

Análisis automático de la posición y del número de plásmidos en una célula bacteriana

P. González-Aledo⁽¹⁾, L. Vielva⁽¹⁾, J. Vía⁽¹⁾, C. Guynet⁽²⁾, J.C. Meile⁽³⁾ y F. de la Cruz⁽²⁾
{pablo.aledo,LuisVielva,jvia}@gtas.dicom.unican.es, {C.Guynet, J.Francois,fernando.cruz}@uncan.es

⁽¹⁾Dpto. de Ingeniería de Comunicaciones, Universidad de Cantabria.

⁽²⁾Dpto. de Biología Molecular, Universidad de Cantabria

⁽³⁾Laboratoire de Microbiologie et génétique Moleculaire, UMR S100 CNRS_UPS, Toulouse, France

Abstract—Plasmids are small chromosomes with the ability to pass from one cell to another. Recent advances in the study of its behavior shows that one of the main ways to assure its long-term survival is to adapt themselves to the cellular division mechanism of the host cell, assuring that at least one copy of the plasmid remains on each of the replicas generated in the cell-division process. Usually these studies have been made by hand by analyzing the images of a microscope and measuring the distance between each plasmid and the centre of the cell in which it is immersed. However, this manual process is slow and tedious, and reduces the applicability of the method to large image data sets. This communication shows an automated tool for cell segmentation based on a pre-emphasis filtering and parametric curve fitting that is able to accurately segment cells in microscopic images, count the plasmids inside it and extract its positions inside the cell.

I. INTRODUCCIÓN

Los plásmidos son pequeños cromosomas (moléculas de DNA) presentes en las bacterias y que son capaces de pasar de una célula a otra. Así propagan la resistencia a los antibióticos y otras características adaptativas de las bacterias. Recientemente se han podido visualizar las moléculas individuales de plásmidos en una célula añadiéndoles un sitio de unión para una proteína fluorescente. Así, la molécula de plásmido se puede ver como un pequeño punto fluorescente dentro de la célula. Esto permite no sólo contar el número de plásmidos de una célula, sino también determinar su posición dentro de la misma. Estos estudios son importantes para determinar cómo se distribuyen las moléculas de DNA en las células, cómo reorganizan sus posiciones en determinados momentos del ciclo celular (por ejemplo durante la conjugación o la división celular), y cómo su posicionamiento es alterado por determinadas mutaciones bien en el propio plásmido o en el cromosoma principal de la bacteria [1]-[5].

Básicamente existen tres estrategias utilizadas por los plásmidos para asegurar su supervivencia en las células hijas tras el proceso de división celular; la más evidente pero también la más costosa es generar un elevado número de copias de modo que exista una alta probabilidad de estar presente en las hijas; la segunda es sintetizar un veneno de largo plazo y un antídoto de corto plazo que asegure que todas las células en las cuales no existe un plásmido capaz de generar el antídoto mueran y por tanto aumente el porcentaje de células con huesped; la tercera y la más compleja es adaptarse al

mecanismo de división de la propia célula, de forma que cuando ésta se divida, el plásmido realice la misma acción, disponiéndose cada réplica de forma uniforme a lo largo del eje central de la célula y de modo que al dividirse ésta, se mantenga un plásmido por cada célula escindida. Un análisis cuidadoso de la posición de los plásmidos, en el que se pueda observar su acoplamiento al mecanismo de división celular del hospedador, permite por tanto detectar si es éste el método que está siendo empleado para asegurar la existencia a largo plazo de los mismos. En la mayoría de los casos, este análisis se realiza de forma manual. Esta tarea suele resultar tediosa para un técnico de laboratorio, lo cual limita habitualmente el número de medidas tomadas, la difusión de las mismas, y la replicabilidad de los experimentos. Este trabajo describe una herramienta automática para la segmentación de este tipo de imágenes.

El uso de contornos activos para la segmentación de células ha demostrado ser una técnica muy útil en los últimos años [6]-[8]. Presenta sin embargo el problema de requerir imágenes bimodales (en blanco y negro) para discernir entre la célula y el fondo de imagen. En la mayoría de las técnicas de microscopía actuales, sin embargo, el contorno de las células suele presentarse difuso y difícil de obtener con precisión. La agrupación de varias células solapándose, las diferentes formas que estas pueden tomar al no tratarse de cuerpos rígidos y el desenfoque producido por la óptica del microscopio dificultan notablemente la extracción de medidas fiables de forma automática.

II. METODOLOGÍA

La herramienta que se describe en esta comunicación para la localización y conteo de plásmidos en las células bacterianas se divide en tres etapas:

- 1) Filtrado de énfasis: A la imagen obtenida por el sistema de adquisición se le aplican varios filtros que resaltan la posición de las células y definen los bordes entre ellas.
- 2) Se obtiene una primera aproximación de los parámetros que definen cada célula. La finalidad de estos primeros resultados es el servir de punto inicial para iterar en el siguiente paso
- 3) Se ajustan los contornos obtenidos inicialmente para dar una solución más exacta que la proporcionada por el segundo paso

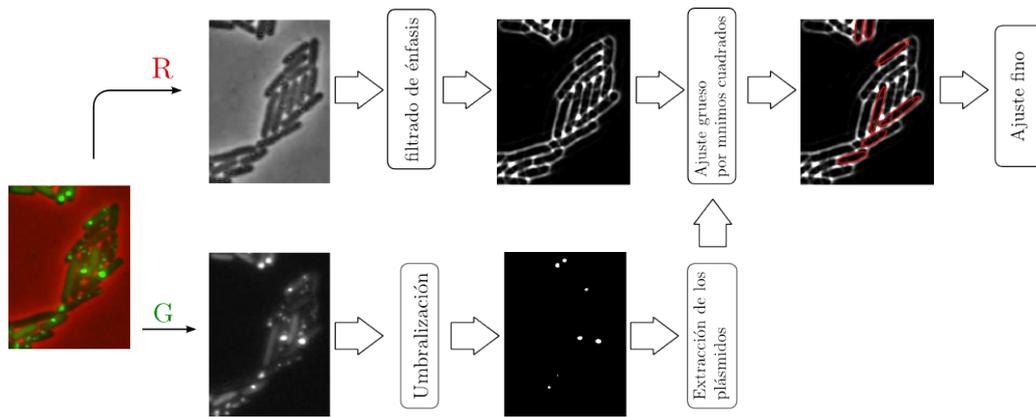


Fig. 1. Proceso de extracción de medidas a partir de la imagen. De la imagen original extraemos los canales verde y rojo. El primero de ellos se puede emplear directamente para obtener la posición de los plásmidos, cuyo brillo resalta notablemente en dicho canal. Dichos plásmidos se emplean como semilla para, a partir de ellos y de una imagen filtrada del canal rojo, obtener un contorno inicial de las células que nos interesan. Este contorno es finalmente refinado para obtener las células y los datos finales del proceso.

A. Filtrado de énfasis

Como ya se ha comentado, entre las dificultades más importantes de la segmentación automática de células se encuentra el desenfoque producido por la óptica del microscopio, que dificulta la detección de los bordes de las células, especialmente cuando estas se encuentran cercanas o solapadas.

Para solventar este problema se propone emplear el conocimiento *a priori* de que disponemos sobre la forma de la célula. Sabiendo el ancho aproximado de la misma podemos ajustar un filtro que convolucionado con la imagen resuene a la frecuencia que determina este ancho.

El concepto de filtrado se entiende en procesado de imagen como la convolución de dos señales bidimensionales y un filtro de convolución como una matriz $K(x, y)$ de dimensiones (M, N) generalmente mucho más pequeñas que la de la imagen.

El kernel K determina las propiedades del filtro. Mediante la aplicación de diferentes kernels se puede desenfocar una imagen, enfocarla, extraer sus bordes, clasificar sus píxeles ...

En este estudio se emplea como kernel para el filtrado una familia de funciones de Gabor. Esta familia de funciones se define como una función armónica multiplicada por una gaussiana:

$$\phi_j(\vec{x}) = \frac{k_j^2}{\sigma^2} \exp\left(-\frac{k_j^2 x^2}{2\sigma^2}\right) \left\{ \exp(i\vec{j}_j \vec{x}) - \exp(-\frac{\sigma^2}{2}) \right\} \quad (1)$$

siendo $\vec{x} = x + jy$, k_j un vector director dado por $(k_j \cos(\theta), k_j \sin(\theta))$ y σ la varianza de la función gaussiana. La representación mostrada, como se puede ver en la ecuación es una representación en números complejos del kernel. En este caso emplearemos la parte real del mismo cuya representación bidimensional se muestra en la figura 2 para varias frecuencias y orientaciones. Los filtros de Gabor son ampliamente utilizados en el tratamiento de imagen, donde se han mostrado eficientes a la hora de realizar diferentes

tareas, tales como segmentación de texturas, compresión, clasificación, extracción de puntos clave...

El uso de los kernels de Gabor viene justificado por su capacidad de representar variaciones sobre la imagen tanto en el espacio frecuencial como en el espacial. Transformaciones puramente frecuenciales como la fft consideran la imagen en su totalidad, siendo útiles para la descripción global de la misma pero no para la descripción localizada en torno a un punto. Por otra parte, operaciones como el desenfoque son operaciones puramente espaciales y por tanto no se puede modelar con ellas las propiedades frecuenciales de la imagen. Mediante el filtrado con esta familia de kernels, podemos obtener una representación de las propiedades frecuenciales de la imagen en torno a un punto dado de la misma. Si elegimos como frecuencia del filtro (k_v) el ancho medio de las células a analizar, al filtrar la imagen con dicho filtro se resaltarán los bordes de la misma que se den con una frecuencia próxima a la seleccionada. Podemos emplear este efecto para resaltar los bordes de las células. Un ejemplo de procesado de una imagen con este método se muestra en la figura 2.

B. Ajuste adaptativo del histograma

Una vez se ha procesado la imagen inicial, se deben de ajustar los niveles de gris de la misma. Este ajuste debe ser adaptativo, considerando que no conocemos *a priori* la densidad de las partículas que vamos a encontrarnos en la imagen, por lo que no se puede elegir un umbral fijo. El algoritmo canónico para binarizar una imagen en escala de grises a una imagen binarizada en blanco y negro es el algoritmo de Otsu [9]. Este algoritmo emplea un contexto estadístico para elegir el umbral óptimo que mejor divide el histograma de la imagen. Para ello se divide este histograma en dos segmentos; el primero incluyendo desde un nivel de intensidad 0 hasta el umbral seleccionado y el segundo desde el umbral hasta el nivel máximo de intensidad de la imagen. El histograma total se modela entonces como la suma de dos funciones gaussianas, cada una de ellas obtenida a partir de

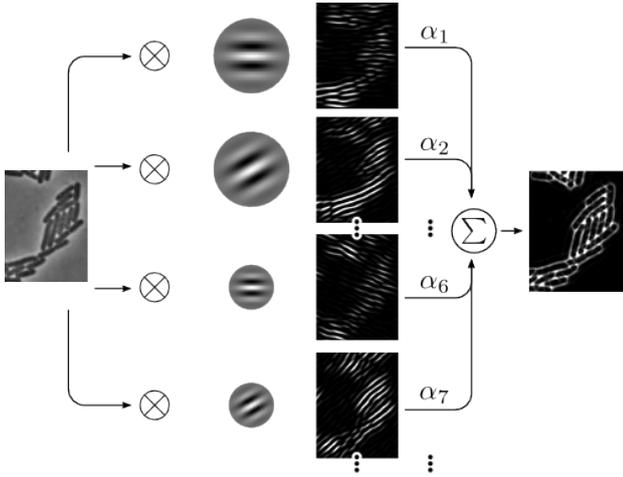


Fig. 2. Filtrado de una imagen de prueba con una familia de kernels de Gabor. El canal rojo es convolucionado con varios kernels de Gabor y las imágenes resultantes, tras ser ponderadas con unos coeficientes de peso α_i se combinan formando la imagen de salida del banco de filtros.

uno de los segmentos. Variando el umbral u , variaremos dichas distribuciones. El objetivo a alcanzar es la minimización de la varianza de cada una de las dos gaussianas, con lo que el error de umbralización se reduce. Para ello se define la función de coste como:

$$C(u) = P_0(u)\sigma_0^2(u) + P_1(u)\sigma_1^2(u) \quad (2)$$

donde $P_0(u)$ es el número de píxeles por debajo del umbral u , σ_0 la varianza de la primera gaussiana y P_1, σ_1 sus homólogos por encima del umbral. Definida esta función de coste, el algoritmo itera sobre los diferentes umbrales u hasta encontrar un mínimo.

C. Aproximación de los parámetros

En el siguiente paso se va a ajustar una curva paramétrica a la imagen obtenida en los pasos anteriores. Este ajuste es un paso complicado, puesto que la curva suele modelarse con un número elevado de parámetros, y el espacio de ajuste suele presentar máximos locales. Para alcanzar un máximo global en el siguiente paso del proceso es necesario el encontrar una aproximación a los parámetros que van a modelar la curva que finalmente ajustemos sobre la imagen. Dada la forma general de las células que vamos a tratar, se ha elegido un ajuste de puntos dado por una elipse. Para obtener los candidatos de la frontera entre la célula y el fondo, se trazan varias rectas partiendo del plásmido en ángulos equiespaciados. El punto en que estos rayos “tocan” el contorno de la célula se establece a partir de la imagen binarizada anterior, cuando la intensidad de la imagen bajo el rayo es máxima. Del primer conjunto así obtenido se eliminan los *outliers*; aquellos puntos que se separan más de una distancia prefijada de sus vecinos. Esta eliminación se repite una vez más y sólo se consideran los puntos que se mantienen después de estas dos iteraciones. En la figura 3 se muestran en diferentes colores las diferentes iteraciones del proceso. Una vez obtenidos los candidatos se

ajusta la elipse y de ella se extraen los parámetros iniciales para el siguiente paso. En este caso extraemos la posición del centro de la elipse, la longitud de su eje mayor y su orientación.

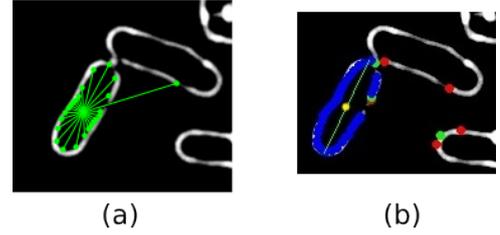


Fig. 3. Eliminación sucesiva de los outliers. A partir de la posición inicial dada por un plásmido, se lanzan rayos en varias direcciones para obtener un contorno aproximado de la célula que lo hospeda. De los puntos iniciales de dicho contorno se eliminan aquellos que se alejan una distancia mayor de un umbral de sus vecinos. Este proceso se realiza dos veces para finalmente ajustar los puntos restantes con una elipse. (Para mejorar la visualización de la imagen no se muestran todas las direcciones en las que se trazan los rayos.)

D. Ajuste final de la curva

Llegados a este punto, se dispone de una imagen en la cual se han enfatizado las diferentes células y una primera estimación de los parámetros que describen la curva que representa la célula. En torno al máximo global hay sin embargo varios máximos locales que representan soluciones válidas aunque no óptimas. Para ajustar finalmente la curva a la imagen se toma como punto inicial los parámetros obtenidos en el apartado anterior. Iterativamente se muestrea el espacio de búsqueda en un entorno próximo a ellos, y si la función de coste es mejor que el mejor resultado obtenido anteriormente, se establecen los parámetros así obtenidos como siguiente punto alrededor del cual iterar. De este modo, la solución converge progresivamente hacia el máximo global. Para este paso se toma como función de coste la suma de las intensidades de los píxeles bajo la curva dividido entre la longitud de la misma. Para evitar alejarnos excesivamente de los parámetros iniciales, se han considerado dos criterios:

- Se centra una gaussiana que modula el coste en torno a los parámetros iniciales calculados en el paso anterior. Las varianzas de dicha gaussiana han sido obtenidas en base a medidas experimentales.
- Cada iteración del método se reduce progresivamente el espacio a explorar. Esta reducción está controlada por una función exponencial decreciente.

Como parámetros de la curva se han tomado θ, l, w, x, y , cuyo significado se muestra en la figura 4.

III. RESULTADOS

Las técnicas presentadas anteriormente se han implementado en C con ayuda de la biblioteca OpenCV. Los resultados que se han obtenido se describen en este apartado.

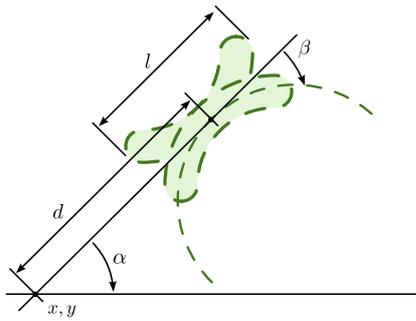


Fig. 4. Parametrización de la curva para ajustar a una célula

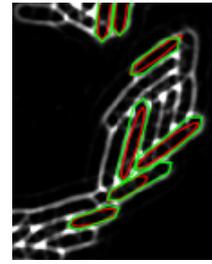


Fig. 6. Primera aproximación y segmentación final de las células

A. Imágenes de entrada

En este trabajo se analiza la distribución celular de las moléculas del plásmido R388 y de algunos mutantes derivados de éste que no logran separarse bien durante la división celular. Las imágenes de entrada han sido obtenidas separando la cadena de DNA del plásmido e introduciendo un sitio de unión para una proteína fluorescente.

Mediante un microscopio conectado a una cámara electrónica se obtienen imágenes de 1392×1040 , con una profundidad de color de 24 bits por píxel. La proteína empleada para destacar los plásmidos hacen que estos destaquen con un color verde sobre las imágenes. Las células también adquieren parte de esta tonalidad, aunque con una intensidad menor.

Tomaremos una zona de 219×298 como ejemplo de imagen para mostrar los diferentes pasos del proceso. Esta imagen se ve en la figura 1

La figura 5 muestra la diferencia entre emplear un nivel fijo de umbralización para segmentar la imagen, obtenido mediante el algoritmo de Otsu, el resultado de emplear el pre-filtrado de varios kernels de Gabor pero eligiendo un umbral fijo para binarizar la imagen (dicho umbral se ha tomado como la media de los umbrales óptimos para varias imágenes diferentes), y el obtenido empleando el filtrado de Gabor y el umbral de Otsu para binarizar.



Fig. 5. Diferentes algoritmos de umbralización

Una vez se ha obtenido la imagen binarizada, se muestra en la figura 6 las elipses que mejor definen un conjunto de células, junto con la segmentación final de las mismas.

En este punto del proceso cada célula ha sido asociada a un plásmido mediante el proceso de trazado de rayos y ajuste de

una elipse que se ha descrito anteriormente. Para los estudios analizados se requiere no obstante el ser capaces de obtener la asociación inversa; el decir qué plásmidos yacen dentro de cada célula. Un último paso de post-procesado de los datos busca entonces para cada elipse qué plásmidos se encuentran dentro de ella considerando que un punto está dentro de una elipse si la suma de su distancia a los dos focos de la elipse es menor que su eje mayor.

IV. CONCLUSIONES

La gran cantidad de datos producida por estudios biológicos en forma de imágenes requiere de algoritmos fiables que sean capaces de extraer de dicha información en forma de píxeles medidas fiables de los organismos que representan. Las mayores dificultades al respecto en el caso que nos ocupa son la agrupación de varias células, el desenfoque y las deformaciones que se producen en la célula por la presión de otras circundantes. Esta comunicación presenta un sistema automático de segmentación que emplea un filtrado de énfasis para resolver el problema de la agrupación y el desenfoque y una aproximación evolutiva para el ajuste de una curva paramétrica a las imágenes de entrada.

REFERENCES

- [1] de la Cruz, F. and J. Davies (2000). "Horizontal gene transfer and the origin of species: lessons from bacteria." *Trends Microbiol* 8(3): 128-33.
- [2] Fernandez-Lopez, R., M. P. Garcillan-Barcia, et al. (2006). "Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution." *FEMS Microbiol Rev* 30(6): 942-66.
- [3] Gordon, G. S., D. Sitnikov, et al. (1997). "Chromosome and low copy plasmid segregation in *E. coli*: visual evidence for distinct mechanisms." *Cell* 90(6): 1113-21.
- [4] Li, Y. and S. Austin (2002). "The P1 plasmid is segregated to daughter cells by a 'capture and ejection' mechanism coordinated with *Escherichia coli* cell division." *Mol Microbiol* 46(1): 63-74.
- [5] Sengupta, M., H. J. Nielsen, et al. "P1 plasmid segregation: accurate redistribution by dynamic plasmid pairing and separation." *J Bacteriol* 192(5): 1175-83.
- [6] Xiaoxu Wang Weijun He Metaxas, D. Mathew, R. White, E. "Cell segmentation and tracking using texture-adaptive snakes", *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 12-15 April 2007*, pp 101 - 104.
- [7] F.Sadeghian, Z.Seman, A.Rahman Ramli, B.Hisham Abdul Kahar. M.Saripan, "A Framework for White Blood cell segmentation in Microscopic Blood images using digital image processing", *Biological Procedures Online, Volume 11, Number 1 / December, 2009*.
- [8] Joost Vromen, Brendan McCane, "Red Blood Cell Segmentation from SEM Images", *Conference on Image and Vision Computing New Zealand, Wellington, November 2009*.
- [9] Otsu, N., "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 1, 1979*, pp. 62-66