# Aplicación de la morfología matemática a la telemedicina y la biotecnología: caracterización morfológica de células de la sangre y análisis de *cDNA microarrays*

Jean Serra and Jesús Angulo
Centre de Morphologie Mathématique - Ecole des Mines de Paris,
35, rue Saint-Honoré, 77305 Fontainebleau, France
{serra,angulo}@cmm.ensmp.fr

**Resumen**

La morfología matemática es una metodología no-lineal de tratamiento de la imagen digital, basada en la aplicación de la teoría de retículos a estructuras espaciales. En este artículo, presentamos dos recientes aplicaciones del análisis morfológico a la imagen en biomedicina. La primera de ellas consiste en el desarrollo de una plataforma integrada para la cuantificación morfológica, el reconocimiento de patrones y la indexación basada en el contenido de células, a partir de imágenes en color de sangre periférica; considerando además el problema desde el punto de vista de la telemedicina. La segunda aplicación consiste en la automatización de la lectura de *cDNA micoraarays*. Esta nueva herramienta experimental de la biotecnología comienza a ser utilizada de manera masiva en los laboratorios y el análisis de imagen es un aspecto fundamental. Nuestro método permite una cuantificación completamente automática y precisa de la intensidad asociada a los spots, siendo además robusto frente a artefactos y variaciones indeseadas.

**Palabras-clave:** tratamiento digital de la imagen, indexación por el contenido, morfología matemática, segmentación morfológica, extracción de características, microscopía cuantitativa, telemedicina, biotecnología, citología hematológica, cDNA microarrays.

## 1 Introducción

En 35 años de desarrollo, la morfología matemática se ha convertido en una de las herramientas más utilizadas para el tratamiento de la imagen. La morfología matemática es una metodología no-lineal de tratamiento de la imagen digital, basada en la aplicación de la teoría de retículos a estructuras espaciales; y cuyos operadores son capaces de realizar sofisticadas tareas de procesamiento y análisis sobre imágenes 2D y 3D, a nivel de gris y en color. El lector interesado en profundizar en esta teoría puede consultar las obras de referencia de Serra [20]; o en el caso de querer conocer los últimos avances en teoría y aplicaciones, las tesis doctorales del *Centre de Morphologie Mathématique* de la *Ecole des Mines de Paris* constituyen una fuente indispensable. La morfología matemática ha demostrado en numerosas aplicaciones constituir una herramienta muy potente para el análisis de las imágenes de microscopía cuantitativa [18] [19].

En este artículo, describimos en detalle dos recientes aplicaciones, muy diferentes pero complementarias, del análisis morfológico sobre la imagen en biomedicina. La primera de ellas consiste en el desarrollo de una plataforma integrada para la cuantificación morfológica, el reconocimiento de patrones y la indexación basada en el

contenido de células a partir de imágenes en color de sangre periférica; considerando además el problema desde el punto de vista de la telemedicina. La segunda aplicación consiste en la automatización de la lectura de *cDNA microarrays*.

# 2 Caracterización morfológica de células de la sangre

## 2.1 Diagnostico morfológico en citología hematológica

Históricamente, el examen bajo microscopio de frotis sanguíneos de calidad ha sido la mejor manera de estimar el número y el tipo de los leucocitos o glóbulos blancos, para estudiar la morfología de los eritrocitos o glóbulos rojos (en el caso del paludismo, para detectar sus inclusiones parasitarias), para caracterizar la morfología de los linfocitos, para el estudio de las plaquetas, etc. En los últimos 30 años, se han desarrollado sistemas automáticos para el reconocimiento de los leucocitos, siendo ampliamente utilizados en rutina de laboratorio. Estos sistemas se basan en la citometría de flujo, pero están limitados a la identificación de células normales o a lo sumo, a la detección de las anormales. Sin embargo, el examen de frotis sanguíneos de sangre periférica continua siendo necesario para el análisis visual y la clasificación de células anormales, particularmente para el estudio de la morfología de los linfocitos, principio básico para la identificación y discriminación de los síndromes linfoproliferativos crónicos [23]. La clasificación de células de la sangre es un problema bien establecido en tratamiento de la imagen y en reconocimiento de formas, sobre el que se continúa trabajando [14] [10] [21].

## 2.2 Telepatología e interés de las láminas virtuales

Desde un punto de vista metodológico, hay dos cualidades necesarias para un frotis: un área suficiente de lectura sobre la lámina; y una calidad adecuada y reproducible del procedimiento de coloración. Ambas cualidades han sido previamente estudiadas [7] [8]; además, los aparatos para la preparación automática y coloración de los frotis, disponibles comercialmente, producen preparaciones relativamente buenas.

Diseñando una estrategia de automatización del estudio de los frotis y dentro de un contexto de telepatología hematológica [12] [13], el uso de estaciones de microscopía robotizadas permite la producción masiva de imágenes digitales microscópicas. La adquisición de estas imágenes permite su posterior almacenamiento, transmisión y tratamiento. La solución clásica pasaba por el concepto de dossier de imágenes (incluyendo una selección de imágenes de la lámina) pero que tiene el inconveniente de la falta de representatividad semántica del contenido de la lámina y que limita algunas aplicaciones. La solución alternativa, actualmente propuesta, es la lámina virtual o *teleslide* [16]: un *teleslide* es una imagen de alta resolución composición espacial de cientos o miles de imágenes del microscopio (al aumento usado para el diagnostico). Este tipo de macro-imágenes aseguran la representatividad del contenido de la lámina aunque, evidentemente, su gran tamaño implica la necesidad del desarrollo de herramientas especiales para su manipulación y análisis.

## 2.3 Plataforma para la hematología morfológica

En esta sección, presentamos la arquitectura de nuestro sistema *software* bajo la forma de una plataforma integrada para el tratamiento de la imagen, reconocimiento de patrones e indexación por el contenido de *teleslides* de sangre periférica. Este sistema de análisis de imagen, basado en operadores de la morfología matemática, proporciona una descripción objetiva de células de la sangre y más específicamente, el sistema ha sido diseñado para asistir a los hematólogos en el diagnostico de los síndromes linfoproliferativos. En la figura 1, se muestra el diagrama general de la

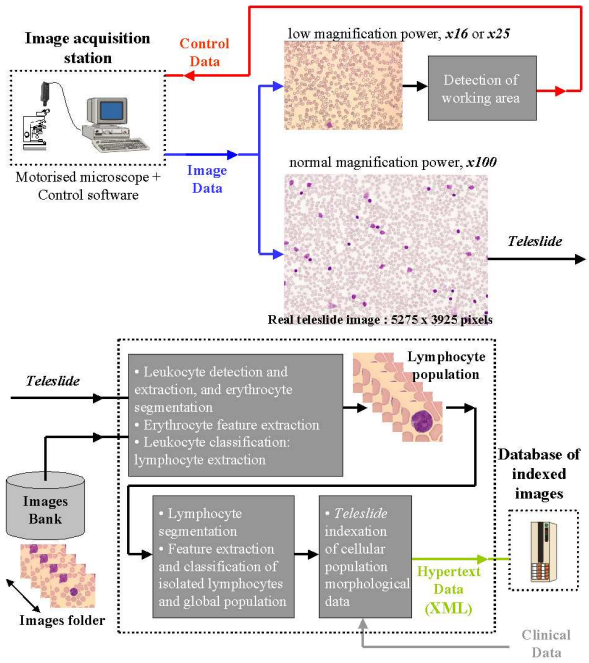



Figura 1: Análisis morfológico automático de células en frotis de sangre periférica: arriba, estación para la producción de los *teleslides*; y abajo, sistema integrado para el tratamiento de la imagen, reconocimiento de patrones e indexación basada en el contenido.

propuesta. Hemos separado claramente la producción de las imágenes *teleslides* (figura 1, arriba), del análisis de éstas (figura 1, abajo). De esta forma ambos subsistemas pueden estar separados físicamente y además, el análisis podrá realizarse sobre un dossier clásico de imágenes.

### 2.3.1 Automatización en la producción de los *teleslides* de frotis

Para facilitar la producción automática de las imágenes mediante un microscopio robotizado, hemos desarrollado un módulo de pretratamiento que implementa la detección automática de la zona de buena lectura del fotis a bajo aumento ($\times 16$ o $\times 25$). Los detalles de la técnica pueden encontrarse en [4]. Cuando la zona óptima es detectada, el objetivo del microscopio se sube al aumento para diagnostico ($\times 100$), adquiriendo toda la serie de imágenes campo que serán compuestas para obtener la imagen *teleslide*, $f_{\times 100}$.

### 2.3.2 Segmentación de las imágenes de células de la sangre

Motivados por la necesidad de una técnica de segmentación óptima y robusta para las aplicaciones de citología hematológica, hemos desarrollado una estrategia jerárquica para la extracción de las entidades visuales de los *teleslides*. El esquema de segmentación consiste en tres módulos independientes (detalles en [1]).

El objetivo del primer módulo es la segmentación de $f_{\times 100}$, ver figura 2 izquierda, obteniendo de una parte la capa de los eritrocitos $f_{\times 100}^{eryth}$ (que puede ser utilizada para el control de calidad de la técnica de coloración así como para el estudio, en aplicaciones posteriores, de la morfología de los eritrocitos) y de otra, la posición (centroide) de cada uno de los $n$ leucocitos presentes, lo cual permite su posterior extracción en

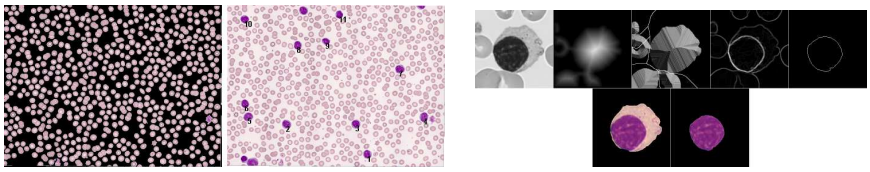


Figura 2: Segmentación de las imágenes: izquierda, segmentación del *teleslide* para obtener la capa de los eritrocitos y para localizar y extraer los leucocitos; y derecha, algoritmo morfológico para segmentar el linfocito, extrayendo su núcleo y su citoplasma.

subimágenes que constituyen la serie de imágenes leucocitos $\{f^{leuko}_{\times 100}(i)\}^{n}_{i=1}$. El algoritmo de esta etapa de segmentación es relativamente simple: sobre la componente verde (la más contrastada) de la imagen color se realizan dos umbralizaciones para obtener todas las células en una imagen, y los núcleos de los leucocitos en otra imagen. Los umbrales óptimos se obtienen a partir de un modelo predefinido para el histograma. Sobre estas imágenes binarias se realizan una serie de filtrados para eliminar artefactos.

En un segundo nivel, el objetivo buscado es la segmentación de los linfocitos $f^{lymph}_{\times 100}$, obteniendo el núcleo $f^{nucl_ly}_{\times 100}$ y el citoplasma $f^{cyto_ly}_{\times 100}$ (ver en la sección siguiente la forma de detectar los linfocitos entre todos los leucocitos). El método para segmentar estas imágenes es la línea divisoria de aguas (*watershed*) aplicada sobre un gradiente suma del gradiente morfológico de la componente verde y del de la saturación (ambas imágenes previamente simplificadas con un filtrado por reconstrucción); y los marcadores del *watershed* son obtenidos utilizando una técnica de umbralizacion similar a la descrita anteriormente. Ver ejemplo con imágenes intermedias en figura 2 derecha.

El último módulo de segmentación consiste en la partición de la cromatina del núcleo en regiones con densidad similar. La organización de la textura de la cromatina es una característica importante a cuantificar. Debido al tamaño reducido del núcleo y de las partes a caracterizar, la utilización de la técnica del *watershed* no garantiza que las regiones sean correctamente separadas, y por eso hemos preferido elegir una técnica alternativa denominada conexión por saltos (*jump connection*), basada en la extensión de la noción de zona plana. La conexión por salto de tamaño $k$ desde los mínimos, se define como una conexión compuesta de todos los conjuntos conexos alrededor de cada mínimo, y donde el valor de la imagen es menor que $k$ encima del mínimo; de forma dual se define para los máximos, o se puede tomar la intersección de ambas conexiones, tal y como nosotros hacemos para la cromatina. Para mejorar esta segmentación hemos implementado a continuación un algoritmo de crecimiento de regiones sobre el retículo de particiones que se obtiene en la conexión por saltos. Todo este tratamiento lo realizamos sobre una imagen a nivel de gris, $f^{nucl_ly*}_{\times 100}$, obtenida como la primera componente (aquella que maximiza la varianza) de un análisis por componentes principales sobre la imagen color $f^{nucl_ly}_{\times 100}$. A partir de la partición final obtenida, cada región podrá ser caracterizada, constituyendo el descriptor de la organización de la cromatina.

### 2.3.3 Clasificación de los leucocitos

Los leucocitos pueden ser clasificados en cinco categorías: (1) monocitos, (2) neutrófilos, (3) basófilos, (4) eosinófilos y (5) linfocitos. Sobre las imágenes $\{f^{leuko}_{\times 100}(i)\}$, las cinco clases de leucocitos pueden ser diferenciadas de acuerdo con las carac-

terísticas morfológicas (forma y textura) y de color. Para este propósito, hemos introducido un algoritmo de clasificación usando técnicas estadísticas (*template matching*, utilizando diferentes distancias sobre los histogramas como medida de similitud entre las plantillas y las entradas a clasificar). El color es caracterizado mediante los histogramas en los espacios HLS y Lab (con una cierta normalización basada en el color del plasma, es decir, en el fondo de la imagen) y la forma-textura es descrita mediante granulometrías morfológicas clásicas (familias de aperturas/cierres de tamaño creciente) extendidas al color. Después de una etapa de aprendizaje, donde diferentes clasificadores has sido evaluados, se ha propuesto un clasificador que utiliza básicamente la información de las granulometrías combinada con las características de color y cuyos resultados son bastante satisfactorios, para más detalles ver [5]. Mediante este módulo, a cada imagen $f_{\times 100}^{leuko}(i)$ se le asignará la categoría de leucocito correspondiente y sobre todo, podremos detectar los linfocitos, cuyas imágenes pasarán a denominarse $\{f_{\times 100}^{lymph}(j)\}_{j=1}^{m}$, y que constituyen la población de $m$ células a caracterizan de forma detallada.

### 2.3.4 Extracción de características de los linfocitos y su clasificación

Tomando las imágenes color segmentadas del núcleo y del citoplasma, $f_{\times 100}^{nucl_ly}(j)$ y $f_{\times 100}^{cyto_ly}(j)$ respectivamente, como punto de partida, todo un conjunto de parámetros cuantitativos deben ser calculados.

Hemos definido alrededor de 30 parámetros de forma, textura y color; incluyendo medidas complejas de textura como las curvas granulométricas para un análisis complementario de la cromatina, la detección y extracción de nucleolos, etc., ver en la figura 3 algunos ejemplos. Estas medidas proporcionan una primera clasificación de cada una de las características morfológicas retenidas (1- tamaño del núcleo, 2- tamaño de la célula, 3- ratio núcleo/célula, 4- densidad de la cromatina, 5- forma del núcleo, 6- nucleolo, 7- basofilia del citoplasma, 8- granulaciones del citoplasma, 9- forma del citoplasma, 10- excentricidad del núcleo) en una categoría: por ejemplo, el tamaño del núcleo podrá ser muy pequeño, pequeño, medio o grande. El clasificador asociado a cada característica morfológica ha sido obtenido mediante técnicas clásicas de aprendizaje a partir de una base de datos de referencia, en un contexto de lógica difusa. Para cada célula, el conjunto de valores de todas las características morfológicas constituyen el descriptor del linfocito.

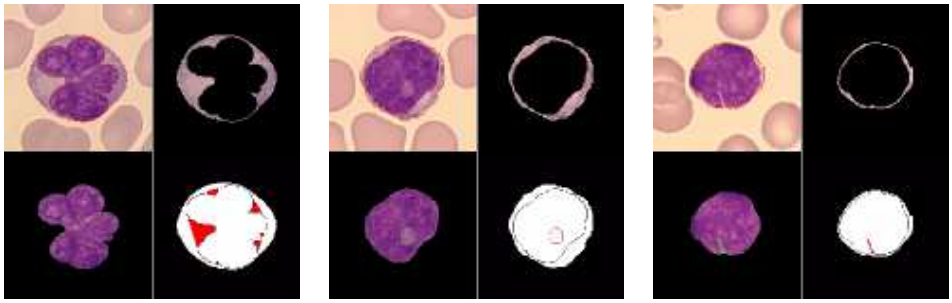


Figura 3: Ejemplos de la extracción de características: izquierda, cuantificación de la forma de un núcleo multilobular; centro, detección y extracción del nucleolo; derecha, detección de hendiduras en el núcleo.

En la etapa siguiente, este descriptor morfológico del linfocito (que ya puede ser incluso interpretado por un utilizador humano) es la entrada para un algoritmo de clasificación del linfocito en una de las 20 categorías celulares que se han definido, en

base al consenso relativo histórico de la hematología. Este clasificador es básicamente un árbol de decisión difuso.

Ya que una categoría celular será asociada a cada uno de los $m$ linfocitos de la población, después de forma estadística deberán definirse las categorías o subpoblaciones presentes así como su presencia relativa.

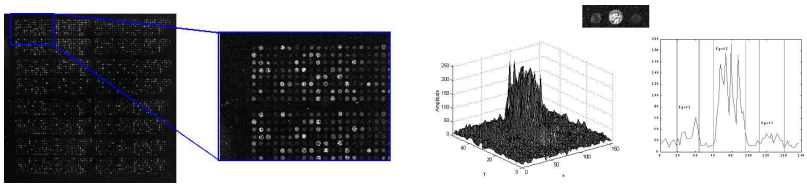
#### 2.3.5 Indexación de la información

Una vez que el conjunto de los descriptores morfológicos y de la información de la clasificación han sido obtenidos, mediante el modulo de indexado toda esta descripción resultante (imágenes y datos alfanuméricos: *metadata*) toma la forma de un documento en formato XML (utilizado en la norma MPEG-7, ya que es muy eficiente para la edición, búsqueda, filtrado y procesado de la información en bases de datos). Esta descripción morfológica proporciona una información objetiva que podrá ser indexada junto con el *teleslide*, así como con otros datos complementarios (características clínicas, inmunofenotipado, genética, etc) en una base de datos.

## 3 Análisis de *cDNA microarrays*

### 3.1 *cDNA microarrays* y análisis de imagen

Los *cDNA microarrays* [1] constituyen una tecnología experimental para identificar y cuantificar niveles de expresión en genes, y que consisten en matrices de miles de secuencias discretas de ADN depositadas sobre laminas para microscopio u otro soporte similar. Para comparar la abundancia relativa de los genes a estudiar en dos muestras de ADN, las dos muestras son etiquetadas usando dos tintes fluorescentes (Cy5 y Cy3), entonces son mezcladas con los spots de ADN depositados, interactuando biológicamente. Después de la hibridación, se realizan separadamente medidas de fluorescencia sobre cada tinte. Estas medidas son utilizadas para determinar el ratio, es decir, la abundancia relativa, de cada secuencia de gen en las dos muestras [9].



(a) (b)

Figura 4: (a) Imagen típica de una matriz (tamaño 2200 × 3000 pixels) y zoom, conteniendo miles de spots. (b) Variacion de la intensidad en tres spots (3D y corte horizontal).

Típicamente, las matrices varían entre algunos cientos hasta muchos miles o más de spots (cada pixel de la imagen es equivalente a $10\mu m$), figura 4(a). El tratamiento de la imagen es un aspecto importante en los experimentos con *cDNA microarrays* ya que las intensidades extraídas pueden tener una impacto potencialmente importante en los subsiguientes pasos de análisis de datos. El objetivo básico es reducir la imagen de los spots de diferentes intensidades a una tabla con una medida de intensidad para cada spot. El principal inconveniente está en el hecho que los spots

[1] En este artículo hemos decidido conservar el termino inglés *cDNA microarray* (en la literatura anglosajona también se designan como *biochips* o *genomic arrays*) debido a la falta de consenso sobre un nombre en español para esta tecnología, causado principalmente por la falta de publicaciones científicas en español.

son construidos utilizando un robot que dispone de una serie de pins que transfieren pequeñas cantidades de ADN. Por consiguiente, el desplazamiento mecánico del robot puede generar una distorsión geométrica y la imposibilidad del alineamiento de una plantilla. Otras técnicas intentan una segmentación automática de los spots usando un algoritmo de umbralización de intensidad adaptativo. Sin embargo, importantes errores son causados por la difícil elección del umbral óptimo (ver figura 4(b)): la frontera entre spot-fondo no es abrupta; el contraste (altura) entre las regiones spot-fondo y el volumen (integral de la intensidad) son muy diferentes de un spot al otro; y además la no homogeneidad del proceso de hibridación hace que las regiones de los spots estén "rotas".

Varias técnicas y herramientas software han sido desarrolladas para ser utilizadas en el procesamiento de las imágenes de las matrices [11] [15] [22] [24]. En es sta sección, presentamos nuestra técnica automática y no-supervisada basada en la morfología matemática para detectar las regiones de la imagen correspondientes a los spots y cuantificar sus intensidades asociadas.

## 3.2 Perspectiva general del algoritmo

La técnica morfológica para el tratamiento de las imágenes de microarrays se divide en seis sub-algoritmos. Los detalles completos de estos pasos son dados en [2]. La entrada es el par de imágenes escaneadas y la salida es la intensidad asociada a cada uno de los spots de cada imagen. Los principales pasos pueden ser resumidos cómo sigue (ver figura 5).

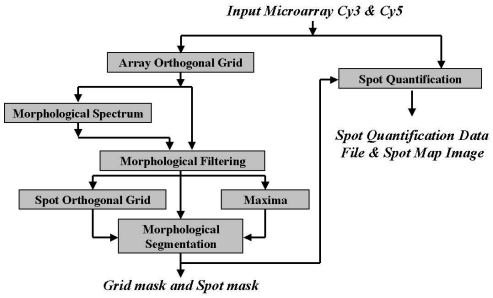



Figura 5: Diagrama de flujo de los algoritmos morfológicos para el procesamiento de las imágenes de microarrays.

### 3.2.1 Rejilla ortogonal del array

Inicialmente, un algoritmo de mallado debe proporcionar la segmentación automática, mediante una rejilla, de la imagen de la matriz completa en sub-matrices, definiendo cada grupo de spots o sub-imágenes, las cuales serán analizadas individualmente. Sean $f_{Cy3}$ y $f_{Cy5}$ las imágenes de 16-bits escaneadas de las fluorescencias Cy3 y Cy5. Nuestro algoritmo para el tratamiento de las imágenes requiere un imagen simple y además, computacionalmente es conveniente que sea de 8-bits. La imagen propuesta $f$ es obtenida por medio de una *combinación lineal sopesada con los valores de mediana* $\nu$, después de una transformación raíz cuadrada; i.e. $f = (\nu_{Cy3}\sqrt{f_{Cy3}} + \nu_{Cy5}\sqrt{f_{Cy5}})/(\nu_{Cy3} + \nu_{Cy5})$. Para segmentar los grupos de spots, la imagen $f$ es reducida en tamaño por medio de un *diezmado con promediado* (tamaño $K = 4$); i.e., $\widetilde{f} = f \div 4$. Entonces, los grupos de spots son realzados mediante el supremo de un cierre vertical $\varphi_n^\pi$ y un cierre horizontal $\varphi_n^0$; i.e, $\widetilde{f}^\bullet = \varphi_n^\pi(\widetilde{f}) \vee \varphi_n^0(\widetilde{f})$. El tamaño $n$ del cierre así como otros parámetros de tamaño dependen de cada microar-

ray; y en la sección siguiente será presentada una técnica espectral morfológica para calcular el tamaño aproximado del spot y por consiguiente, el valor de los parámetros de tamaño. A partir de la imagen realzada, son calculadas las proyecciones horizontales y verticales. El objetivo del subsiguiente filtrado morfológico unidimensional es simplificar la señal proyección, eliminando la contribución de los spots y del ruido de fondo. Sea $P^{\tilde{f}^{\bullet}}(i)$ la proyección horizontal o vertical, el tratamiento de la señal para obtener la rejilla se ejecuta en tres pasos: *filtrado intra-bloque* usando una apertura de tamaño $n_{ib}$; i.e., $P_{ib}^{\tilde{f}^{\bullet}} = \gamma_{n_{ib}}(P^{\tilde{f}^{\bullet}}(i))$; *extracción de los bloques* por medio de una transformación de chistera (*top-hat*) de tamaño $n_b$, que extrae los bloques; i.e., $P_b^{\tilde{f}^{\bullet}} = P_{ib}^{\tilde{f}^{\bullet}} - \gamma_{n_b}(P_{ib}^{\tilde{f}^{\bullet}})$; y *umbralización*, el valor óptimo para el umbral $u_P$ es definido como el 20% del valor medio de $P_b^{\tilde{f}^{\bullet}}$. Después de la umbralización, una señal binaria unidimensional (bloques y fondo) es utilizada para definir el origen $(x_0(j), y_0(j))$ y las dimensiones $[xsize(j), ysize(j)]$ para cada bloque $j$ que son extraídos a partir de la imagen $f$. Entonces, el análisis de las imágenes des los grupos de spots $\{f^j\}$ se lleva a cabo en cinco pasos.

### 3.2.2 Ley de distribución de tamaño de spot

En [3], ha sido presentado el espectro morfológico de extinción (histograma de los extremos) caracterizado por tres parámetros o medidas (contraste, área o volumen) así como su aplicación para el análisis de imágenes de matrices genómicas. Una nueva herramienta: *la ley de distribución de tamaño de spot* ha sido introducida, basada en el espectro de extinción por área en escala logarítmica, $SS[\lambda, f^j]$, donde $\lambda$ es el tamaño del spot (área) y $SS[\lambda, f^j] = n_\lambda$ denota el número normalizado de ocurrencias del valor de extinción $\lambda$ (se trata de una función densidad de probabilidad). El $SS[\lambda, f^j]$ está usualmente constituido de varios modos (el modo más interesante es el primero y los otros pueden considerarse como armónicos de la frecuencia morfológica fundamental). Este modo fundamental, o tamaño medio de spot, proporciona el umbral de área para el subsiguiente filtrado.

### 3.2.3 Filtrado morfológico por área

La contribución del ruido de fondo puede ser importante e introducir errores en la construcción de la rejilla y en la definición de la frontera de los spots. La extracción de este ruido de fondo se obtiene usando un *filtrado morfológico por área* $\gamma_n^a$. Usando esta técnica de filtrado sobre la imagen grupo de spots $f^j$; i.e., $f_s^j = \gamma_{\lambda_s}^a(f^j)$, las estructuras con un área mayor o igual que un umbral dado $\lambda_s$ son preservadas en $f_s^j$. Así pues, el valor del umbral $\lambda_s$ óptimo para cada imagen es una elección importante que depende del tamaño de los spots, siendo esta la justificación del calculo de $SS[\lambda]$. Otra ventaja importante del filtrado morfológico por área es la selección implícita de los máximos: después de este filtrado por reconstrucción hipotéticamente hay un y solamente un máximo asociado a cada spot.

### 3.2.4 Rejilla ortogonal de los spots

Dentro de un grupo de spots, éstos están situados de acuerdo con un alineamiento ortogonal y de nuevo, utilizando las proyecciones horizontal y vertical, se debe obtiener una rejilla correspondiente a los spots. El algoritmo es como sigue. Sea $P(i)$ la proyección horizontal y vertical: (1) calcular el valor medio de los elementos de $P(i)$; i.e, $\overline{P} = \frac{1}{N}\sum_{i=1}^{N} P(i)$; (2) sustraer la media de la proyección; i.e, $P_\eta(i) = P(i) - \overline{P}$; (3) reconstrucción morfológica de $P(i)$ utilizando $P_\eta(i)$ cómo marcador; i.e, $P^{rec}(i) = \gamma^{rec}(P(i); P_\eta(i))$; (4) tomar el residuo de la proyección inicial $P(i)$ y la reconstrucción $P^{rec}(i)$; i.e, $P_{TH}(i) = P(i) - P^{rec}(i)$; (5) estimar el

valor del umbral óptimo $u_P$, definido como $\alpha\%$ de la media del residuo $P_{TH}(i)$; i.e., $u_P = \frac{\alpha}{100}\frac{1}{N}\sum_{i=1}^{N} P_{TH}(i)$; (6) encontrar la señal binaria referencia $P_u(i)$ por medio de la umbralización en $u_P$ sobre la señal residuo $P_{TH}(i)$ y usar el $i$ medio de cada intervalo igual a 1 en $P_u(i)$, trazando la líneas rectas que corresponderán a la rejilla ortogonal. Después de diferentes tests en nuestra base de datos, la elección de $\alpha = 50\%$ aparece como un valor relativamente bueno.

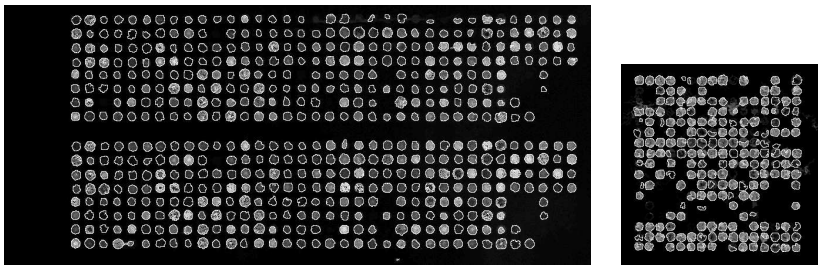


Figura 6: Dos ejemplos de segmentación de los spots.

### 3.2.5 Segmentación morfológica de los spots

La segmentación de los spots o definición de sus fronteras, se lleva a cabo utilizando la transformación de la línea divisoria de aguas (*watershed*): (1) la *función a inundar* es el gradiente externo $g^+$, definido cómo la diferencia entre la imagen dilatada y la imagen original; i.e, $g^+(f_{ss}^j) = \delta_B(f_{ss}^j) - f_{ss}^j$ (antes, la imagen filtrada $f_s^j$ es simplificada de nuevo por medio de un filtro de nivelado (*leveling*); i.e., $f_{ss}^j = \Lambda(f_s^j)$); (2) los *marcadores externos* $g_r$ serán los bordes de la imagen rellenados hasta el principio de la rejilla (marcadores del fondo) junto con la rejilla misma (marcadores de las regiones de influencia de los spots) y para los *marcadores internos* $mki$ se ha desarrollado un algoritmo específico que se basa en un análisis individual de la región de influencia rectangular de cada spot definida por la rejilla, y de los máximos después del filtrado por área; los marcadores globales a imponer vendrán dados por $mk = g_r \vee mki$; (3) construcción de la *línea divisoria de aguas* de cada spot $sm = Wshed(g^+(f_{ss}^j, mk)$, donde $sm$ son los contornos cerrados de cada spot. En la figura 6 son mostrados dos ejemplos de segmentación de los spots.

### 3.2.6 Cuantificación de los spots y extracción de ruido

La capa de segmentación $sm$, obtenida anteriormente se utiliza sobre las imágenes iniciales de 16-bits $f_{Cy3}$ y $f_{Cy5}$ para la cuantificación de los spots y la extracción del ruido. La intensidad calculada (integral en la región del spot) puede ser expresada como la suma de un *valor de intensidad de señal*, $s_i$, y un *valor de intensidad de ruido*, $n_i$, tal que $\hat{s}_i = s_i + n_i = s_i + \bar{N}_i A_i$, dónde $\bar{N}_i$ es el *ruido medio* del spot $i$ y $A_i$ es el *área* (número de pixels) del spot $i$. Brevemente, hay dos alternativas a estudiar. Primero, considerar que el ruido de fondo es uniforme sobre toda la matriz (más precisamente sobre todo el bloque $j$); o segundo, considerar que el ruido de fondo no es uniforme, y por consiguiente es necesario una estimación local por spot. Obviamente, la técnica de *ruido de fondo global* es más simple: hay que cuantificar solamente un valor medio, mientras que para el modelo de *ruido de fondo local* es necesario cuantificar el ruido en una región para cada spot. Basándose en la teoría de la geoestadística de Matheron [17], si suponemos que la imagen $f(x, y)$ es una variable regionalizada, para estimar la media de esta variable en el área $S$ cómo $\mu = 1/S \sum\sum f(x, y)$, la varianza estimada debe seguir la ley $\sigma^2(0/S) = 1/S^2C$ (la varianza deber variar inversamente con respecto al área al cuadrado). Para poder

verificar esta hipótesis, se han realizado algunos tests que implican el muestreo del fondo utilizando diferentes áreas de muestra y el cálculo de la varianza asociada a estas áreas (ver ejemplo numérico en la figura 7). Como conclusión, podemos afirmar que las variaciones locales del ruido de fondo a la escala del tamaño del spot no permiten una estimación global del ruido medio. La rejilla ortogonal de los spots proporciona una segmentación adicional: las regiones-cajas rectangulares $BS_i$ que son consideradas las regiones de influencia de los spots $S_i$. Estas regiones pueden ser usadas para cuantificar el ruido local asociado a cada spot. De forma práctica, una zona envolvente de protección (para evitar el sesgo introducido en la vecindad de la región del spot) es obtenida mediante el residuo de una dilatación de la región del spot; y el ruido $\check{n}_i$ se estima en la región $BS_i - \delta_n(S_i)$ que tiene un área de $\check{A}_i$ pixels (el tamaño típico de la dilatación es $n = 3$). La expresión global para la señal de intensidad viene dada por $s_i = \hat{s}_i - n_i = \hat{s}_i - \bar{N}_i A_i = \hat{s}_i - \frac{\check{n}_i}{\check{A}_i} A_i$.

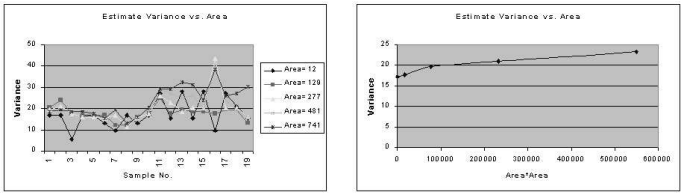



Figura 7: Muestreo del ruido de fondo utilizando diferentes áreas de muestra y estimación de la varianza: izquierda, valores estimados de la varianza y derecha, área al cuadrado con respecto a la varianza.

## 4 Conclusiones

En este artículo hemos presentado la utilización satisfactoria de operadores morfológicos para el análisis cuantitativo de la imagen en microscópica aplicado a la citología hematológica y a los *cDNA microarrays*.

El conjunto de algoritmos de la plataforma de hematología constituyen una herramienta global para el análisis e indexación de imágenes *teleslides* de frotis sanguíneos, la cual proporciona una descripción-clasificación objetiva y automática de células de la sangre; habiendo sido diseñada para asistir a los hematopatólogos en el diagnóstico de afecciones hematológicas. En una fase posterior, la extensión de nuestra plataforma al estudio de la malaria (paludismo) sería desarrollada; por ejemplo, la morfología de los eritrocitos (forma y color) y las inclusiones eritrocitarias arrojan un conocimiento fundamental para la detección del estadio y gravedad de la malaria. Los algoritmos se han basado en tests metodológicos en una amplia base de imágenes (más de 3000). Añadiendo capacidades telemáticas al sistema, es decir, integrándola en red, ver figura 8, pueden desarrollarse nuevas aplicaciones tal como: bases epidemiológicas, diagnostico colaborativo, tele-hematología, etc. Además, actualmente se está llevando a cabo una experiencia de tele-consenso, implicando la opinión de cuatro expertos, con el objetivo de validar los descriptores y los resultados de la clasificación [6].

La metodología descrita para el estudio de los *cDNA microarrays* permite resolver los problemas principales de estas imágenes: forma de los spots, fuerte ruido global de fondo, señales localmente débiles, etc. Como ya hemos dicho, nuestra propuesta es completamente automática y la segmentación de los spots consiste en una técnica adaptativa en posición, forma y tamaño. La evaluación experimental del resultado de la segmentación y cuantificación de los spots muestra que el uso de estos algo-

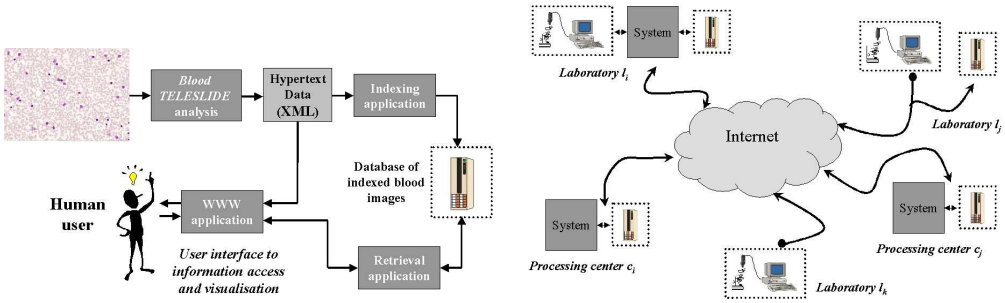



Figura 8: Aplicaciones de la plataforma de hematología: izquierda, Asistente para el diagnóstico y la consulta basada en el contenido morfológico de bases de datos epidemiológicas; y derecha, escenario en red (laboratios con o sin capacidad de tratamiento, bases de datos centralizadas y locales, capacidad de procesamiento distribuida, etc.)

ritmos son generalmente iguales o mejores que aquellos obtenidos mediante técnicas convencionales manuales, ver estudio comparativo en [2].

## References


[1] Angulo, J. and Serra, J. (2001) Semantic-based hierarchical segmentation of color blood smear images using mathematical morphology. *Internal Note N-50/01/MM, CMM-Ecole des Mines de Paris*, August 2001, 28 pp.

[2] Angulo, J. and Serra, J. (2002a) Automatic analysis of DNA microarray images using mathematical morphology. *Submitted to Bioinformatics*, January 14th 2002 (revised version on August 26th), 50 pp.

[3] Angulo, J. and Serra, J. (2002b) Morphological spectrum applied to genome arrays quantification. *Submitted to Journal of Visual Communication and Image Representation*, January 14th 2002, 20 pp.

[4] Angulo, J. and Flandrin, G. (2002c) Automated detection of working area of peripheral blood smears using mathematical morphology. *To appear in Analytical Cellular Pathology*, July 1st 2002, 35 pp.

[5] Angulo, J. and Serra, J. (2002d) Morphological color size distributions for image classification and retrieval. *Proc. of ACIVS'2002 (Advanced Concepts for Intelligent Vision Systems)*, Ghent, Belgium, pp. 46–53.

[6] Angulo, J., Benattar, L., Lusina, D., Martelli, J.M. and Flandrin, G. (2002e) Inter-observer consensus on lymphoid cell morphology, compared to obective measurement by mathematical morphology procedure. *In preparation.*

[7] Benattar, L. and Flandrin, G. (1999a) Morphometry and Quality Control for a May-Grünwald Giemsa stained preparation. A 40 centers cooperative study. *Leukemia and Lymphoma*, Vol. 33, pp. 587-589.

[8] Benattar, L. and Flandrin, G. (1999b) Comparison of the classical manual pushed wedge films, with an improved automated method for making blood smears. *Hematology and Cell Therapy*, Vol. 41, pp. 1-5.

[9] Brown, P. O. and Botstein, D. (1999) Exploring the New World of the genome with DNA microarrays. *Nature Genetics*, Vol. 21 (supplement), pp. 33–37.

[10] Comaniciu, D., Meer, P. and Foran, D. J. (1999) Image-guided decision support system for pathology. *Machine Vision and Applications,* Vol. 11, pp. 213–224.

[11] Chen, Y., Dougherty, E. R. and Bittner, M. L. (1997) Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Journal of Biomedical*

*Optics*, Vol. 2, pp. 364–374.

[12] Flandrin, G. (1995) Haematological cytology image bank and teletransmision for microscopic diagnosis. *Arch. Ana. Cytol. Path.*, Vol. 43, pp. 257-261.

[13] Flandrin, G. (1997) Image Bank, diagnostic codification and telediagnosis in hematology. *Leukemia and Lymphoma*, Vol. 25, pp. 97-109.

[14] Haus, H. M., Harms, H., Ter Meulen, V. and Gunzer, U. (1987) Statistical evaluation of computer extracted blood cell features for screening populations to detect leukemias. Pattern Recognition Theory and Applications. *Edited by P. A. Devjver and J. Kittler, Spring-Verlag Berling,* Heidelberg.

[15] Hirata JR, R., Barrera, J., Hashimoto, R. F. and Dantas, D. O. (2001) Microarray Gridding by Mathematical Morphology. *Proc. SIBGRAPI, International Symposium on Computer Graphics, Image Processing and Vision 2001*, Florianopolis, IEEE Computer Society, pp. 112–119.

[16] Klossa, J., Flandrin, G. and Hémet, J. (2002) Teleslide: Better slide representativeness for digital diagnostic microscopy applications. *6th Congress of European Group of Telepathology*, Crete, Greece, 14-21 September.

[17] Matheron, G. (1975) Random Sets and Integral Geometry. *Wiley*, New York.

[18] Meyer, F. (1986) Automatic Screening of Cytological Specimens. *Computer Vision, Graphics, and Image Processing*, Vol. 35, pp. 356-369.

[19] Seron, D., Moreso, F., Gratin, G. *et al.* (1996) Automated Classification of Renal Interstitium and Tubules by Local Texture Analysis and a Neural Network. *Analytical and Quantitative Cytology and Histology*, Vol. 18, pp. 410-419.

[20] Serra, J. (1982,1988) Image Analysis and Mathematical Morphology. Vol I and II. *Academic Press,* London.

[21] Theera-Umpon, N. and Gader, P. D. (2000) Counting white blood cells using morphological granulometries. *Journal of Electronic Imaging,* Vol. 9, pp. 170–177.

[22] Vesanen, P., Tiainen, M. and Yli-Harja, Olli (2002) On Calibration-Free Methods in Segmentation of cDNA Microarray Images. *IS&T/SPIE Symposium Image Processing: Algorithms and Systems. Proceedings of SPIE Vol. 4667*, 12 pp.

[23] World Health Organization (2001) Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues. *Edited by E. S. Jaffe, N. L. Harris, H. Stein and J. W. Vardiman, IARC Press*, Lyon, France, 2001.

[24] Yang, Y. H., Buckley, M. J., Dudoit, S. and Speed, T. P. (2002) Comparison of methods for image analysis on cDNA microarray data. *Journal of Computational and Graphical Statistics*, Vol. 11, pp. 108–136.