

Filtros Morfológicos y Clasificación No Supervisada Aplicados a la Identificación de Tejido en Biopsias de Médula Ósea.

Guillermo N. Abras[†], Juan I. Pastore^{†‡}, Virginia L. Ballarin[†] y Manuel L. Gonzalez[†]

[†]Lab. de Procesos y Medición de Señales, Facultad de Ingeniería U.N.M.d.P., [‡] CONICET
gabras@fi.mdp.edu.ar

Resumen—Este trabajo presenta un método semiautomático de detección de tejidos en biopsias de médula ósea utilizando técnicas de Procesamiento Digital de Imágenes. La técnica utilizada combina filtros morfológicos y técnicas de reconocimiento de patrones, con el fin de realizar un cálculo preciso de la celularidad medular. Los informes anatómo-patológicos de estos cortes histológicos entregan resultados porcentuales de la celularidad medular, indicando la presencia de trabéculas, células adiposas y hematopoyéticas. Dichos porcentajes permiten evaluar la presencia y/o el grado de algún desorden metabólico, estableciendo comparaciones entre los valores normales y los patológicos. Generalmente, estas mediciones se realizan por simple inspección visual. El método propuesto permite calcular el porcentaje de trabéculas, células adiposas y hematopoyéticas utilizando como entrada al algoritmo de clasificación k-medias ponderado imágenes preprocesadas mediante operadores morfológicos.

El método fue probado con 26 imágenes de biopsias de médula ósea que presentan distribuciones variadas de celularidad medular. Se presenta un análisis comparativo entre los resultados obtenidos con el método propuesto y los calculados por un patólogo especialista donde la diferencia entre ambas estimaciones es de 5.87% promedio. Tanto los resultados visuales como los numéricos indican que el método es altamente eficiente para esta aplicación.

Palabras clave—Médula ósea, Filtros Morfológicos, k-medias ponderado, Segmentación.

I. INTRODUCCIÓN

Los informes de biopsias de médula ósea entregan resultados porcentuales de la presencia de trabéculas, células adiposas y hematopoyéticas. Dichos porcentajes se toman como criterio genérico mediante el cual se pueden determinar, entre otros factores, diversos desórdenes metabólicos [1]. Sin embargo, no es un problema sencillo realizar una correcta identificación de estos tres tipos diferentes de tejido, dado que los niveles de gris que los representan son muy similares [2]. Por lo tanto, es de interés contar con una herramienta que, a través de técnicas de Procesamiento Digital de Imágenes (PDI), contribuya a la cuantificación de las estructuras presentes, para realizar cálculos posteriores [3].

Las técnicas de reconocimiento de patrones han sido ampliamente utilizadas para la identificación de regiones en imágenes con diferentes texturas. Luego de obtener características para las diferentes regiones, un indicador de la textura existente se asigna a cada región. La variante del algoritmo k-medias utilizado en este trabajo tiene en cuenta el número de veces que un patrón específico aparece en la imagen cuando calcula los centroides de las clases. De esta

manera el número de cálculos involucrados resulta minimizado y como consecuencia los tiempos en la clasificación disminuyen. Este algoritmo resultó ser estable con respecto a los puntos de comienzo, previniendo además que aquellos patrones con poca presencia en la imagen tengan incidencia significativa en la clasificación de las diferentes regiones [4].

Para poder aplicar este algoritmo eficientemente es necesario asignar más de un característico a cada píxel de la imagen a segmentar. En el caso de imágenes de biopsias de médula ósea, obtenidas a través de un microscopio, se cuenta con una sola imagen y no múltiples. Para solucionar este problema se propone aplicar Filtros Frecuenciales Alternados (ASFs) por reconstrucción a las imágenes de biopsias de médula ósea y utilizar las sucesivas imágenes filtradas como característicos de entrada al algoritmo k-medias ponderado. Los operadores de apertura y cierre morfológicos, han demostrado ser útiles en aplicaciones de filtrado de imágenes debido a que los objetos brillantes u oscuros son eliminados, sin embargo carecen de utilidad por sí solos como característicos de texturas. Los ASFs presentan la ventaja de filtrar los objetos deseados sin alterar la forma original, lo que produce menor distorsión en la imagen y por ende son útiles para caracterizar una textura [5].

En las siguientes secciones se describen las principales características de las técnicas de PDI utilizadas y el método propuesto que permite la segmentación esperada, como así también los resultados y conclusiones.

II. FUNDAMENTOS TEÓRICOS

A continuación se presentan las principales definiciones de interés para el método propuesto.

A. Operadores Morfológicos: Filtros Secuenciales Alternativos (ASFs) por Reconstrucción.

La Morfología Matemática (MM) es una teoría basada en conceptos de geometría, álgebra, topología y teoría de conjuntos, creada para caracterizar propiedades físicas y estructurales de diversos materiales. La idea central de la MM es examinar las estructuras geométricas de una imagen por superposición con pequeños patrones localizados, denominados elementos estructurantes, en distintas partes de la misma. Las dos operaciones básicas de la MM son la dilatación y la erosión. De su composición se pueden generar operadores de filtrado, como por ejemplo, la

apertura y el cierre que a su vez dan lugar a nuevos filtros. [6]. Los ASFs consisten en la repetición de operaciones de aperturas (γ^{rec}) y cierres (Φ^{rec}) por reconstrucción con elementos estructurantes de tamaño creciente, interviniendo un segundo elemento estructurante usado en la reconstrucción [7].

Formalmente se define:

el filtro secuencial “OC” (opening - closing) (n iteraciones) como:

$$nb, bc - \gamma^{rec} \Phi^{rec}(f) = \gamma^{rec}(\Phi^{rec}(\gamma^{rec}(\Phi^{rec}(\gamma^{rec}(\Phi^{rec}(\dots \gamma^{rec}(\Phi^{rec}(f))\dots)))))) \quad (1)$$

el filtro secuencial “CO” (closing - opening) (n iteraciones) como:

$$nb, bc - \Phi^{rec} \gamma^{rec}(f) = \Phi^{rec}(\gamma^{rec}(\Phi^{rec}(\gamma^{rec}(\Phi^{rec}(\dots \Phi^{rec}(\gamma^{rec}(f))\dots)))) \quad (2)$$

donde n es un número entero positivo que representa el factor de escala del elemento estructurante b :

$$nb = \underbrace{\delta_b(\delta_b \dots (\delta_b(b)))}_{(n-1) \text{ veces}} \quad (3)$$

Convencionalmente $nb = \{(0,0)\}$ para $n = 0$.

Los ASFs por reconstrucción descomponen la imagen original en un conjunto de imágenes filtradas, las cuales pasarán a formar los vectores de característicos que alimentan el algoritmo de clasificación k-medias ponderado.

B. Algoritmo k-medias ponderado

El clustering es una de las técnicas de Reconocimiento de Patrones no Supervisadas que han sido ampliamente utilizadas para identificar regiones en una imagen. Son métodos determinísticos para agrupar datos de acuerdo a distintos tipos de características, cuando el nivel de información previa que se tiene es muy bajo. Uno de los métodos más difundidos de clustering es el algoritmo k-medias.

Dado un conjunto de objetos Ω y una cantidad n de características observables para cada objeto de la muestra, que forman un vector de características F en el espacio n -dimensional, y $W = \{W_1, W_2, \dots, W_p\}$ una partición de Ω , donde cada subconjunto W_i de Ω se llama clase, se define una partición $H = \{H_1, H_2, \dots, H_p\}$ sobre \mathbb{R}^n y una función D o medida de similitud, que asigna cada vector F de \mathbb{R}^n la clase $D(F) = H_i$, que refleje en la mejor forma posible la partición W . Esto quiere decir que si la observación $F(w)$ de un nuevo objeto w se encuentra en la región H_i de \mathbb{R}^n , se puede afirmar con cierto grado de error, que el objeto w pertenece a la clase W_i . Si la partición H de \mathbb{R}^n se correspondiera en forma biunívoca con la partición W de Ω se podría clasificar con absoluta certeza los objetos w de W con el conocimiento de F , lo cual normalmente no ocurre.

La medida de similitud usualmente está determinada por la proximidad de los puntos en \mathbb{R}^n de acuerdo a una función distancia definida, aunque puede basarse en otras propiedades. Una de las funciones distancia más utilizadas es la Euclídea:

$$D = \|v - z\| = \sqrt{(v - z)^t (v - z)} \quad (4)$$

donde v y z son dos vectores de dimensión n . Cuanto menor es la distancia, mayor es la similitud.

El algoritmo k-medias se basa en la minimización de una función criterio (o índice de performance) la cual consiste, en este caso, de la suma de los cuadrados de las distancias euclídeas de todos los puntos de un cluster al centro del mismo. Se eligen arbitrariamente centros iniciales para los p clusters y se asigna cada muestra al cluster cuyo centro está a menor distancia. Se recalcula el centro de cada cluster minimizando la función criterio, el cual resulta ser el promedio de los vectores F de las muestras pertenecientes al cluster. El algoritmo termina cuando en la siguiente iteración no cambian los centros de clusters. El comportamiento del algoritmo k-medias está poco influenciado por la elección de los centros iniciales. No se necesitan prototipos, sino que surgen de los propios datos.

Para este trabajo se utiliza el algoritmo k-medias ponderado, que es una variante del algoritmo k-medias, en el cual se tiene en cuenta para el cálculo de los centros de cluster, la cantidad de veces que aparece un determinado patrón F . Es decir, se utiliza una media ponderada en el cálculo de los centros.

Las imágenes a procesar están representadas por n matrices de igual dimensión. Los elementos de cada matriz representan los niveles de gris de los píxeles de posición (s, t) en cada una de las n imágenes. El vector de características $F = \{f_1(s, t), f_2(s, t), \dots, f_n(s, t)\}$, donde $f_i(s, t)$ es el nivel de gris de la imagen o matriz i -ésima, representa un punto (o patrón) en el espacio \mathbb{R}^n . El objetivo es formar con estos puntos, tres clusters o grupos H_1, H_2 y H_3 . Todos los píxeles de las imágenes cuyo vector F pertenezca a H_i son asignados a la clase W_i , obteniéndose de esta forma la imagen segmentada.

III. MATERIALES Y MÉTODOS

A. Materiales

El material fue procesado con las siguientes especificaciones:

- Las muestras se fijaron en líquido de Bouin durante 24 hs. (ó 48 hs., según los casos), procediendo a su descalcificación y a su inclusión en parafina.
- Los cortes histológicos fueron realizados en un micrótopo rotativo con un espesor de 4 μm .
- Las tinciones histológicas utilizadas fueron: Hematoxilina-Eosina, como técnica de rutina, coloraciones para Fibras Reticulares (reticulina de Gomori), coloraciones para hierro (Perls) y coloración de Giemsa.

Para este trabajo se estudiaron veintiséis (26) imágenes. Éstas se obtuvieron de un Microscopio Óptico Medilux-12 con un objetivo 4X Plano 4X, N.A. 0.10 digitalizadas a través de una cámara color CCD Hitachi KP-C550. Esta cámara tiene una dimensión eficaz de 682 (H) \times 492 (V) píxeles y un rango de longitud de onda de 400 a 700 nm.

Soporta una resolución de video de 430 líneas, suficientes para capturar imágenes de las biopsias de 640×480 píxeles en una PC. Después de seleccionar el área de interés en el microscopio se adquiere una única imagen en cada caso. Las imágenes fueron almacenadas en Mapa de bits y convertidas a escala de grises de 8 bits.

B. Método propuesto

El método propuesto para la detección de tejidos presenta los siguientes pasos:

Paso 1: Filtrado de la imagen mediante ASFs por Reconstrucción. Se aplica el filtro secuencial "CO", con elemento estructurante creciente. Este filtrado, junto con otras operaciones morfológicas y de ecuilización permiten generar 3 imágenes filtradas que formarán los vectores de características que alimentarán al algoritmo k-medias ponderado.

Paso 2: Selección de los centros de cluster: los centros de cluster iniciales se obtienen en forma sencilla desde la pantalla en la imagen original (Fig. 1).

Paso 3: El algoritmo de clasificación es alimentado con las tres imágenes obtenidas en el paso 1, por lo que el vector de características queda definido por $F = \{f_1(s,t), f_2(s,t), f_3(s,t)\}$ donde cada componente representa los niveles de gris de las imágenes filtradas (Fig. 2).

Paso 4: Visualización. El resultado del algoritmo, es una imagen segmentada con tres colores que representan cada uno de los tres tejidos presentes: trabécula, tejido adiposo y tejido hematopoyético (Fig. 3).

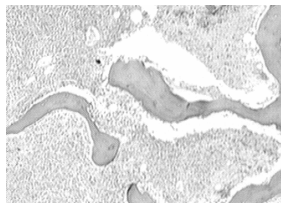


Fig. 1: Imagen de Biopsia de Médula Osea

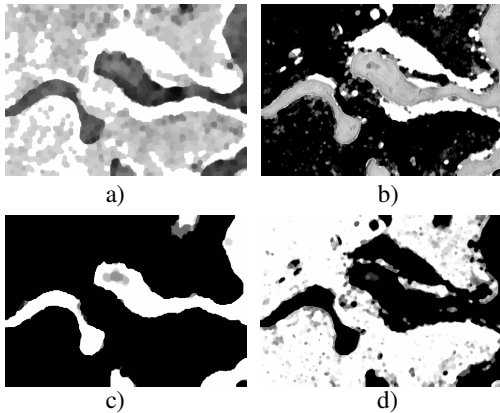


Fig. 2: Imágenes resultantes de aplicar a la imagen original los operadores morfológicos.

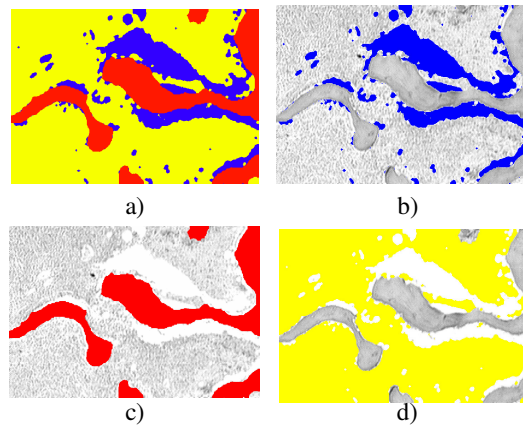


Fig. 3: Segmentación resultado. a) imagen resultado con los tres tejidos clasificados b) c) y d) resultados parciales para cada uno de los diferentes tejidos.

Para ilustrar los pasos del algoritmo propuesto se presentan los resultados obtenidos en las diferentes etapas de la segmentación en una imagen de biopsia de médula ósea. En primer lugar en la Fig. 1 se muestra la imagen original, a continuación en la Fig. 2 el resultado de aplicar a la imagen original los ASFs por reconstrucción. Las figuras 2-b a 2-d son las imágenes resultantes de los procesos morfológicos que alimentan el algoritmo de clasificación propuesto. En la figura 3 se pueden visualizar las imágenes que devuelve el algoritmo de clasificación; en rojo la trabécula, en azul las células adiposas y finalmente en amarillo el tejido hematopoyético.

IV. RESULTADOS

En la Figura 4 puede observarse los resultados obtenidos en cuatro de las veintiséis imágenes procesadas. En la primera columna se muestra la imagen original, en la segunda la segmentación objetivo o "Gold Standard" y finalmente en la tercer columna la segmentación lograda con el método propuesto.

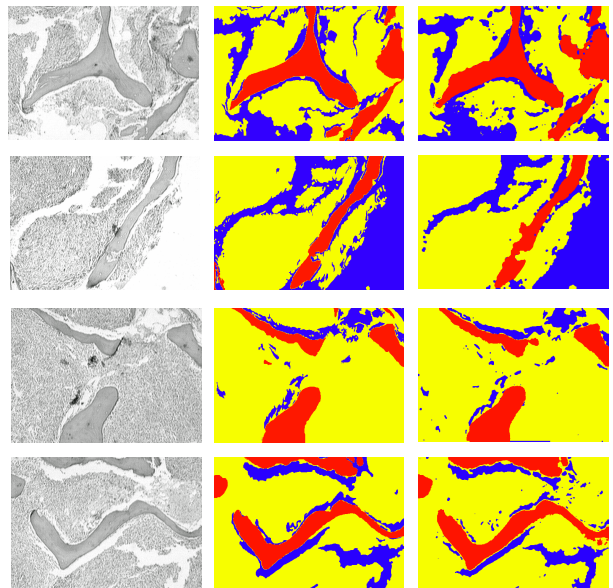


Fig. 4: Diferentes ejemplos de segmentaciones de Biopsias de Médula Osea. En la primer columna la imagen original, en la segunda el gold estándar, tercer columna la imagen resultado propuesta.

La segmentación objetivo fue obtenida manualmente por un experto patólogo. Puede apreciarse en todos los casos, comparando con la segmentación de referencia, que los resultados obtenidos son visualmente similares.

Sin embargo, cuando se define un nuevo algoritmo de segmentación es importante poder medir la exactitud de la técnica empleada de manera que el usuario final de la imagen clasificada pueda utilizar la información con mayor confiabilidad. Cuando se trata de segmentación de áreas, una de las formas más comunes de expresar la exactitud de una clasificación es la presentación del error de clasificación como matriz de error o matriz confusión [8], la cual, en este caso en particular, permite comparar el porcentaje de píxeles que han sido correctamente clasificados en cada uno de los diferentes tejidos que componen las imágenes, respecto de los clasificados por los expertos.

La Tabla I muestra los valores porcentuales promedio de las 26 imágenes procesadas. Los casilleros de la diagonal principal de la matriz muestran los porcentajes de píxeles correctamente clasificados. Se puede ver que los porcentajes para el tejido trabecular y el hematopoyético son muy elevados (97.5201% y 95.8290 respectivamente), mientras que los correspondientes al tejido graso son menores (77.1261%). También se desprende de los resultados, que el mayor grado de confusión se presenta en los píxeles correspondientes a tejido graso que son clasificados como tejido hematopoyético (22.0594%), lo cual es esperable dadas las características de los mismos en la imagen.

TABLA I

RESULTADOS FINALES Y REDUCCIÓN RELATIVA DE LOS ERRORES

% promedio de píxeles por clases	% de píxeles que deben ser clasificados como tejido trabecular	% de píxeles que deben ser clasificados como tejido graso	% de píxeles que deben ser clasificados como tejido hematopoyético
% de píxeles que fueron clasificados como tejido trabecular	97.5201 %	0.8144 %	2.6059 %
% de píxeles que fueron clasificados como tejido graso	0.3833 %	77.1261 %	1.5651 %
% de píxeles que fueron clasificados como tejido hematopoyético	2.0966 %	22.0594 %	95.8290 %

V. CONCLUSIONES

Se presentó un método de segmentación que permite la identificación de trabéculas, células adiposas y hematopoyéticas en imágenes microscópicas de biopsias de médula ósea. La segmentación en este tipo de imágenes es un proceso complejo ya que éstas presentan zonas difusas que dificultan la tarea de extracción de objetos. Con el método propuesto se puede lograr una eficiente segmentación de imágenes similares, lo que facilitaría el trabajo del profesional, disminuyendo la subjetividad del análisis.

Los algoritmos se desarrollaron en MatLab® 7. Luego de realizar las pruebas necesarias y recibir las opiniones de los especialistas, se podrá desarrollar en la plataforma que éstos requieran.

AGRADECIMIENTOS

Los autores agradecen a los especialistas: Dr. Ulises Zanetto y Dr. Fernando Pagani por su permanente asesoramiento y provisión de material histológico e instrumental.

REFERENCIAS

- [1] E. Rubin, J. Farber, "Patología: Fundamentos", Buenos Aires: Editorial Médica Panamericana. 1992.
- [2] B. Frisch, *Biopsy Pathology of Bone and Bone Marrow*, New York: Raven Press. 1985.
- [3] P.G. Bullough, M. Bansal y E.F. DiCarlo, "The tissue diagnosis of metabolic bone disease. Role of histomorphometry", *The Orthopedic Clinics of North America*, v. 21, p. 65-79. 1990.
- [4] G. Abras y V. Ballarin, "A weighted K-means algorithm applied to brain tissue classification", *Journal of Computer Science and Technology*, vol. 5, pp. 121-126, 2005.
- [5] J. Pastore, E. Moler y V. Ballarin, "Segmentation of brain magnetic resonance images through morphological operators and geodesic distance", *Digital Signal Processing*, vol. 15, pp. 153-160, 2005.
- [6] J. Serra, *Image Analysis and Mathematical Morphology*, Vol. I, London: Academic Press. 1982.
- [7] J. Serra, L. Vincent, "An overview of morphological filtering", *IEEE Transactions on Circuits, Systems and Signal Processing*, v. 11, p. 47-108. 1992.
- [8] F. Provost and T. Fawcett, "Analysis and Visualization of Classifier Performance: Comparison Under Imprecise Class and Cost Distributions," *Proc. Third International Conference on Knowledge Discovery and Data Mining*, pp. 43-48, 1997.