

APLICACIÓN DE ALGORITMOS PARA EL PROCESAMIENTO DE IMÁGENES A LA TECNOLOGÍA DE MICROARREGLOS PARA EL DIAGNÓSTICO.

Celia M. Laza Dueñas¹, Alfredo Rego Días, Niurka Carlos Pías.

Departamento Central de Matemática y Programación Aplicada. Centro de Inmunoensayo

. Calle 134 y Ave. 25, Apartado Postal 6653, Cubanacán, Playa, Ciudad de La Habana, Cuba.

ABSTRACT

Algorithms for the processing of Microarray images are presented in this work. A special emphasis is made on the stage of image segmentation presenting three segmentation methods: method of Court, method of Canny, and method of Segmentation for Region Growing (Seeded Region Growing). All of them are implemented with the corresponding adaptations to the Microarray technology for diagnosis.

Several examples of the application of these methods are shown, and a discussion of the results obtained is carried out by means of a brief comparison among them; considering the computational speed of each method and the proximity of the results obtained with the prospective results.

KEY WORDS: Microarray for diagnosis, Image Processing, Image Segmentation.

MSC: 62P10

RESUMEN

En este trabajo se presentan algoritmos para el procesamiento de imágenes de Microarreglos. Se hace un especial énfasis en la etapa de segmentación de la imagen y se presentan tres métodos de segmentación: método de Corte, método de Canny y método de Segmentación por Crecimiento de Regiones (Seeded Region Growing), implementados con las respectivas adaptaciones a la tecnología de Microarreglos para el diagnóstico.

Se muestran varios ejemplos de aplicación de dichos métodos y se realiza una discusión de los resultados obtenidos mediante una breve comparación entre ellos teniendo en cuenta la rapidez computacional de cada uno y la cercanía de los resultados obtenidos a los resultados esperados.

1. INTRODUCCIÓN

Los Microarreglos (MA) [8, 10, 14] están íntimamente ligados al desarrollo y miniaturización de las técnicas de afinidad empleadas en la Biología Molecular. Los primeros ensayos de afinidad con muestras fijas sobre soportes sólidos se realizaron con ensayos inmunológicos, inmovilizando sobre una superficie, antígenos o anticuerpos para su detección. Ejemplo de estos ensayos son los llamados Microarreglos para el Diagnóstico (MAD), sobre los que se dispensan varios antígenos de forma matricial y luego de hibridarlos con un material fluorescente, se pueden ver las reacciones entre estos y los anticuerpos presentes en las muestras de suero como resultado de alguna enfermedad [10].

Este tipo de tecnología es muy útil para el diagnóstico clínico y el monitoreo de estadios de varias enfermedades [11]. La evidencia de reacción entre los anticuerpos presentes en el suero de una muestra con los antígenos inmovilizados en la superficie del MA nos da la posibilidad, no sólo de determinar a que enfermedad resulta reactiva dicha muestra, sino también con cuales antígenos reacciona lo que permite aplicar terapias específicas en el momento de tratar la enfermedad.

En este trabajo se realiza un estudio de varios métodos para el procesamiento de imágenes [9] y se aplica a la tecnología de MAD. En esta dirección, hacemos un especial énfasis en la etapa de segmentación de la imagen [9] (clasificación de píxeles en píxeles Fondo y píxeles Señal). Se presentan tres métodos para segmentar una imagen de MAD, con las respectivas adaptaciones que responden a las características de los MAD desarrollados para la tecnología SUMA.

¹ matprogram2@cie.sld.cu.

Posteriormente se realiza una comparación preliminar de estos en cuanto a rapidez computacional y cercanía de los resultados obtenidos a los esperados, fundamentalmente, para tener idea de cómo se comportan los mismos una vez aplicados a las imágenes. Por último se presentan varias conclusiones obtenidas a partir de dicha comparación.

2. MATERIALES Y MÉTODOS

2.1 Imágenes usadas, software y etapas para el procesamiento de imágenes de MAD

Las imágenes de microarreglos usadas en este trabajo, fueron obtenidas en el Centro de Inmunoensayo. La Tabla 1 muestra el orden en que se dispensaron los antígenos correspondientes a las diferentes proteínas de los virus: HCV (virus de la hepatitis C), VIH (virus de la inmunodeficiencia humana), HTLV (virus linfotrópico de las células T humanas) y del parásito *Trypanosoma cruzi*, causante del mal de Chagas.

Tabla 1: Dispensado de un Microarreglo de dimensiones 8x7 según el orden en que se dispensaron los antígenos correspondientes a las diferentes proteínas de los virus: HCV (virus de la hepatitis C), VIH (virus de la inmunodeficiencia humana), HTLV (virus linfotrópico de las células T humanas) y del parásito *Trypanosoma cruzi*, causante del mal de Chagas.

HCV	HCV	HCV	HCV	Neg	IgG	IgG	IgG
Chagas	Chagas	Chagas	Chagas	HCV	HCV	IgG	IgG
Chagas							
VIH	VIH	VIH	VIH	VIH	VIH	Chagas	Chagas
VIH							
HTLV	HTLV	HTLV	HTLV	VIH	VIH	VIH	VIH
HTLV							

Se estudiaron 19 muestras de suero. De estas, cuatro eran negativas, tres positivas al HTLV, una positiva a Chagas, seis positivas al VIH, cuatro positivas a HCV y una positiva al HCV y al VIH.

Se utilizó un *software* desarrollado por el departamento de Matemática y Programación Aplicada del Centro de Inmunoensayo: MRS versión 1.0, diseñado con el objetivo de brindar una herramienta para la interpretación de los resultados por parte de los analistas y de los laboratorios donde se implemente la tecnología. Como se puede ver en la figura 1, tiene una interfaz que permite una fácil manipulación por parte de los usuarios y está preparado para escoger el método de segmentación a usar para procesar la imagen.

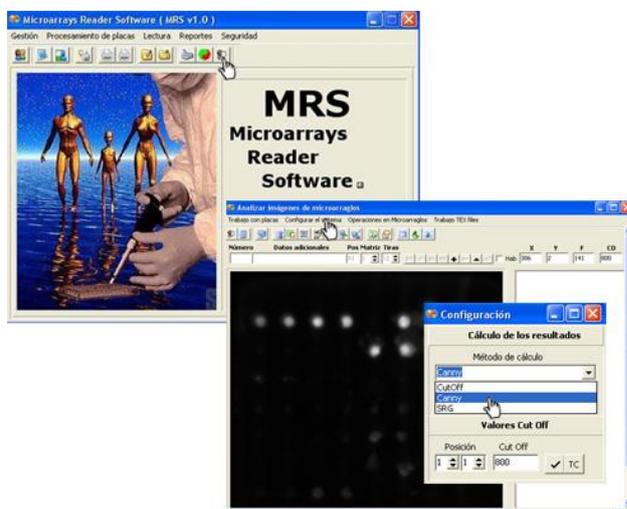


Figura 1: Ventana Principal del MRS versión 1.0 y Ventana para análisis de la imagen de MA.

El análisis de una imagen de MA esta dividida en tres etapas: ubicación de los puntos del MA, segmentación de la imagen y extracción de características [9, 16].

En la etapa de ubicación de los puntos del MA se realiza una asignación de coordenadas al centro de cada punto, o se escoge un elemento estructurante (círculos, rectángulos u otras forma libres) que los contengan. Esto debe realizarse de forma confiable, rápida y lo más automática posible [3].

En nuestro caso, para la asignación de coordenadas se utilizaron rectángulos como elementos estructurantes y un conjunto de puntos de referencias (puntos dispensados que siempre reaccionaran con la muestra de suero), conformándose una especie de rejilla, que se posiciona automáticamente sobre la imagen. Para ello se aprovechó el hecho de que en el proceso de dispensado se hace contacto sobre la placa de MA, dando lugar a que la mayor intensidad de cada gota caiga siempre en el centro del punto. Luego, para determinar las dimensiones de los cuadrantes se usó la distancia calculada entre los centros de los puntos de referencia y para darle ubicación a la rejilla sobre la imagen se usó el centro del punto de referencia ubicado en un extremo superior izquierdo del MAD. En la figura 2 se muestra el procedimiento antes descrito.

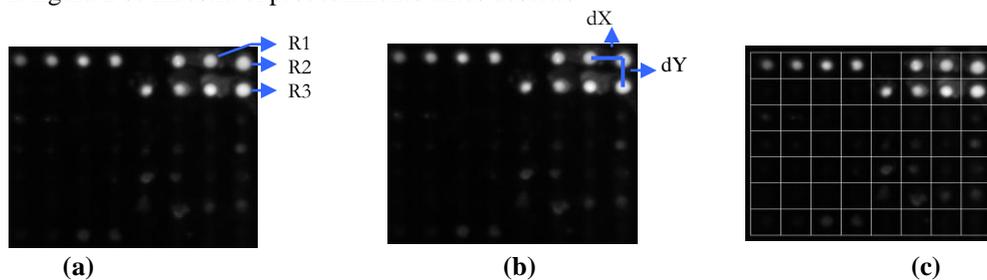


Figura 2: (a) Imagen de un Microarreglo para el Diagnóstico, en la que se señalan los puntos de Referencia. (b) dX = distancia en X entre el centro de la referencia R1 y el centro de la referencia R2 y dY = distancia en Y entre el centro de la referencia R2 y el centro de la referencia R3. (c) Imagen con la rejilla posicionada (dX , dY : dimensiones de cada cuadrante).

Luego de posicionada la rejilla, el proceso de segmentación y extracción de características se realiza dentro de cada cuadrante de la misma.

En la etapa de segmentación de la imagen se realiza la clasificación de los píxeles en fondo o señal. Los métodos de segmentación de imágenes están agrupados en cuatro tipos fundamentales: técnicas basadas en cotas, métodos basados en bordes, métodos basados en regiones y técnicas que combinan los métodos basados en bordes y regiones [9].

Los métodos de corte se basan en el postulado de que todos los píxeles cuyo valor de intensidad caiga en un cierto rango pertenecen a una clase determinada [13]. Para ello se usa un valor de corte o umbral que permite definir cuando un píxel pertenece a la señal o al fondo. Los métodos de bordes se basan en el postulado de que los valores de los píxeles cambian bruscamente alrededor de los bordes de los objetos [6, 7]. Permiten la detección de bordes en las imágenes, con una precisión entre el borde detectado y el real dependiente del grado de variabilidad de ciertos parámetros. Los métodos de regiones se basan en el postulado de que píxeles vecinos dentro de una región dada tienen valores de intensidad similar [1, 9]. El procedimiento general compara cada píxel con los de cierta región y si se satisface la condición de homogeneidad de interés se agrega a la misma. Parte de un conjunto de regiones semillas (representativas de la característica de interés) y a partir de estas comienza el crecimiento mediante la adición de píxeles [1]. Los métodos que combinan estos dos últimos incluyen segmentación morfológica [2, 9] y ajuste de superficie. Se aplica la transformada de Watershed [4] al gradiente de la imagen y se analiza la topografía de dicho gradiente con fronteras entre regiones como crestas.

Por último, tenemos la etapa de extracción de características en la que se calculan diferentes medidas de interés para las regiones segmentadas en la imagen: áreas, volúmenes, longitudes, promedios. Generalmente, para estos cálculos se suele usar las intensidades de los píxeles de la región a medir, pero pudieran usarse también la cantidad de píxeles de los puntos pertenecientes a una región dada [5, 15]. Por ejemplo, el área de un objeto de la imagen puede definirse a partir de la cantidad de píxeles que conforman dicho objeto. También puede definirse la distancia entre dos puntos de la imagen a partir de la cantidad de píxeles que conforman la línea recta que los une. Puede resultar también de interés evaluar la calidad de la segmentación a partir de ciertos parámetros conocidos de antemano y que deben cumplirse en la imagen segmentada. Por ejemplo, que un objeto tenga determinadas dimensiones o que este ubicado en determinado lugar dentro de la imagen.

En nuestro caso, utilizamos la media de las intensidades de los píxeles clasificados señal como valor para asignar a cada punto del MA y se rechaza todo punto que no cumpla con estar en un entorno de ± 10 píxeles respecto al centro del cuadrante, o con un área (en píxel) de valor menor que 50.

2.2 Métodos de Segmentación Implementados

Método de Corte: A partir de un valor de intensidad escogido, usualmente dado por un experto, se dice que el píxel es fondo si su valor está por debajo de la cota y es señal si está por encima. Este valor se le denomina cota o valor de corte.

La ventaja de este método es que es un método útil para realizar inspecciones preliminares sobre la imagen respecto a la determinación de zonas fluorescentes a partir de un valor de corte dado. Pero depende del criterio de expertos, el cual, sobre todo en las primeras etapas del desarrollo, es imposible definir con exactitud.

En la figura 3 se muestra un ejemplo de segmentación por Corte, aplicado a un punto dispensado de un MAD. Para este caso se usó un valor de corte de 800 u.

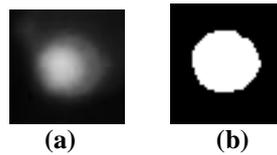


Figura 3: (a) Imagen de un punto de un Microarreglo para el Diagnóstico. (b) Imagen segmentada.

Método de Canny [6]: Esta basado en un proceso de optimización de ciertos objetivos tales como:

1. Maximizar la relación señal-ruido: mediante un filtrado de la imagen, la cual es suavizada usando un filtro Gaussiano.
2. Minimizar la diferencia borde real y borde detectado: para disminuir la diferencia entre uno y otro, y evitar falsas detecciones, se eliminan los puntos que no sean de máximos locales de la imagen filtrada.
3. Identificar bordes usando un conjunto de píxeles con cierta conectividad. Se construyen dos imágenes binarias a partir de ciertas condiciones y se usan complementariamente para encontrar un borde final.

Su ventaja principal es que es independiente del criterio de expertos, pero en la determinación del borde final pueden quedar tramos de bordes abiertos. Para solucionar este problema se usó la media de los píxeles bordes como valor de corte para definir los píxeles señal y fondo.

Algoritmo:

Sea $f(x,y)$ la intensidad de un píxel que se encuentra en la posición (x,y) de la imagen.

1. Aplicar un filtro Gaussiano a la imagen para suavizarla:

$$g(x, y) = f(x, y) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2 + x^2}{2\sigma^2}} \rightarrow \text{Suavizado de la imagen.}$$

2. Calcular el gradiente de g para maximizar la diferencia entre el ruido y la señal:

$$\nabla g(x, y) = \nabla \left[f(x, y) h_\sigma(x, y) \right] = f(x, y) \nabla h_\sigma(x, y) = f(x, y) \begin{pmatrix} \frac{\partial h_\sigma(x, y)}{\partial x} \\ \frac{\partial h_\sigma(x, y)}{\partial y} \end{pmatrix}$$

3. Se calcula el módulo y el argumento de la imagen suavizada y filtrada

$$G^{\text{mod}}(x, y) = \sqrt{g_x^2(x, y) + g_y^2(x, y)} \rightarrow \text{Módulo}$$

$$G^{\text{arg}}(x, y) = \tan^{-1} \left[\frac{g_y(x, y)}{g_x(x, y)} \right] \rightarrow \text{Argumento}$$

4. Y entonces en la dirección del gradiente, se eliminan aquellos píxeles que no sean de máximos locales del módulo. Ya que el operador gradiente, aplicado a un píxel (x,y) , devuelve un vector que indica la dirección de máxima variabilidad de la intensidad luminosa y su nivel de variación.
5. Se construyen dos imágenes binarias auxiliares: I_1 e I_2 a partir de dos umbrales $T_1 < T_2$, tales que, con T_1 habrían muchos más píxeles bordes que con T_2 .

$$I_{i_{(i=1,2)}} = \begin{cases} 1, & \text{si } G^{\text{mod}}(x, y) > T_i \\ 0, & \text{si } G^{\text{mod}}(x, y) < T_i \end{cases}$$

6. Entonces a partir de píxeles bordes de la imagen I1, se van construyendo bordes, mediante la adición de píxeles bordes de la imagen I2.
7. Calcular la media (M_b) de los píxeles que definitivamente fueron bordes.
8. Si $f(x,y) > M_b$, entonces el píxel (x,y) pertenece a la señal, si no pertenece al fondo.

En la figura 4 se muestra un ejemplo de segmentación por Canny aplicado a un punto dispensado de un MAD. Los valores de umbral usados fueron 0.3 y 0.7.



Figura 4: (a) Imagen de un punto de un Microarreglo para el Diagnóstico. (b) Imágenes I1 (borde fino) e I2 (borde grueso). (c) Imagen del borde final, obtenido mediante la agregación de píxeles de I2 a I1. (d) Imagen segmentada.

Método de crecimiento de regiones por agregación de píxeles (SRG: Seeded Region Growing) [1]: Dado un conjunto de píxeles generadores (semillas) se definen regiones, que crecerán agregando otros píxeles con propiedades similares. En nuestro caso la característica a considerar es la intensidad.

Los valores de máxima y mínima intensidad se usaron como semilla. El criterio de parada es que todos los píxeles hayan sido ubicados en alguna región. Este método también tiene la ventaja de ser independiente del criterio de expertos.

Algoritmo:

1. Crear conjuntos A_1, \dots, A_n a los que inicialmente solo se agregaran los píxeles semillas. Cada paso del algoritmo envuelve la adición de un píxel a alguno de los conjuntos A_i .
2. Definamos el conjunto de todos los píxeles que aún no han sido ubicados en ningún A_i y que son vecinos de las semillas: T.
3. Para cada elemento x del conjunto T, definimos una medida $\delta(z)$, que nos dará la mínima diferencia de x con respecto a todas las A_i :
$$\delta(x) = \min_y |g(x) - \text{mean}_{y \in A_i(x)} [g(y)]|$$
4. Ordenar los píxeles de T, de menor a mayor según $\delta(x)$.
5. Tomar el primer elemento de T, llámémosle z , agregarlo al conjunto A_i mas cercano y eliminarlo de T.
6. Guardar en T los vecinos de z que aún no han sido agregados a ningún conjunto, ni a T.
7. Repetir los pasos del 3 al 6 hasta que ya no queden píxeles en T.

En nuestro caso, sólo tenemos dos conjuntos: Señal y Fondo. Los píxeles escogidos como semillas son el de máxima y mínima intensidad respectivamente. A partir de ellos comienza el crecimiento de las regiones. Una vez concluido el procedimiento antes descritos quedan definidos los conjuntos de interés.

En la figura 5 se muestra un ejemplo de segmentación por SRG aplicado a un punto dispensado de un MAD.

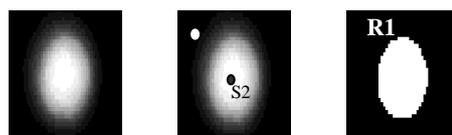


Figura 5: (a) Imagen de un punto de un Microarreglo para el Diagnóstico. (b) Semillas para el crecimiento de las regiones. (c) Imagen segmentada.

3. RESULTADOS Y DISCUSIÓN

En la figura 6 se muestran ventanas del MRS versión 1.0 con la aplicación del algoritmo para el procesamiento de imágenes de MAD. El método de segmentación seleccionado fue el de Canny y la imagen usada muestra la

reacción entre las proteínas del parásito causante del mal de Chagas presentes en el suero con los antígenos dispensados en la matriz. Se esperaba ver valores por encima de 800 unidades (nivel de corte dado por los analistas) para los aquellos correspondientes a dicha enfermedad (Ver Tabla 1).

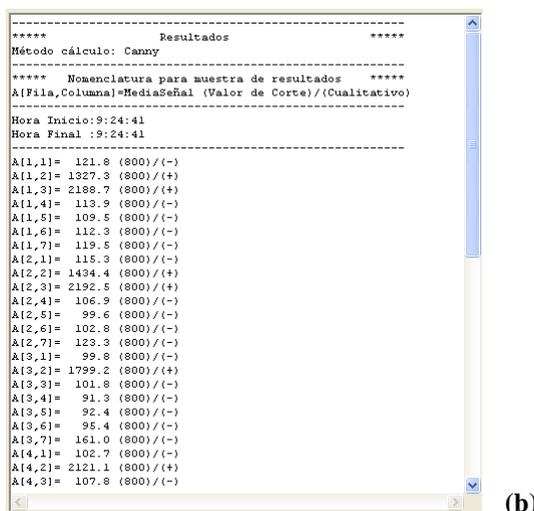
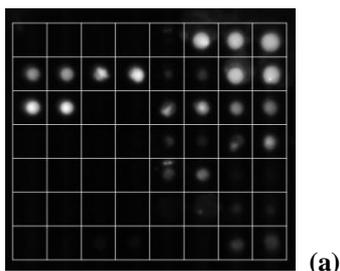


Figura 6: (a) Posicionamiento Automático de la Rejilla. (b) Resultados usando el método de Canny.

La figura 6 (a), muestra un posicionamiento automático de la rejilla bastante eficiente, en cuanto a ubicación de los puntos del MAD [12]. Dentro de cada cuadrante quedó recogido un único punto del MAD, lo que permite un mejor procesamiento a la hora de aplicar cualquiera de los métodos de segmentación. La figura 6 (b) muestra que se obtuvieron, efectivamente, valores por encima de 800 unidades para aquellos antígenos correspondientes a la enfermedad de Chaga. Este procedimiento fue aplicado a todas las imágenes de pruebas que se tenían.

A continuación se muestra una tabla que resume los resultados que ofrece cada método, según el ensayo, teniendo en cuenta el por ciento de imágenes en las que dio un resultado cercano a lo que se esperaba para aquellos antígenos de los ensayos que debían reaccionar según la muestra y bajos para el resto. Dicha tabla también recoge el tiempo promedio de ejecución de cada método, dado que es de interés también que este sea bajo para garantizar el trabajo cuando se tienen grandes volúmenes de placas de MAD, como sucede en muchos laboratorios.

Tabla 2: Resumen de resultados por cada método en por cientos y tiempo medio de ejecución de los mismos en segundos.

Método	Negativo	VIH	HTLV	HCV	Chagas	Tiempo Medio (seg)
Corte	66	100	100	100	100	0.2
Canny	100	100	100	100	100	0.4
SRG	100	100	100	100	100	0.9

Según estos resultados podemos concluir que el método de Corte es un método computacionalmente rápido, sólo que, es muy importante tener un buen criterio de cota para la clasificación de los píxeles, sobre todo para imágenes correspondientes a muestras negativas. A pesar de ello, este le permite al diseñador del ensayo realizar una inspección preliminar de la imagen. El método SRG presenta una leve demora, comparado con los otros dos, la cual conspira con la necesidad de trabajo con grandes volúmenes de muestras en un laboratorio pero da

resultados cercanos a los esperados en un 100 por ciento de los casos. El método de Canny parece ser el más eficiente de los tres pues combina rapidez computacional y proximidad de resultados a los resultados esperados.

4. CONCLUSIONES

En este trabajo se mostraron algoritmos para el procesamiento de imágenes de MAD. Los mismos fueron aplicados a varias imágenes de las que se sabía las características del dispensado y a que enfermedad era reactiva la muestra evaluada.

Los resultados obtenidos muestran que los métodos implementados, de manera general, responden a la necesidad de obtener un resultado lo mas cercano posible a lo esperado y con una velocidad computacional aceptable, aunque alguno tiene mejor comportamiento que otro según sea el parámetro que se tome en cuenta.

RECEIVED MARCH 2010

ACCEPTED JUNE 2010

REFERENCIAS

- [1] ADAMS, R. and BISHOP, L. (1994): Seeded Region Growing, **IEEE PAMI**, 16, 641-647.
- [2] ANGULO, J. and SERRA, J. (2003): Automatic analysis of DNA microarray images using mathematical morphology, **Bioinformatics**, 19, 553-562.
- [3] BAJCSY, P. (2004): Gridline: automatic grid alignment in DNA microarray scans, at URL: <http://alqdocs.ncsa.uiuc.edu/PB-20040101-1.pdf>, Ultima consulta 01-2010.
- [4] BEUCHER, S. and MEYER, F. (1992): The morphological approach to segmentation: the watershed transformation, **Mathematical Morphology in Image Processing**, Marcel-Dekker, New York.
- [5] BUHLER, J., IDEKER, T. and HAYNOR, D. (2000): Dapple: Improved Techniques for Finding Spots on DNA Microarrays. **UV CSE Technical Report UWTR 2000-08-05**.
- [6] CANNY, J.F. (1986): A Computational Approach to Edge Detection, **IEEE PAMI**, 8, 679-698.
- [7] DAVIS, L.S. (1975): A survey of edge detection techniques, **Comput. Graphical Image Processing**, 4, 248-270.
- [8] FRIEND, S.H. and STOUGHTON, R.B. (2002): **The Magic of Microarray**, Scientific American.
- [9] GONZÁLES, R.C. and WOODS, R.E. (2002): **Digital Image Processing**, Prentice Hall, New Jersey.
- [10] GUIDO, L. and MANRIQUE, C. (2005): Microarreglos: herramienta para el conocimiento de las enfermedades, **Revista Colombiana de Reumatología**, 12, 263-267.
- [11] HALL, D.A., PTACEK, J. and SNYDER, M. (2007): Protein Microarray Technology, **Mechanisms of Ageing and Development**, 128, 161-167.
- [12] KATZER, M., KUMMERT, F. and SAGERER, G. (2002): Robust Automatic Microarray Image Analysis. **Proceedings of the International Conference on Bioinformatics**, North-South Networking, Bangkok.
- [13] Sahoo, P.K., SOLTANI, S. and WONG, A.K.C. (1988): A survey of thresholding techniques, **Comput. Graphical Image Processing**, 41, 233-260.
- [14] SHENA, M. (2000): **Microarray Biochip Technology**, Eaton Publishing, Eaton.
- [15] WANG, X., GHOSH, S. and GUO, S.W. (2001): Quantative quality control in microarray image processing and data acquisition, **Nucleic Acids Research**, 29, 2-8.
- [16] YANG, Y.H., BUCKLEY, M.J., DUDOIT, S. and SPEED, T.P. (2002): Comparison of methods for image analysis on cDNA microarray data, **J. Comput. Graphical Statistics**, 11, 108-136.