

Descripción del borde utilizando medidas de compacidad aplicado al diagnóstico automático de cáncer de piel

Raul Santiago-Montero, Noemí Barajas and J. L. Fernando Suárez Y Gómez

Instituto Tecnológico de León, León, Gto,
México
rsantiago66@gmail.com

Resumen Medir la irregularidad del borde en regiones digitales ha despertado gran interés en la comunidad científica. Aplicaciones en el área médica, visión por computado, geometría digital colocan como un rasgo importante la descripción cuantitativa del borde de las regiones. Una aplicación donde se vuelve indispensable la cuantificación de esta propiedad en la detección temprana del melanoma, que es el tipo de cáncer de piel más letal; tres de cada 4 muertes por cáncer de piel se le atribuyen al melanoma. En el presente trabajo se presenta una propuesta para medir la irregularidad del borde de regiones digitales utilizando el concepto de compacidad. Muchas medidas de compacidad se han planteado anteriormente, pero solo algunas han probado una eficaz relación entre la irregularidad del borde y su valor cuantitativo. En el presente trabajo se muestran un conjunto de medidas de compacidad aplicadas a medir la irregularidad del borde de regiones digitales de lesiones de piel. Las descripciones obtenidas general un vector n-dimensional que es utilizado sobre un modelo de inferencia automática que genera un diagnóstico de la lesión; mostrando la eficacia de la propuesta.

Palabras clave: Compacidad, análisis de forma, geometría digital, reconocimiento de patrones.

1. Introducción

El cáncer de piel tiene tasas de crecimiento alarmantes por encima 400% según el reporte anual sobre cáncer [1]. En México, el Instituto Nacional de Cancerología estima el crecimiento en 500 % [1,2].y ocupa el tercer lugar entre los cánceres de la piel a nivel mundial [2]. Dentro de los cánceres de piel, el melanoma es el cáncer más mortal con el 80 % de las muertes y solo el 14 % de los pacientes con melanoma tiene un pronóstico de vida de más de 5 años. [3].

El único tratamiento para el melanoma es la extirpación de la lesión en etapas tempranas. Una vez que el melanoma se desarrolla es altamente resistente a tratamientos, tanto de quimio como radio-terapia y las expectativas de vida para el paciente son escasas. Es por ello la alta prioridad en la detección de este tipo de lesiones en piel apenas su aparición [1-3].

El diagnóstico de melanoma se sospecha por clínica, es decir, si la lesión cumple con ciertas características clínicas que la hagan sospechosa de ser maligna mediante la

dermatoscopia u otros métodos no invasivos y se confirma el diagnóstico con el estudio de las células del tejido enfermo usando procesos de histopatología [3].

La metodología no invasiva más usada para el diagnóstico de melanoma es la regla clínica ABCDE (A= Asimetría, B= Bordes irregulares, C= Color irregular, D= Diámetro, E= Evolución; este último fue introducido para describir los cambios en la forma de las lesiones a través del tiempo) [4-6]. El método se basa en evaluar un conjunto de rasgos o características geométricas y de color o textura de la lesión.

Dado el alto crecimiento en la incidencia del melanoma, el enorme número de regiones susceptibles a ser evaluadas que presenta el cuerpo humano (se estima en al menos 15) y la factible de cuantificar las características; los sistemas de diagnóstico asistido por computadora han despertado un enorme interés [5]. Se ha mostrado que estos sistemas tienen el mismo porcentaje de efectividad que el diagnóstico por clínica por dermatólogos con 2 o más años de experiencia [7]. Existe una cantidad significativa de propuestas que pueden ser divididas en tres segmentos: procesos automáticos de adquisición de imágenes de lesiones, procesamiento de las imágenes y análisis e inferencia [5].

La eficacia en la inferencia del sistema está determinada por la naturaleza de la descripción de la característica que se pretende medir. El presente trabajo se enfoca en la descripción cuantitativa de la característica B de la lesión. Esta característica intenta describir el borde mediante alguna relación o expresión que correlacione la irregularidad del bordo con un valor específico.

Existen una cantidad de artículos que detallan el proceso de encontrar una medida para la irregularidad del borde de una lesión en piel. Dos caminos para llegar a una medida de irregularidad han sido el índice de compacidad y la dimensión fractal. El índice de compacidad es el más utilizado por su facilidad de calcular [5, 8, 9].

Sin embargo, se argumenta que es inapropiado para describir el borde de la lesión. En general, el argumento se sustenta en dos hechos: la poca sensibilidad a evaluar la presencia de picos y concavidades, y su fácil alteración a la presencia de ruido [8, 9].

Otra propuesta para medir la irregularidad del borde consiste en dividir de la región de la lesión utilizando dos pares de ejes ortogonales que pasen por el centro de masa de ésta. Una variante de este proceso se describe en [6, 9].

En el 2003, Lee propuso los índices de irregularidad. Estos índices son obtenidos a partir de una transformación de la región bidimensional y obtener una función continua. La transformación utiliza de los puntos en el contorno y mide las diferencias que existen frente a una curva patrón [8].

El presente trabajo se enfoca en obtener una descripción cuantitativa del borde de una lesión en piel mediante medidas basadas en el concepto de compacidad de las formas y diseñadas para el espacio discreto. Este conjunto de medidas permiten generar un vector de características 4-dimensional que es sometido a proceso de clasificación y se obtengan diagnósticos sobre la presencia de un melanoma utilizando solo la característica B de la regla ABCD.

El resto del trabajo está dividido por la descripción de las herramientas utilizadas, el modelo propuesto, los resultados obtenidos y finaliza con una serie de conclusiones.

2. Compacidad y descripción del borde de una lesión

En esta sección se describen las propuestas utilizadas para cuantificar el concepto de compacidad. Además se presenta el modelo de descripción a través de un vector 3-dimensional que será la materia básica en el proceso de automatización del diagnóstico.

2.1. Medidas de compacidad

El concepto de compacidad hace referencia al grado de densidad-porosidad de un objeto o al grado disperso-compacto de la forma de este objeto [10]. En el caso de la forma, el concepto se aplica tanto al plano como al espacio. Cuantificar el concepto ha recaído sobre la teoría isoperimétrica desde hace varias décadas [11] y su relación base se ha aplicado como medida de compacidad de las formas [11-15], que se determina como:

$$C = \frac{P^2}{A} \quad (1)$$

No obstante su amplio uso, se detectaron efectos no deseables cuando la medida es aplicada sobre formas contenidas en imágenes digitales [8,16, 15]. La medida, en ámbito de las imágenes digitales, es afectada de manera importante por la resolución y la rotación de la forma bajo análisis. Esto es, la forma de un objeto contenida en una imagen digital a diferente resolución produce diferentes valores de compacidad, bajo la medida isoperimétrica. Esta inconsistencia se incrementa cuando se agrega una rotación a la forma.

Con el incremento y uso de imágenes digitales en diversas aplicaciones, el generar una medida adecuada para la compacidad se ha convertido, en si mismo, en un problema de investigación. Hasta ahora se encuentran reportadas 12 propuestas de medir el concepto de compacidad sobre formas en imágenes digitales [17,18]. De todas ellas, 4 propuestas se han tomado en función de su sencillez de implementación y que realizan la evaluación de la compacidad aplicando el enfoque de similitud entre la forma bajo estudio y una forma referencia; mismo enfoque que presenta la medida clásica de compacidad.

La primera propuesta que incluimos en este trabajo es la medida de circularidad de Haralick [19]. Esta medida determina la compacidad de una región utilizando el conjunto de distancias entre en centroide de la región y las celda en su borde. Una vez que es determinado el conjunto de distancias, se obtiene un valor de compacidad mediante la media de las distancias entre su desviación estándar. No obstante el mayor proceso de cómputo de la medida, la medida de circularidad de Haralick se muestra sensible a los cambios en la morfología de región. El primer paso para determinar la medida es calcular el centroide de la región digital. Shapiro define el centroide como la localización promedio de los elementos de una región digital [20], denotado como (\bar{r}, \bar{c}) donde:

$$\bar{r} = \sum_{(r,c) \in R} r \quad (2)$$

$$\bar{c} = \sum_{(r,c) \in R} c \quad (3)$$

Donde r y c son las posiciones en renglón y columna de las celdas de la región digital. A pesar de que los renglones y columnas son números enteros, el centroíde frecuentemente resulta en una posición en el espacio continuo, por lo tanto, se ubicará en la celda más próxima a su valor de renglón y de columna.

A partir de este punto de la región digital se calculan distancias euclidianas hacia cada celda en el borde de la región. La medida de circularidad de Haralick queda entonces expresada como:

$$C = \frac{\mu_R}{\sigma_R} \quad (4)$$

donde:

$$\mu_R = \frac{1}{K} \sum_{k=0}^{K-1} \|(r_k, c_k) - (\bar{r}, \bar{c})\| \quad (5)$$

y

$$\sigma_R = \left(\frac{1}{K} \sum_{k=0}^{K-1} (\|(r_k, c_k) - (\bar{r}, \bar{c})\|)^2 \right)^{1/2} \quad (6)$$

Por otra parte, Danielsson utiliza un enfoque diferente en el diseño de una medida que evalué la compacidad de las regiones digitales. El propone usar la información que proporciona cada celda a través de la transformada de distancia [21]. Al aplicar la transformada de distancia, Danielsson utiliza el principio del momento de inercia y lo normaliza para que el círculo produzca el valor mínimo. La propuesta es nombrada factor G y se expresa, para el espacio discreto, como:

$$G = \frac{N^3}{9\pi(\sum_1^N x_i)^2} \quad (7)$$

Donde N es el número de celdas en la región digital y x_i es el valor de la transformada de distancias en cada celda de la región.

La última medida de compacidad que se emplea en este trabajo es la propuesta por Peura [22]. Esta medida establece una relación entre dos regiones. La relación compara la región digital bajo estudio contra un círculo en el espacio continuo. La relación de circularidad de Iivarién y Peura se expresa como una relación entre el perímetro de la región digital y un círculo con igual área. La medida de circularidad de Peura está dada por:

$$C = \frac{2\sqrt{A\pi}}{P_D} \quad (8)$$

Donde A es el número de celdas de la región digital y P_D es el perímetro de la región digital.

2.2. Knn

El K-vecino más cercano (K-nearest neighbor) es un método de estimación no paramétrica muy bien conocido en el área de reconocimiento de patrones bajo representaciones vectoriales n-dimensionales de objetos a clasificar. Este método es particularmente efectivo cuando la distribución de probabilidad del conjunto de patrones es desconocida [23].

Entonces, sea L el número de clases, N el número total de patrones, N_i el número de patrones pertenecientes a cada clase, conocidos como prototipos. Sea Φ el conjunto de los prototipos y donde $y \in \Phi$ es un vector de dimensión n .


La inferencia o clasificación del KNN se determina por una función $f(x, y)$ y es una función no negativa y simétrica; donde el valor más alto denota la mayor similitud entre un el vector y y el vector x . Es común utilizar la función de distancia euclidiana como función de similitud.

Este método se emplea en el presente trabajo para generar un diagnóstico, sobre la lesión presente en una imagen digital binaria, entre melanoma (maligna) y una lesión benigna.

2.3. Descripción del borde de la lesión por compacidad


Las lesiones que se emplean son: imágenes segmentadas, binarizadas y diagnosticadas previamente. La descripción se realiza median un vector 3-dimensional de compacidad.

Tabla 1. Valores obtenidos para una lesión diagnosticada por dermatólogos como benigna.

Region benigna	Descripción
	0.731971643857869
	19.08214965799585
	1.687351050412136

En la Tabla 1 se obtiene la descripción de una lesión clasificada como benigna a través de un vector 3-dimensional de compacidad utilizando las medidas descritas y empleadas en este estudio; donde el primer valor es el obtenido por (8), la segunda por (4) y finalmente, se muestra el valor obtenido por (7). Por otra parte, la Tabla 2 muestra la descripción para una lesión diagnosticada como maligna o melanoma.

Tabla 2. Descripción obtenida por las medidas incluidas en el estudio para una lesión maligna

Región maligna (melanoma)	Descripción
	0.5319088810047378
	5.7119639900771855
	2.708315581066739

3. Resultados

En este estudio se utilizaron un conjunto de 20 imágenes binarias de las regiones digitales que contienen una lesión en piel humana. Las imágenes están tienen una clasificación previa conocida y divididas en dos categorías, diez son benignas y 10 malignas. Del conjunto de imágenes se seleccionaron 5 con regiones clasificadas como benignas y 5 como malignas como conjunto de entrenamiento del modelo KNN.

Tabla 3. Regiones usadas como conjunto de entrenamiento del clasificador KNN en el diagnóstico de melanoma.

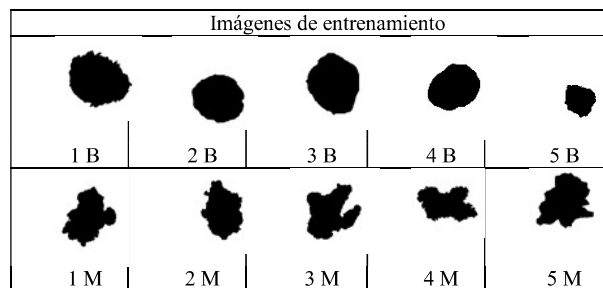


Tabla 4. Descripción de las regiones benignas mostradas en el primer renglón de la Tabla 3 mediante sus valores de compacidad.

Datos de entrenamiento: benignos			
Imagen	Peura	Haralick C	Factor G-
1	0.601	10.4558	1.894
2	0.731	19.0821	1.687
3	0.743	14.51696	1.769
4	0.752	10.5891	1.681
5	0.696	10.15409	1.527

En la Tabla 3 se muestran las regiones digitales binarizadas las lesiones de este conjunto y en la Tabla 4 y 5, sus vectores de características en función de las medidas de compacidad.

Las imágenes restantes se propusieron como imágenes de prueba, las cuales se presentan en la Tabla 6 y 7 con sus diagnósticos previos y el resultado arrojado por el modelo de clasificación en el último renglón. La eficacia de la representación propuesta de compacidad se puede observar en la Tabla 8 que muestra la matriz de confusión del experimento realizado.

4. Conclusiones

No obstante lo limitado de las muestras de los conjuntos de entrenamiento y prueba, los resultados obtenidos por los experimentos realizados muestran que la propues-

ta puede realizar una efectiva clasificación de lesiones de piel humana y de generar un diagnóstico del cáncer de piel, conocido como melanoma. Hay que destacar que la propuesta solo emplea una característica de la regla ABCD para el diagnóstico de melanoma empleada por expertos en cáncer de piel.

Tabla 5. Descripción de las regiones malignas ilustradas en el segundo renglón de la Tabla 3.

Datos de entrenamiento: malignos			
Imagen	Peura	Haralick C	Factor G
1	0.53190	5.711963	2.708319
2	0.58264	5.97185	2.09514
3	0.48831	4.61046	3.314071
4	0.4785	4.00625	3.07477
5	0.579341	6.30134	2.54245

Tabla 6. Conjunto de imágenes de melanoma, su descripción mediante compacidad y el resultado obtenido por el clasificador junto a un diagnóstico previo.






Imagen					
Peura	0.4601	0.5325	0.5195	0.4365	0.5663
Haralick	5.8355	3.3935	4.5333	5.6488	6.2934
Factor G	2.7250	2.4890	2.9476	2.3757	2.4715
Resul	M-M	M-M	M-M	M-M	M-M

Tabla 7. Conjunto de imágenes de lesiones de piel benignas, su descripción mediante compacidad y el resultado obtenido por el clasificador junto a un diagnóstico previo.






Imagen					
Peura	0.60443	0.60713	0.72447	0.71845	0.76838
Haralick	16.48476	12.69999	11.17549	9.19046	14.66675
Factor G	1.79331	1.74325	1.72806	1.88039	1.71360
Resul	B-B	B-B	B-B	B-B	B-B

Table 8. Matriz de confusión resultante del proceso de clasificación.

Lesión	Benigna	Maligna
Benigna	5	0
Maligna	0	5

Además se muestra que el concepto de compacidad es útil en la generación de una descripción del borde de una región digital. Las medidas aquí empleadas evitan los efectos nocivos que investigadores detecta en la medida clásica de compacidad.

El presente trabajo abre un campo para probar la sencillez, eficacia y lo robusto del concepto de compacidad en diversas aplicaciones donde el borde de las regiones sea un factor necesario para procesos de evaluación y clasificación.

Referencias

1. R.M. Gutierrez Vidrio y N. Cortés Lozano: Confrontando al melanoma en el siglo XXI. *Med Cutan Iber Lat Am*, 35(1), pp. 3-13 (2007)
2. A. De la Fuente-García y J. Ocampo-Candiani: Melanoma cutáneo. *Gac. méd. Méx.*, 146(2), pp. 126-135 (2010)
3. A. Alfaro, L. Castrejón y M. Rodríguez Ortíz: Cáncer de piel. Estudio epidemiológico a 10 años de derechohabientes del ISSSTE en Nuevo León. *Dermatología Rev Mex*, 54(6), pp. 321-325 (2010)
4. M. d'Amico, M. Ferri, and I. Stanganelli, I.: Qualitative asymmetry measure for melanoma detection. In: *ISBI2004: Proceedings of the IEEE International Symposium on Biomedical Images* (Arlington, VA) (2004)
5. I. Maglogiannis, C. N. Doukas: Overview of Advanced Computer Vision Systems for Skin Lesions Characterization. *Information Technology in Biomedicine, Trans on.*, 13(5), pp. 721-733 (2009)
6. A. Bono et al.: The ABCD System of Melanoma Detection, A spectrophotometric Analysis of the Asymmetry, Border, Color, and Dimension. *Cancer*, 85(1), pp. 72-77 (1998)
7. S.M. Rajpara, A. P. Botello, J. Townend y A. D. Ormerod: Systematic review of dermoscopy and digital dermoscopy/ artificial intelligence for the diagnosis of melanoma. *British Journal of Dermatology*, 161, pp. 591-604 (2009)
8. T.K. Lee and Ela Claridge: Predictive power of irregular border shapes for malignant melanomas. *Skin Research and Technology*, 11, pp. 1-8 (2005)
9. Tor-Björn Holmström: A Survey and Evaluation of Features for the Diagnosis of Malignant Melanoma. *Mater Thesis* (2005)
10. Real Academia Española: *Diccionario de la lengua española*. Ed. 23. [11] R. O.
11. K.R. Castleman: *Digital Image Processing*. Prentice Hall (2010)
12. L.F. Costa and R. M. Cesar: *Shape Analysis and Classification: Theory and Practice*. CRC Press, Boca Raton, Florida, USA, (2009)
13. R.C. Gonzalez and R. E. Woods: *Digital Image Processing*. Prentice Hall, Upper Saddle, New Jersey, USA (2007)
14. A.K. Jain: *Fundamentals of Digital Image Processing*. Prentice Hall, USA (1989)
15. A. Rosenfeld: Compact figures in digital pictures. *Systems, Man and Cybernetics, Trans on.*, 4, pp. 221-223 (1974)
16. M.J. Bottema: Circularity of objects in images. In: *International Conference on Acoustic, Speech and Signal Processing, ICASSP 2000, Istambul*, pp 2247-2250, (2000).
17. N. Ritter and James Cooper: New Resolution Independent Measures of Circularity. *J Math Imaging Vis*, 35, pp. 117-127 (2009)
18. Santiago Montero, R. and Ernesto Bribiesca: State of the art of compactness and circularity measures. *International of Mathematical Forum*, 4(27), pp. 1305-1335 (2008)
19. R.M. Haralick: A measure for circularity of digital figures, *Systems. Man and Cybernetics, Trans on.*, SMC- 4(4), pp. 334-336 (1974)
20. L.G. Shapiro and G. C. Stockman: *Computer Visión*. Prentice Hall, Upper Saddle River, New Jersey, USA (2001)
21. P.E. Danielsson: A new shape factor. *Computer Graphics and Image Processing*, 7(2):292-299 (1978)

22. M. Peura and J. Iivarinen: Efficiency of simple shape descriptors. In: C. Arcelli, P. Cordella and G Sanniti de Baja, editors, *Advances in Visual Form Analysis*, pp. 443-451, World Scientific, Singapore (1997)
23. E. Fix and J. L. Hodges Jr.: Discriminatory Analysis: Nonparametric Discrimination Consistency Properties. Project 21-49-004, Report Number 4, USAF School of Aviation Medicine, pp 261-279 (1951)