendencia tiene pendiente igual a 0.2609. Se calculó que el grado de correlación que existe entre el porcentaje de tejido denso del patrón de oro con el tejido denso calculado por la técnica propuesta llega a ser moderado y alto.

Tabla 8: Descripción del índice de Pearson para la comparación de técnicas

ENTRADAS					
Comparación	Segmentación 1	Segmentación 2	% Pearson (P)	Intervalo	Intensidad de Correlacion
1	Manual	Bden	0.184	-0.5<P<0.2	Pobre correlación
2	Manual	C-Medias difusas	0.344	0.75<P<1	Pobre correlación
3	Manual	Bimodal	0.018	P<0.25	No hay correlacion
4	Manual	Otsu	-0.195	P<0.25	Pobre correlación
5	Manual	Técnica Propuesta	0.458	0.450<P<0.75	Moderada correlación

La tabla 8 muestra los resultados luego de haber calculado el coeficiente de Pearson. El grado de correlación más alto calculado con respecto al estándar de oro de los médicos es el de la técnica propuesta para la segmentación de tejido denso con P=0.458

4.4. Discusión

La técnica de segmentación de Otsu calcula una área mayor de tejido denso a las técnicas de Bden, C-medias difusas, bimodal y la técnica propuesta desarrollada para este trabajo de tesis. Esto se debe a que Otsu calcula un porcentaje de densidad mamaria que se encuentra entre el 60 y 80% (Figura 65). Otsu tiene una pobre correlación de 0.1954 (Tabla 6). Para la Tabla 7 los resultados de Otsu resultan ser los peores para los estimadores de intersección sobre unión y SSIM (Tabla 7). Sin embargo, para el estimador de norma, Otsu es la mejor técnica.

El algoritmo de Otsu utiliza a la función de probabilidad de cada intensidad de píxel, calcula su varianza y luego utiliza la intensidad del píxel que maximice su varianza para umbralizar la imagen. Es esta cualidad de Otsu es la que produce la minimización de los valores del estimador “norma” o “norma L2”.

En la Figura 75 se ha calculado la varianza máxima para seis mamografías no procesadas. La varianza forma parte del procedimiento que se utiliza en la segmentación de Otsu. También se les ha calculado el estimador global “norma” a estas seis mamografías. Se ha establecido que un aumento en la varianza genera la disminución de la norma, mientras que la disminución de esta genera el aumento de la norma para el estudio. Este comportamiento entre la varianza y la norma puede aproximarse a una relación inversamente proporcional.

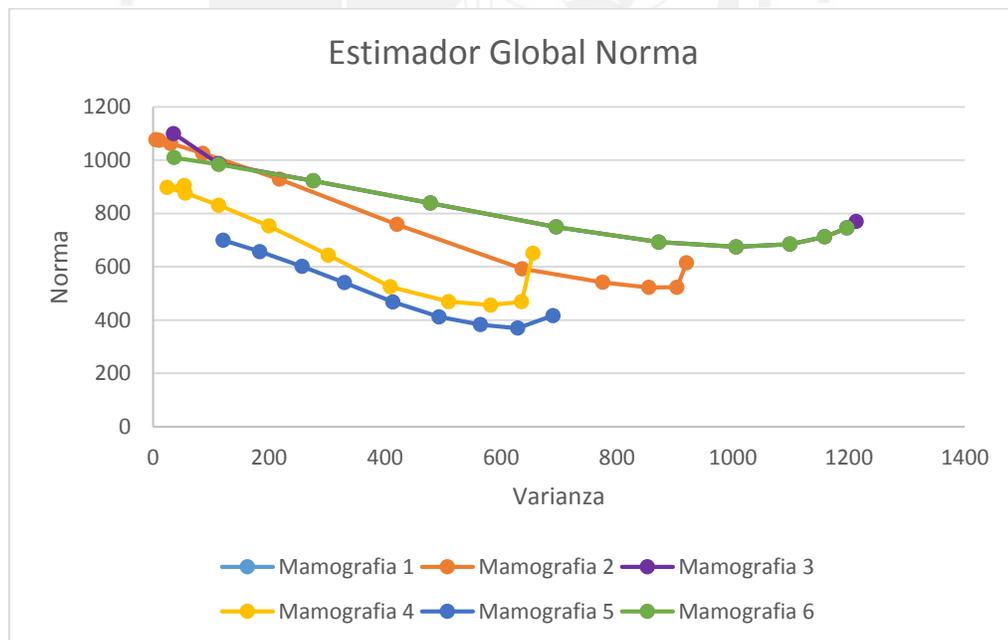


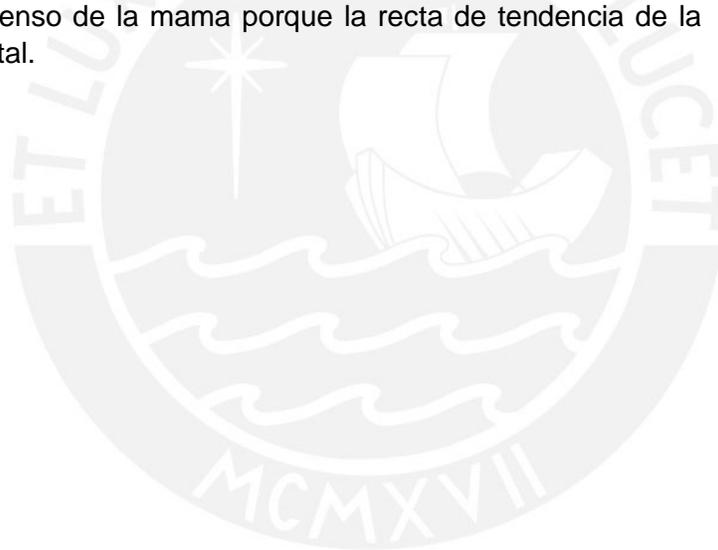
Figura 75. Estimador Global Norma versus Varianza en seis mamografías

No existe una relación proporcional entre el los estimadores de intersección sobre unión, SSIM o norma L2. Es decir si se tienen los valores “a”, “b” y “c” para las métricas intersección sobre unión, SSIM y norma L2, respectivamente, y el valor de “a” aumenta entonces no necesariamente “b” y “c” aumentan.

En cuanto a métricas, la técnica de segmentación fuzzy C-means no brinda buenos resultados porque es la cuarta técnica que se utilizaría en intersección sobre unión, tercera en SSIM y quinta en norma L2 (Tabla 7). Sin embargo, en cuanto al grado de correlación de tejido denso resulta ser la segunda mejor por encima de todas la técnicas, excepto por la técnica propuesta (Tabla 8).

La segmentación bimodal brinda mejores resultado en el cálculo de las métricas que en el de correlación de tejido denso. La métrica de intersección sobre unión es la tercera mejor después de la técnica propuesta y Bden (Tabla 6).

La segmentación de Bden es la técnica que le sigue en exactitud a la técnica propuesta ya sea en el promedio del cálculo de intersección sobre unión con un 52.08% versus un 66.82% (vence la técnica propuesta) como también en SSIM. También es la mejor técnica que se utilizaría en base a la métrica de intersección sobre unión y SSIM si la técnica propuesta no estuviera (Tabla 6). El índice de Pearson para la técnica de Bden resultó ser una pobre correlación. Si los puntos de la Figura 70 no estuvieran tan dispersos se obtendría un mucho mejor resultado comparable al de la técnica de segmentación propuesta. Bden sobreestima el tejido denso de la mama porque la recta de tendencia de la Figura 70 es casi horizontal.



CONCLUSIONES

En la presente tesis se han implementado cinco técnicas de segmentación de tejido denso de la mama en mamografías digitales de alta resolución a partir de las cuales se ha desarrollado una propuesta de segmentación propia. Las mejores técnicas de segmentación de estas 5 han sido la segmentación de Bden y fuzzy C-medias difusas. La primera debido a que tuvo los mejores resultados en estimadores globales y la segunda porque tuvo el más alto grado de correlación de tejido denso con respecto al patrón de oro. La técnica de segmentación propuesta utilizó las mejores características de Bden y C-medias difusas que permitió obtener como resultado el tejido denso de la mama más cercano al patrón de oro del médico experto.

Se han utilizado 3 estimadores globales: intersección sobre unión, SSIM y norma L2 para medir la exactitud de las 5 técnicas de segmentación implementadas y una adicional de la propuesta realizada. Los estimadores de intersección sobre unión y SSIM indicaron de manera correcta que la técnica propuesta, Bden y C-medias difusas eran superiores a las demás técnicas. En cambio el estimador norma L2 eligió a Otsu como el mejor, lo cual fue un resultado erróneo. Este comportamiento se debe a que Otsu maximiza su varianza lo cual resulta en minimizar su norma L2 (Figura 78). Intersección sobre unión y SSIM son mejores indicadores del tejido denso de la mama que norma L2.

La propuesta de segmentación de tejido denso tiene un valor promedio de intersección sobre unión de 66.82%, SSIM promedio de 99.54% y norma de 404.3 (Tabla 6). Es la mejor técnica para el cálculo de tejido denso elegida por intersección sobre unión y SSIM (Tabla 7). Finalmente, su porcentaje de tejido denso tiene un moderado grado de correlación con el de los médicos. El coeficiente de Pearson es de 0.458 lo cual lo cataloga como una moderada correlación. Se concluye que la propuesta de técnica de segmentación es superior a Bden, bimodal, fuzzy C-means y Otsu.

La técnica propuesta tiene mejores resultados cuando se utiliza en pacientes que tienen un porcentaje de tejido denso mayor al 40% porque la pendiente de la recta crece y el coeficiente de Pearson aumenta a valores mayores de 0.46. La herramienta diseñada se puede utilizar en casos que el tejido denso se haya desarrollado más en la mama debido a una mayor agresividad de la neoplasia en la mama.

RECOMENDACIONES

- La orientación de las mamografías sin procesar de la base de datos están invertidas. Comienzan en el lado derecho. Se recomienda invertir la posición de todas estas mamografías para que comiencen en el lado izquierdo.
- La segmentación unimodal es muy útil porque calcula todo el tejido de la mama a pesar de que no la segmenta. Al área que calcula también se le llama máscara de la mamografía y se utiliza para filtrar los píxeles que no se desean como aquellos que se encuentran fuera del área de la mama.
- Se recomienda utilizar la función “fcm” de MATLAB ya luego de haberse implementado el algoritmo C-medias difusas utilizando funciones básicas el primero obtuvo menores tiempos de procesamiento y mejores resultados que el segundo.
- Todas las mamografía tienen una marca en la parte superior, esta dice “CC” o “MLO” que añaden píxeles a la mamografía durante el procesamiento, es recomendable filtrar la mamografía previamente para que no aparezcan.
- Durante la tesis se trabajó con un tamaño estándar de mamografía de 4640x3560 píxeles. El procesamiento para imágenes de mayor tamaño tarda más tiempo. Por ello se recomienda reducir los tamaños para lograr que los tiempos de procesamiento sean menores.
- Para este trabajo de tesis se utilizaron los algoritmos de segmentación de imágenes básicos; sin embargo, existe una gama de técnicas adicionales que permitirían mejorar los resultados tales como “support vector machine” y redes neuronales. Se recomienda seguir el acercamiento de implementar técnicas y utilizar las mejores características de cada una de estas para diseñar una propuesta de segmentación de tejido denso que supere a la actual.
- Dentro de este trabajo de tesis se han desarrollado comparaciones de las técnicas implementadas con el patrón de oro del médico experto. Para estas comparaciones se han utilizado imágenes de la vista cráneo caudal de la mama. Se recomienda utilizar imágenes de la vista medio-lateral oblicua de la mama para las comparaciones ya que la vista cráneo-caudal oscurecen lesiones en el pecho [9].
- La segmentación bimodal es también llamada segmentación por minimización del error (del inglés “minimum error thresholding”) [26]. Las revisiones de segmentación bimodal en la IEEE se encuentran también bajo ese nombre.

BIBLIOGRAFÍA

- [1] Cancer Research UK
“Breast Cancer”, <http://www.cancerresearchuk.org/cancerinfo/cancerstats/world/breast-cancer-world/#Mortality> (Consultado el día 11 de Octubre del 2013).
- [2] Instituto Nacional de Enfermedades Neoplásicas
Neoplasias Malignas más Frecuentes Diagnosticadas en Ambos Géneros, http://www.inen.sld.pe/portal/documentos/pdf/estadistica/datos_estadisticos/22042013_NEOPLASIAS_MALIGNAS_FRECUENTES_1985.pdf (Consultado el 11 de diciembre de 2013).
- [3] Datos del Departamento del INEN
2011 <http://www.inen.sld.pe/portal/estadisticas/datos-epidemiologicos.html> (Consultado el 21 de abril de 2014).
- [4] SAFTLAS AF, Hoover RN, Brinton LA, Szklo M, Olson DR, Salane M, Wolfe JN.
1991 “Mammographic densities and risk of breast cancer,” *Cancer*, 67:2833-2838.
- [5] KOPANS, Daniel B.
2007 “Mammography – Equipment and basic Physics”. *Basic Principles of X-Ray Mammography. Breast Imaging. 3rd edition.* pp. 244.
- [6] KOPANS, Daniel B.
2007 “Mammography – Equipment and basic Physics”. *Breast Imaging. 3rd edition.* Figure 8-1, pp. 244.
- [7] KOPANS, Daniel B.
2007 “Breast Anatomy and Basic Histology, Physiology”, *Breast Development. Breast Imaging. 3rd edition.* Figure 2A – 2, pp. 10.
- [8] KOPANS, Daniel B.
2007 “Breast Anatomy and Basic Histology, Physiology”, *Supporting Structures of the Breast. Breast Imaging. 3rd edition.* Figure 2A – 18, pp. 24.
- [9] KOPANS, Daniel B.
2007 “Mammography Positioning”. *Breast Imaging. 3rd edition.* pp. 282.

- [10] Laboratorio de Imágenes Médicas
2014 Base de datos mamográfica.
- [11] American Cancer Society
2013 Breast Density and your breast mammogram Report.
<http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-039989.pdf>.
- [12] American College of Radiology,
1993 “Breast Imaging Reporting and Data System (BIRADS),” American College of Radiology, Reston, VA.
- [13] KOPANS, Daniel B.
2007 “Breast Anatomy and Basic Histology”. Adipose tissue in the Breast. Breast Imaging 3rd edition. PP. 28.
- [14] TAKAAKI Takei, Mitsuru Ikeda, Kuniharu Imai, Katsuhiko Kato, Haruo Isoda
2013 “A New Automated Method for Contrast Detail Images by Applying Support Vector Machine and its Robustness to Nonlinear Image Processing”. Australasian Physical & Engineering Sciences in Medicine vol. 16 pp. 313-322.
- [15] KOWAL Marek
2013 “Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images”. IEEE Transactions on medical imaging, Vol. 32, N. 12.pp. 2169-2178.
- [16] GUBERN, Mérida
2014 “Breast Segmentation and Density Estimation in breast MRI: A fully automatic framework”. IEEE Journal of Biomedical and Health Informatics. JBHI: 2014.2311163.
- [17] MESELHI, Mohamed
2013 “An Adaptive Threshold Method for Mass Detection in Mammographic Images. IEEE International Conference on Signal and Image Processing Applications (ICSIPA)”. PP. 374-378.
- [18] ROSIN Paul L.
2001 “Unimodal Thresholding”, Department of Information Systems and Computing. Brunel University. UK.

- [19] ROSIN Paul L.
2001 “Unimodal Thresholding”. Figure 1. Department of Information Systems and Computing, Brunel University. UK.
- [20] Elaboración propia de diagramas y tablas.
- [21] DUNN, J.C.
1973 A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. Journal of Cybernetics. Vol. 3 N° 3. PP. 32-57.
- [22] BEZDEK, J.C.
1981 Pattern Recognition with Fuzzy Objective Function Algorithms, New York, Plenum Press, New York.
- [23] KAUR Prabhjot, Anjana Gosain, I.M.S. Lambda
2011. “Kernelized Type–2 Fuzzy C-Means Clustering Algorithm in Segmentation of Noisy Medical Images”. IEEE Conference Publication. pp. 493 – 498.
- [24] Rajalakshmi Thyagarajan
“Segmentation of Digital Breast Tomograms using Clustering Tecniques”. Research Schollar, Sathyabama University. Figure 3, input image. Figure 5, Result of Fuzzy C-means Segmentation.
- [25] RODRIGUEZ Valderrama, Paul
2013 Segmentación Bimodal, Apuntes del curso “Procesamiento de Imágenes Digitales”.
- [26] PENG Tao, Cao Wei, Cao Wei, Huang Yi, Lou Yangge, Zou Ying
2013 “A Method Based on Minimum Error Thresholding. 25th Chinese Control and Decision Conference.
- [27] NINGBO Zhu,
2007 “A Fast 2D Algorithm Based on Improved Histogram”. Two Dimensional Otsu Method. School of computer and communication of Hunan University, China.
- [28] NINGBO Zhu, Gang Wang, Gaobo Yang, Weiming Dai
2007 “A Fast 2D Algorithm Based on Improved Histogram”. Figure 1: two-dimensional histogram. School of computer and communication of Hunan University, China.

- [29] MCAVOY, Steven
2010 "Thesis: Feasibility Of Determining Breast Density Using Processed Mammogram Images". I.K. Barber School of Arts & Sciences. The University of British Columbia.
- [30] ARGYLE Bernell
1971 "Techniques for Edge Detection". IEEE Proceedings, volume 59. pp. 285-286.
- [31] D.G. Macleod
1979 Comments on "A quantitative study of the orientation bias of some edge detector schemes". IEEE transactions on pattern analysis and machine intelligence, Vol. PAMI-1, N° 4.
- [32] The MathWorks
2010 Documentation Center. imerode function.
- [33] SHIQI Wang
2012 "SSIM-Motivated Rate-Distortion Optimization for Video Coding". IEEE Transactions on Circuits And Systems For Video Technology Vol. 22 N°4 pp. 516-527.
- [34] The MathWorks
2010 Documentation Center, norm function.
- [35] Wolfe JN, Saftlas AF, Salane M.
"Mammographic parenchymal patterns and quantitative evaluation of mammographic densities: a case-control study", 148:1087-1092.
- [36] Saftlas AF, Hoover RN, Brinton LA, Szklo M, Olson DR, Salane M, Wolfe JN.
1991 "Mammographic densities and risk of breast cancer," Cancer 1991, 67:2833-2838.

[37] Wolfe JN.

"Breast patterns as an index of risk of developing breast cancer" Am J Roentgenol 1976, 126:1130-1139.

[38] Wolfe JN.

1976 "Risk for breast cancer development determined by mammographic parenchymal pattern, 37:2486-2492.

[39] Z. Wang

2004 "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612.

[40] Rafael Gonzalez, Richard Woods

2008 Digital image processing. Tercera Edición. NJ: Pearson Education.

