

Estimación Semi-automática de Fracción Glandular Mamaria en Imágenes de Mastografía

M. P. R. Zamudio-Arteaga^{1*}, A. R. Mejía-Rodríguez¹,
M. P. Pérez-Badillo², H. A. Galván-Espinoza²

¹Facultad de Ciencias, Universidad Autónoma de San Luis Potosí (UASLP), San Luis Potosí, México

²Departamento de Radiodiagnóstico, Instituto Nacional de Cancerología (INCan), Ciudad de México, México.

*mel.zaart@alumnos.uaslp.edu.mx

Abstract— Breast cancer is a priority public health problem due to its global magnitude and importance, that develops mainly in the glandular tissue. On mammography imaging, the presence of a large amount of glandular tissue could conceal lesions. Due to this, the estimation of glandular fraction (FG) is a tool that allows evaluating the risk of developing breast cancer. Having knowledge of the different tissues that constitute the anatomy of the breast (glandular, connective and adipose tissues), on a mammography image there are structures that should not be considered for the estimation of the FG, such as skin or pectoral muscle. In the clinical practice, a proper differentiation between glandular and connective tissues is a challenging task, and a discrimination of extra-mammary structures from glandular tissue is particularly difficult due to an intensity similarity. In this work, a strategy to properly isolate the principal breast tissues from the extra-mammary structures, and to perform a robust semi-automatic segmentation of glandular, connective and adipose tissues by using the K-means algorithm in order to provide a quantitative estimation of the mammary glandular fraction is presented. Additionally, a comparison with the Density-based Spatial Clustering of Applications with Noise (DBSCAN) and an empirical glandular fraction estimated by a clinical expert, to demonstrate the convenience of the strategy is made.

Palabras clave— Estimación semi-automática, Fracción Glandular, Mastografía, Predictor de cáncer, Tejidos mamaros.

I. INTRODUCCIÓN

Una mastografía (o mamografía) es simplemente una imagen de rayos-X que se toma de cada mama mientras se comprime, de forma que la apariencia de una mama en la imagen depende de las diferentes características de atenuación del tejido mamario. Los componentes fibrosos y glandulares de la mama absorben el haz de rayos-X de una forma significativamente mayor en comparación con la grasa, por lo que, el tejido fibroso y glandular aparece en tonalidades más blancas y la grasa en tonalidades más oscuras [1]. Dado que la cantidad relativa entre áreas blancas y negras determina la densidad mamográfica, una mamografía que parece principalmente blanca se considera *densa* y una que es principalmente negra o gris no.

Un tejido mamario denso puede dificultar la detección de cánceres en la imagen de mastografía, además, se ha mostrado en la literatura un aumento significativo del riesgo

de cáncer de mama para las mujeres con los niveles más altos de densidad mamaria al número de células que están en riesgo de sufrir una transformación maligna [2],[3]. Por estas razones, la densidad mamográfica tiene el potencial de ser usada como predictor de riesgo [4].

En la práctica clínica, la inspección y estimación visual de la fracción glandular (FG), definida como la proporción de tejido glandular presente en la mama, por parte de los especialistas clínicos tiene un alto grado de subjetividad ya que depende en gran medida de la experiencia del médico, el equipo utilizado y el expertise de la adquisición de la imagen. Por ejemplo, en el Instituto Nacional de Cancerología (INCan), México, una de las formas de obtener la FG, como primera aproximación, es mediante una interfaz gráfica donde es posible modificar algunos parámetros de la imagen con el fin de contrastar los diferentes tejidos mamaros y facilitar la estimación de la FG. Sin embargo, no hay un registro preestablecido de estos parámetros los cuales cambian en cada imagen incluso si se trata de imágenes del mismo paciente. Por estas razones, definir una estrategia para identificar adecuadamente los diferentes tejidos constitutivos en una mamografía sigue siendo un problema abierto.

La densidad mamográfica ha sido evaluada por métodos totalmente automatizados [5], especificando la variabilidad entre observadores y la impracticidad para el cribado de la población, lo que sugiere el desarrollo de una medida semi-automatizada para aplicaciones dentro de la clínica. Se han realizado también esfuerzos para incorporar escalas visuales y umbrales basados en el histograma, matrices de co-ocurrencia, longitud de corrida, composición arquitectónica, complejidad o transformaciones multiresolución y multispectrales [6][7], pero estos enfoques requieren etapas de procesamiento más complejas o inexactas, las cuales son aplicadas a la imagen completa sin discriminar primero ciertas estructuras, que no deberían ser consideradas en la estimación de la FG.

En este trabajo se propone una estrategia que inicialmente aísla los tejidos mamaros de interés (adiposo, conectivo y glandular) de los que no lo son (piel, pezón y músculo pectoral), para posteriormente mejorar el contraste de la imagen, lo que facilita la segmentación de la mama, y así lograr una estimación del FG mamaria de forma robusta y semi-automática.

II. METODOLOGÍA

A. Población de estudio y base de datos

Las imágenes de mastografía digital fueron proporcionadas por el INCan en formato médico DICOM, correspondientes a estudios bilaterales de un grupo de 10 mujeres con diagnóstico positivo de cáncer y 10 con diagnóstico negativo. Las imágenes se adquirieron con los mamógrafos FUJIFILM AMULET Innovality y HOLOGIC Selenia Dimensions.

Las cuarenta imágenes tienen un tamaño de 5928x4728 píxeles, con un tamaño de píxel de 0.050 mm. Para normalizar automáticamente el rango de intensidad de todas las mamografías, se utilizó un paquete de procesamiento de imágenes de código abierto basado en ImageJ llamado Fiji [8], y todas las imágenes normalizadas se guardaron en formato TIFF. Además, considerando que la extracción de las máscaras binarias de las regiones extramamarias no requiere una gran cantidad de información como la que poseen las imágenes originales, estas fueron redimensionadas a una cuarta parte de su tamaño original y ecualizadas localmente para reafirmar la distribución de los 4 principales niveles de intensidad.

B. Eliminación de piel y músculo pectoral

A la imagen original se aplicó la técnica de enmascaramiento sin enfoque (Fig. 1), que consiste en aumentar el contraste de la imagen en los bordes donde se encuentran los diferentes niveles de intensidad, restando una versión borrosa (sin enfoque) de sí misma usando un filtro Gaussiano [9] y así una primera máscara binaria (Fig. 1(b)) se obtiene a partir de la imagen contrastada. La segunda máscara binaria se obtiene utilizando el operador Prewitt [10], para la detección de bordes en la imagen contrastada como podemos ver en la Fig. 1(c). De esta manera, el algoritmo genera una única máscara binaria que detecta ciertas regiones como el pezón, que son naturalmente indiferentes a las transiciones de la detección de bordes, y además se resuelven las discontinuidades generadas en la región de la piel debido al enmascaramiento sin enfoque. Esta única máscara binaria, como se muestra en la Fig. 1(d), nos permite ubicar los píxeles que serán eliminados de la imagen original, permitiéndonos obtener una imagen sin piel.

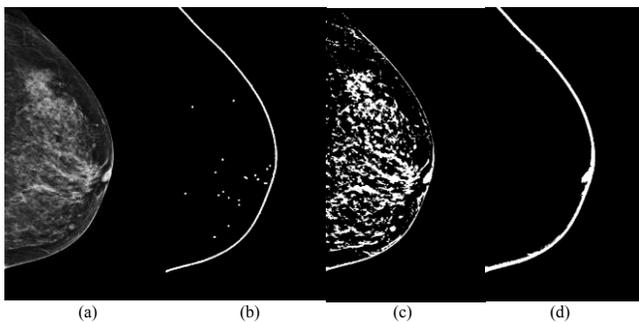


Fig. 1 Estimación de máscaras binarias para eliminación de la piel.

A diferencia de las imágenes con proyecciones oblicuas medial-laterales, en las imágenes con proyecciones craneocaudales, como las estudiadas en este trabajo, no suelen estar presentes pliegues axilares o fragmentos del músculo pectoral, por lo que se emuló el proceso que suelen realizar los radiólogos de retirar 3 mm del borde de la imagen.

C. Segmentación y estimación de fracción glandular

Después de eliminar las regiones extra-mamarias, como podemos apreciar en la Fig. 2(a), las imágenes se redimensionaron a su tamaño original y se llevó a cabo la segmentación. Para resaltar los tejidos internos de la mama, se aplicó un filtro Laplaciano local rápido el cual aproxima el algoritmo al discretizar el rango de intensidad en un número n de muestras definidas [11] como podemos ver en la Fig. 2(b). Suponiendo la presencia de píxeles de fondo y píxeles correspondientes a 3 tejidos principales: adiposo, conectivo y glandular, se seleccionó $n = 4$. Posteriormente, se utilizó el algoritmo k-medias [12], para establecer, en función del nivel de intensidad, las regiones correspondientes a los tejidos antes mencionados (ver Fig. 2(c)), donde los colores gris claro, gris oscuro y azul marino representan tejidos adiposo, glandular y conectivo respectivamente. Con esta información es posible conocer el porcentaje que ocupa cada tejido en la mama y así estimar la fracción glandular.

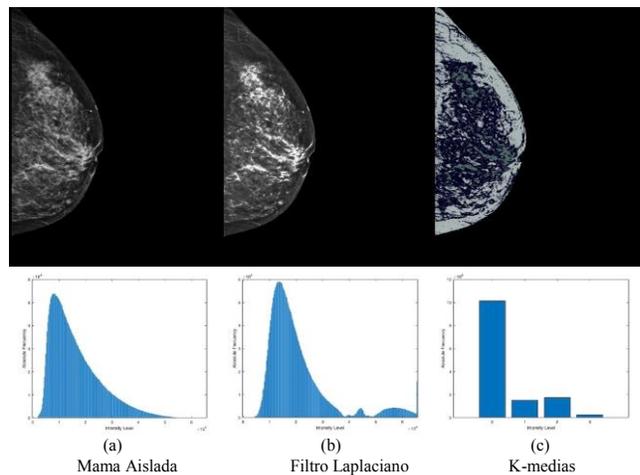


Fig. 2 Proceso de mejora de contraste y segmentación de la mama aislada.

Los histogramas mostrados en la Fig. 2 nos permiten confirmar que existe una mayor diferenciación de los niveles de intensidad en la imagen contrastada, lo que nos permite generar una imagen bien segmentada con un histograma remuestreado al número de etiquetas correspondientes a los tejidos de interés.

III. RESULTADOS

Las segmentaciones y estimaciones de FG obtenidas con el enfoque propuesto se compararon con las obtenidas en un trabajo previo de forma manual utilizando una Interfaz Gráfica de Usuario (GUI), donde solo estaban disponibles 2

clases (tejido glandular y adiposo), así como con uno de los algoritmos de agrupamiento más usados y citados en la literatura científica, el Agrupamiento Espacial Basado en Densidad de Aplicaciones con Ruido (DBSCAN por sus siglas en inglés) [13] y, finalmente, con la referencia empírica de una especialista. Las comparaciones se detallan a continuación.

A. Segmentación y estimación de fracción glandular

Comparando un caso positivo como el que se muestra en la Fig.3, se puede observar que en todas las segmentaciones de k-medias se suprimió la parte de la región del pliegue axilar, de la misma manera que el pezón, el borde de la piel y otros pixeles dispersos antes considerados lo que nos permitió obtener una mamografía donde solo estén presentes los tejidos de interés. La región gris oscuro que corresponde al tejido glandular en las Fig. 3 (b), 3 (c) y 3 (d) considera un área mucho mejor delimitada que la observada en la Fig. 3 (a), en donde la región en rojo representa el tejido glandular y es claro que se tiene una sobreestimación. Sin embargo, la aproximación más aceptable visualmente para la determinación de tejido glandular corresponde a la de la Fig. 3 (c).

Las segmentaciones de cada tejido en cada imagen se realizaron considerando 4 medias, asignando una media a cada tejido mencionado en la aplicación del filtro Laplaciano; 5-medias consideró 2 de estas medias como tejido conectivo (2TC) y 5-medias consideró 2 de estas medias como tejido glandular (2TG). Las fracciones estimadas se compararon con las obtenidas en un trabajo previo y las empíricas de la especialista. Los resultados para todos los casos se registraron en la Tabla 1.

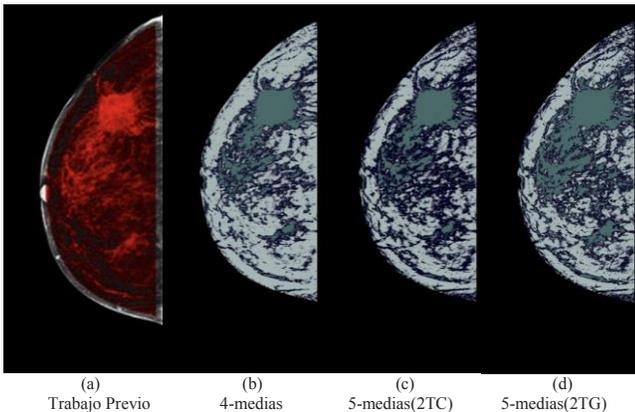


Fig. 3 Segmentación de un caso positivo con el algoritmo previo y el actual con k-medias.

B. DBSCAN

Es un algoritmo de agrupación en clústeres basado en densidad que está diseñado para encontrar clústeres y ruido en los datos. En términos generales, se define un radio de búsqueda ϵ alrededor de un punto y si esa vecindad contiene al menos $minpts$ vecinos, entonces se identifica el punto como central.

A pesar de que el algoritmo fue aplicado a las imágenes que fueron redimensionadas para evitar una sobre-

segmentación debido a la variabilidad de intensidad, y la variación de los parámetros de control, podemos observar en la Fig. 4 que realmente no se aprecia una buena diferenciación de los tejidos internos como con el algoritmo propuesto.

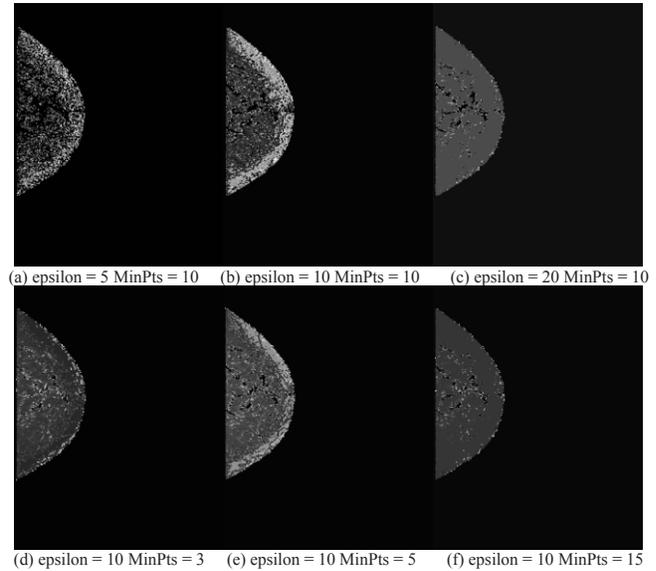


Fig. 4. DBSCAN aplicado variando parámetros epsilon y minpts.

IV. DISCUSIÓN

Los tejidos glandulares estimados a partir de imágenes segmentadas con 4-medias no estaban tan lejos de la realidad, visualmente la distribución de las regiones fue un primer criterio de descarte. Por otro lado, en las imágenes segmentadas con 5-medias (2TG) hubo aumentos en las fracciones glandulares y en las imágenes segmentadas con 5 medias (2TC) hubo disminuciones, por lo que se procedió a la estimación numérica para establecer resultados concluyentes.

La fracción glandular estimada a partir de casos negativos en el trabajo anterior estuvo en una media de $59 \pm 27\%$, porcentaje que disminuyó a $23 \pm 4\%$ para el caso de 5-medias (2TG) que considerada la mayor área de tejido glandular, esto representa una disminución considerable, de más del 45% en la media (μ) y más del 20% en la desviación estándar (σ). En el caso de 4 medias, la disminución fue aún mayor, alcanzando un $11 \pm 5\%$ y, en el caso de 5-medias (2TC), alcanzó un 84%. Considerando la baja proporción de tejido glandular en pacientes sanos, se reforzó el buen abordaje de la fracción glandular.

Por otro lado, las estimaciones de los casos positivos en el trabajo anterior fueron del $41 \pm 16\%$, lo que en principio es inconsistente al registrar porcentajes inferiores a los casos negativos. De hecho, el porcentaje disminuyó a $25 \pm 5\%$ para el caso de 5-medias (2TG), esto representa una disminución de aproximadamente 16% en μ y 11% en σ . Para el caso de 4-medias, la disminución alcanzó el $11 \pm 7\%$ y, en el caso de 5 medias (2CT), alcanzó el $10 \pm 5\%$. Nótese que la disminución en estas proporciones fue menor que la

Caso	Mama Derecha					Mama Izquierda				
	T. Previo	Especialista	4-medias	5-medias(2TC)	5-medias(2TG)	T. Previo	Especialista	4-medias	5-medias(2TC)	5-medias(2TG)
Neg 1	68.7 %	3.0 %	7.5 %	7.2 %	21.4 %	84.3 %	3.0 %	7.7 %	7.2 %	17.3 %
Neg 2	85.5 %	65.0 %	9.4 %	8.4 %	17.9 %	97.6 %	65.0 %	15.6 %	13.8 %	24.4 %
Neg 3	81.5 %	45.0 %	13.3 %	12.2 %	24.4 %	99.9 %	40.0 %	13.8 %	12.1 %	24.3 %
Neg 4	98.6 %	70.0 %	15.3 %	14.1 %	24.4 %	89.2 %	65.0 %	7.9 %	7.6 %	22.2 %
Neg 5	36.6 %	35.0 %	16.6 %	15.8 %	23.7 %	35.4 %	40.0 %	15.3 %	13.8 %	26.0 %
Neg 6	62.7 %	1.0 %	20.9 %	18.3 %	32.0 %	43.6 %	1.0 %	18.4 %	16.2 %	28.9 %
Neg 7	24.5 %	35.0 %	16.9 %	13.9 %	24.3 %	20.6 %	30.0 %	14.8 %	12.8 %	25.3 %
Neg 8	13.3 %	30.0 %	3.3 %	3.2 %	16.4 %	24.3 %	30.0 %	6.9 %	6.4 %	27.2 %
Neg 9	55.7 %	40.0 %	11.8 %	10.1 %	21.2 %	42.1 %	25.0 %	8.1 %	7.8 %	24.4 %
Neg 10	67.6 %	45.0 %	6.1 %	6.0 %	19.9 %	61.7 %	65.0 %	4.5 %	4.4 %	15.7 %
Pos 1	28.9 %	30.0 %*	1.1 %	1.0 %	21.1 %	22.5 %	50.0 %	11.2 %	0.6 %	20.0 %
Pos 2	62.6 %	70.0 %*	17.9 %	15.4 %	28.8 %	55.5 %	70.0 %	17.7 %	15.4 %	28.7 %
Pos 3	58.3 %	55.0 %	9.6 %	9.0 %	24.4 %	52.3 %	55.0 %*	9.6 %	8.9 %	22.6 %
Pos 4	64.4 %	50.0 %	12.5 %	11.6 %	27.9 %	65.9 %	45.0 %*	11.3 %	10.8 %	29.2 %
Pos 5	32.5 %	55.0 %	12.9 %	12.1 %	20.5 %	43.3 %	70.0 %*	8.7 %	8.2 %	19.2 %
Pos 6	15.6 %	80.0 %*	22.0 %	0.6 %	31.2 %	19.6 %	75.0 %	30.6 %	11.7 %	49.0 %
Pos 7	34 %	4.0 %*	6.2 %	6.0 %	24.2 %	42.9 %	5.0 %	12.0 %	10.0 %	22.9 %
Pos 8	40.1 %	12.0 %	3.7 %	3.5 %	15.8 %	38.3 %	10.0 %*	5.0 %	4.8 %	17.1 %
Pos 9	28.2 %	65.0 %*	7.5 %	7.3 %	21.7 %	25.4 %	60.0 %	9.1 %	8.6 %	18.9 %
Pos 10	51.3 %	35.0 %	11.0 %	10.2 %	25.0 %	53.8 %	40.0 %*	7.5 %	6.9 %	19.4 %

disminución en los casos negativos, lo que sugiere que se mantiene la gran cantidad de tejido denso esperado en los casos positivos.

A pesar de la discrepancia entre la estimación empírica de FG por parte de la especialista clínica y el valor estimado por el algoritmo propuesto en este trabajo, tanto las estimaciones numéricas como las visuales fueron validadas según las consideraciones de los especialistas y coincidieron en que los resultados obtenidos con 5-medias (2TC) reflejaban una FG más realista. Por tal motivo resulta importante estudiar la metodología propuesta en una mayor cantidad de imágenes y evaluar la forma empírica en como el especialista realiza la estimación de FG, para tener mayor evidencia de las bondades de la estrategia propuesta en este trabajo. Cabe mencionar que, además de los altos niveles de densidad mamaria, es importante analizar la distribución de las regiones densas presentes en la mama, así como considerar si el carcinoma se encuentra en ambas mamas o solo en una.

V. CONCLUSIONES

El enfoque propuesto para el aislamiento de la mama basado en la detección de bordes y las operaciones morfológicas sobre una mejora del contraste se realizó con éxito en todas las imágenes disponibles. Para la etapa de segmentación, la selección de 4 niveles para el filtro Laplaciano permitió una mejor diferenciación de los principales tejidos que constituyen a la mama (glandular, conectivo y adiposo), y junto con el uso de 5 clases para el algoritmo de k-medias considerando 2 clases intermedias como tejido conectivo, hicieron posible segmentar las cuatro regiones de interés definidas de una forma más robusta y congruente para la estimación de FG mamaria. Además, la estimación numérica refuerza las segmentaciones visuales y la estandarización de parámetros arrojó resultados más prometedores y precisos que los obtenidos en el trabajo anterior en donde solo se consideraban dos clases (tejido adiposo y glandular).

Como trabajo futuro, se considera realizar pruebas no solo en imágenes 2-D sino también imágenes volumétricas como las resonancias magnéticas, para poder tener un punto de referencia numérico que evite la necesidad de tomar como estandar de oro una estimación empírica.

REFERENCIAS

- [1] R. Pinsky, M. Helvie, and A. Arbor, "Mammographic Breast Density: Effect on Imaging and Breast Cancer Risk," *Journal of the National Comprehensive Cancer Network*, vol. 8, 2010, pp. 1157–1165.
- [2] G. Torres-Mejía, et al., "Mammographic Features and Subsequent Risk of Breast Cancer: A Comparison of Qualitative and Quantitative Evaluations in the Guernsey Prospective Studies," *Cancer Epidemiology, Biomarkers and Prevention*, vol. 14, 2005, pp. 1052–1059.
- [3] J. Harvey, and V. Bovbjerg, "Quantitative Assessment of Mammographic Breast Density: Relationship with Breast Cancer Risk," *Radiology*, vol. 130, 2004, pp. 29–41.
- [4] M. Carmen, et al., "Mammographic Breast Density and Race," *AJR Women's Imaging*, vol. 438, 2007, pp. 1147–1150.
- [5] S. Astley, et al., "A comparison of five methods of measuring mammographic density: a case control study," *Breast cancer research*, vol. 20, 2018, pp. 1–13.
- [6] A. Gastouniotti, et al. "Beyond breast density: a review on the advancing role of parenchymal texture analysis in breastcancer risk assessment," *Breast Cancer Research*, vol. 18, 2016, pp. 1–13.
- [7] E. Fowler, et al., "Generalized breast density metrics," *Physics in Medicine and Biology*, vol. 64, 2018, pp. 1–18.
- [8] J. Schindelin et al., "Fiji: an open-source platform for biological-image analysis," *Nature Methods*, vol. 9, 2012, pp. 676–682.
- [9] A. Zaafouri, M. Sayadi, and F. Fnaiech, "A Developed Unsharp Masking Method for Images Contrast Enhancement," *Eighth International Multi-Conference on Systems, Signals*, 2011, pp. 1–6.
- [10] J. Dim, and T. Takamura, "Alternative Approach for Satellite Cloud Classification: Edge Gradient Application," *Advances in Meteorology*, 2012, pp. 1–8.
- [11] M. Aubry, et al. "Fast local Laplacian filters: Theory and applications," *ACM Transactions on Graphics*, 2014, p. 167.
- [12] D. Arthur, and S. Vassilvitskii, "K-means++: The Advantages of Careful Seeding," *SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2007, pp. 1027–1035.
- [13] M. Ester, et al. "A density-based algorithm for discovering clusters in large spatial databases with noise," In *Proceedings of the Second International Conference on Knowledge Discovery in Databases and Data Mining*, 1996, pp. 226–231