

# Segmentación e identificación automática del *Mycobacterium tuberculosis* en imágenes de esputo\*

Manuel Guillermo Forero\*\*

## Automatic segmentation and identification of the *Mycobacterium tuberculosis* in sputum images

### RESUMEN

La tuberculosis es una grave enfermedad cuyo control está basado en el diagnóstico presuntivo. Un método frecuentemente utilizado para su diagnóstico es el análisis del esputo con el objetivo de detectar el *Mycobacterium tuberculosis*. El examen del esputo ocupa una gran cantidad de tiempo y se requiere un buen entrenamiento del especialista para evitar cometer un gran número de errores. Las técnicas de procesamiento de imágenes pueden ser de gran ayuda para realizar un examen. Así, se presenta una nueva técnica que pretende mejorar la precisión y disminuir el tiempo empleado en el análisis de muestras de esputo. Esta técnica emplea el conocimiento lingüístico acerca de las características de los bacilos, utilizando la información de color para la segmentación y un árbol de clasificación para la identificación de los bacilos con el fin de establecer si una muestra es positiva o negativa.

### PALABRAS CLAVE

Segmentación en color, procesamiento de imágenes, microscopía de fluorescencia, reconocimiento de patrones, tuberculosis

### ABSTRACT

Tuberculosis is a very serious disease whose control is based on early diagnosis. A method frequently employed in its diagnosis consists of the sputum analysis in order to detect the *Mycobacterium tuberculosis*. The sputum examination demands a great amount of time and a good training of the specialist is required to avoid to commit a great numbers of errors. Image processing techniques can be helpful in examinations. Thus, this paper presents a new technique, it tries to improve the precision and diminish the time used in the analysis sputum samples. This techniques uses the linguistic knowledge about the characteristics of the bacilli, using the color information for segmentation and a classification tree for bacilli identification to establishing if a sample is positive or negative.

### KEY WORDS

Color segmentation, fluorescence microscopy, image processing, pattern recognition, tuberculosis

\* Las gráficas de este artículo tienen la mejor calidad que se pudo lograr, a partir del archivo suministrado por el autor (nota del editor).

\*\* Instituto de Óptica Daza de Valdés CSIC, Serrano 121 28006, Madrid, España. Profesor adscrito Departamento de Ingeniería de Sistemas de la Universidad Nacional de Colombia, correo electrónico: mforero@optica.csic.es.

## INTRODUCCIÓN

El *Mycobacterium tuberculosis* es el causante de la tuberculosis pulmonar, aunque puede llegar a infectar otros órganos o tejidos como el cerebro, los riñones, los huesos y la piel. La tuberculosis es la principal causa de muerte por enfermedad infecciosa en el mundo. Según un estimativo de la Organización Mundial de la Salud (OMS), un tercio de la población mundial (1722 millones de personas) está infectado por esta bacteria. Anualmente se producen 10 millones de casos nuevos de tuberculosis activa en todo el mundo y fallecen 3 millones de personas. Algunos factores que han favorecido el incremento de la incidencia de esta enfermedad son las infecciones con el virus VIH (en el mundo, la tuberculosis es la causa principal de muerte entre personas infectadas con el VIH) y el aumento de la resistencia del bacilo a las drogas.

Además de la sospecha clínica, el diagnóstico de la micobacteriosis debe ser hecho a través del cultivo de muestras de esputo. La identificación del *Mycobacterium tuberculosis* se logra rutinariamente usando técnicas fluorocromáticas en muestras de esputo tinturadas con auramina fluorescente empleando un microscopio de fluorescencia. Estas técnicas carecen de sensibilidad, por lo cual los especialistas deben esperar los resultados de los cultivos, los cuales pueden demorar hasta 2 meses debido al hecho de que este bacilo requiere de 5 a 20 horas para duplicarse. El análisis manual para la identificación de los bacilos involucra un trabajo intensivo en el cual se presenta una rata alta de falsos negativos [1]. El estudio automático puede conllevar varias ventajas; por ejemplo la reducción sustancial en el trabajo de los especialistas, mejorando la sensibilidad de la prueba y una mayor precisión en el diagnóstico, e incrementando el número de imágenes que pueden ser analizadas.

El proceso de segmentación de las bacterias es complejo pues su forma no es suficiente para usarla como característica discriminante, debido a que otros bacilos, especies y partículas tienen la misma morfología. Por este motivo el proceso de segmentación requiere mejorarse empleando la información dada por el color de la bacteria.

Se han realizado algunos trabajos sobre la segmentación de bacterias. Veropoulos *et al.* [1,2] desarrolla-

ron un método de identificación basado en la forma de los descriptores y en clasificadores por redes neuronales, obteniendo una sensibilidad (rata entre las decisiones positivas encontradas como tales contra el número total de casos positivos) de 94.1%. Wilkinson [3] propone una técnica de segmentación de multiresolución rápida basada en la obtención de diferentes umbrales para diferentes zonas de una imagen monocromática. Otros autores usan la información de color como factor discriminante clave en la segmentación e identificación de bacterias [4,5] o la segmentación de células en el diagnóstico del cáncer de pulmón [6,7].

Como una continuación de los trabajos previos [8,9], en este artículo se presenta una nueva técnica para la segmentación e identificación de bacterias de la tuberculosis con el fin de establecer una muestra como positiva o negativa. El método de segmentación incluye la información de color; la técnica de identificación está basada en el uso de varios descriptores, empleando el conocimiento heurístico acerca de su forma.

## MATERIAL

El proceso de tinción de las muestras fue realizada con especímenes clínicos tanto respiratorios como no respiratorios, exceptuando la orina y la sangre. Estos especímenes fueron teñidos con auramina fluorocromada O y escaneados usando un microscopio de fluorescencia con una magnificación x250. Debido a los artefactos que pueden presentarse en las muestras, es necesario revisar cuidadosamente la morfología de las células.

La confirmación de las muestras fluorocromas positivas fueron hechas observando el crecimiento del *Mycobacterium tuberculosis* en cultivos de especímenes en medios sólidos y líquidos. Las tomas de las muestras fueron analizadas con un fotomicroscopio Zeiss Axiophot iluminado con un sistema de iluminación fluorescente de intensidad variable Zeiss Attoarc. La adquisición de las imágenes fue hecha con una cámara digital Coolsnap de Photometrics.

Para el desarrollo de la investigación se recolectaron 397 imágenes negativas y 75 positivas. Para realizar la descripción de los bacilos fueron utilizadas 110 muestras de bacilos, encontradas en 15 de las imágenes positivas; las 60 restantes fueron empleadas para la verificación del método de identificación de bacilos.

## MÉTODOS

Si el *Mycobacterium tuberculosis* aparece en una muestra, presenta una fluorescencia verdosa que puede llegar a ser amarilla y blanca. Estas bacterias miden entre 1 y 10 micrómetros de largo, y 0.2 y 0.6 micrómetros de ancho aproximadamente, pudiendo aparecer curvadas o rectas. Individualmente los bacilos pueden mostrar zonas fuertemente teñidas y áreas de tinción alternada que le dan una apariencia moteada. Esta información es importante en el desarrollo de la segmentación y el proceso de identificación, como se ve más adelante. Vale la pena anotar que para la correcta identificación de los exámenes, no nos centramos en la detección de todos los posibles bacilos, sino en la correcta evaluación del examen. Así, a partir de la experiencia expresada por el especialista, se busca desarrollar un método que permita una correcta evaluación de cada examen.

El proceso de segmentación consta de varios pasos. El primero, similar al propuesto con el mismo fin en [1,2], consta de un operador de Canny, empleando un valor de  $\alpha = 1$ , usado para detectar los bordes en la imagen. Este paso es seguido por un proceso de selección de los máximos locales del gradiente y una umbralización por histéresis [10].

Debido a que algunas estructuras pueden resultar abiertas en las imágenes obtenidas, se aplica una opera-

ción de cerrado morfológico. Como resultado se observa, tal como lo ilustra la figura 2c, pueden aparecer bordes abiertos no pertenecientes a bacilos. Puesto que estos bordes no tienen más que un pixel de ancho, su eliminación se logra ejecutando una operación de llenado de todos los contornos cerrados y, posteriormente, haciendo una apertura morfológica.

Dado que las imágenes están en formato RGB, es necesario estudiar su comportamiento espectral. Como se observó en [8], las principales características de color de estas imágenes son:

- Cada región de la imagen, perteneciente al fondo o al bacilo presentan un color, intensidad y textura similar.
- Los bacilos presentan una fluorescencia verde que pasa por el amarillo y puede llegar al blanco.

Se observó en los histogramas de cada canal y el perfil de los bacilos que aquellos que presentan una coloración verde muestran un valor alto en ese canal, y muy bajo en los otros, como ilustra la figura 1. Los amarillos muestran un valor alto y similar en los canales rojo y verde. Finalmente, cuando los bacilos aparecen con una coloración blanca, el valor es alto y similar en los tres canales. También se vio en todos los bacilos analizados que siempre hay, al menos, un pixel en el cual el valor del color verde es mayor que el del rojo. Además,

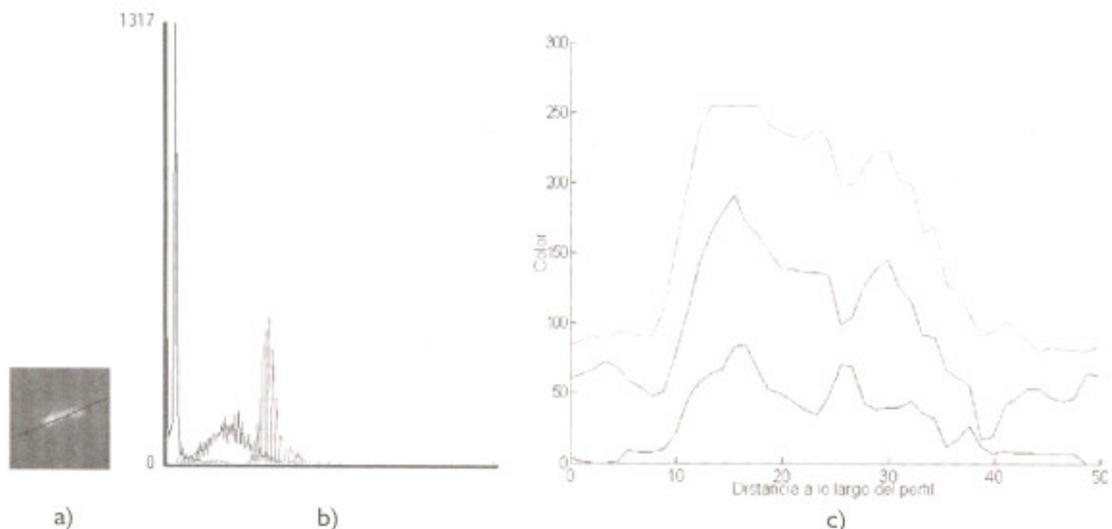


Figura 1. Características cromáticas de un bacilo. a) Imagen original. b) Histograma. c) Perfil.

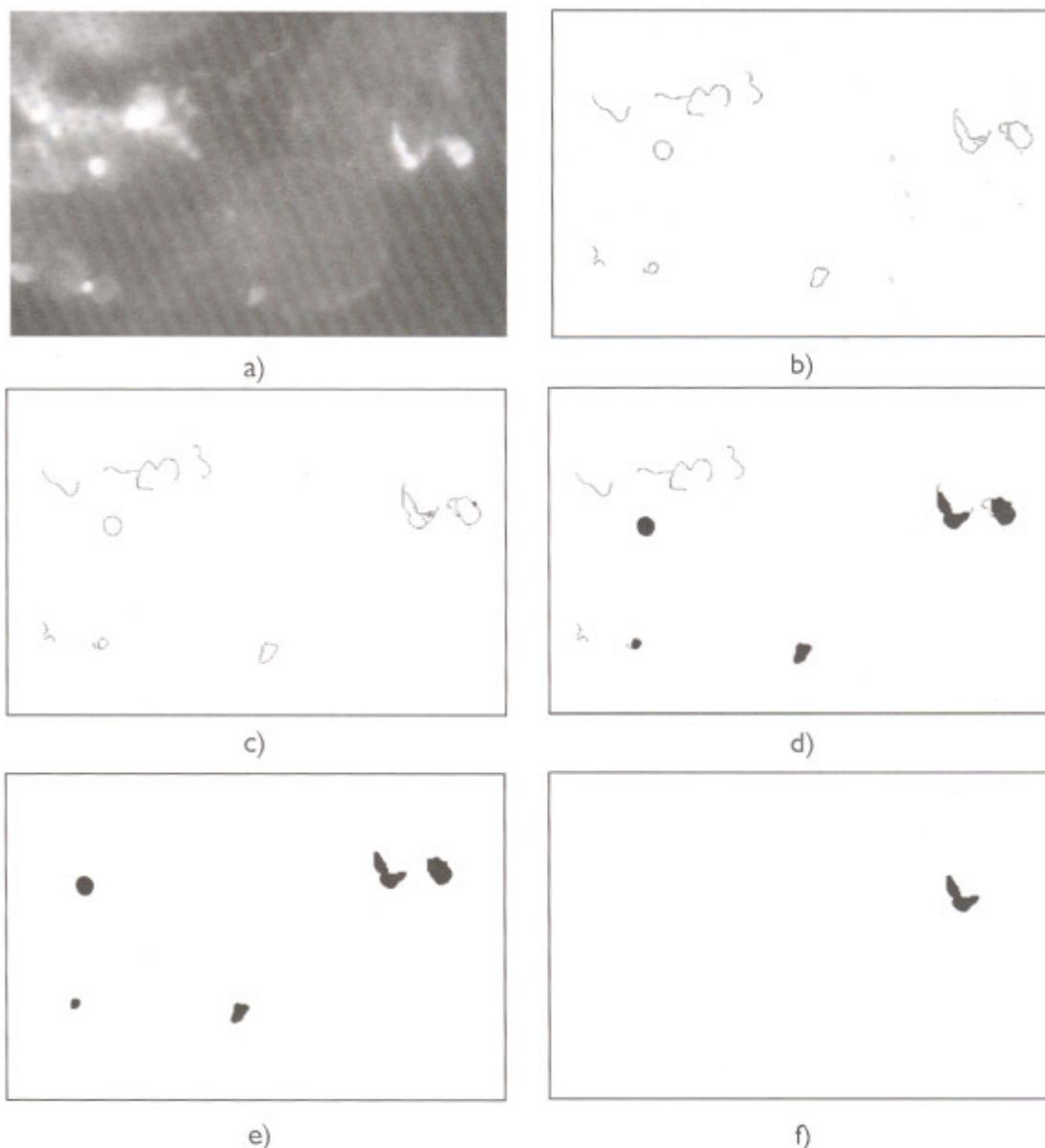


Figura 2. Muestra negativa. a) Imagen original. b) Detección de bordes. c) Cerrado. d) Llenado de regiones cerradas. e) Apertura. f) Segmentación por color.

se encontró que el canal azul no aporta información importante para distinguir a los bacilos, lo cual concuerda con las conclusiones de Demantova *et al.* [5].

Basados en estas características y en un análisis estadístico de las intensidades de las muestras, se definió la siguiente etapa en la técnica de segmentación. Se barre la imagen y se estudian los píxeles de cada objeto. Un objeto es conservado si tiene al menos un píxel cuyo valor en el canal verde es superior a 180 y al valor del canal rojo; en caso contrario, el objeto es eliminado.

Las figuras 2 y 3 muestran los resultados obtenidos con esta técnica aplicados a una imagen negativa y una positiva.

Una vez segmentada la imagen, se realiza un proceso de identificación con el fin de averiguar si los objetos restantes corresponden a bacilos (en cuyo caso la imagen será dada como positiva) o no, caso en el cual la imagen es negativa. Varios métodos pueden ser empleados para clasificar los objetos en la imagen, éstos incluyen las redes neuronales, ya em-

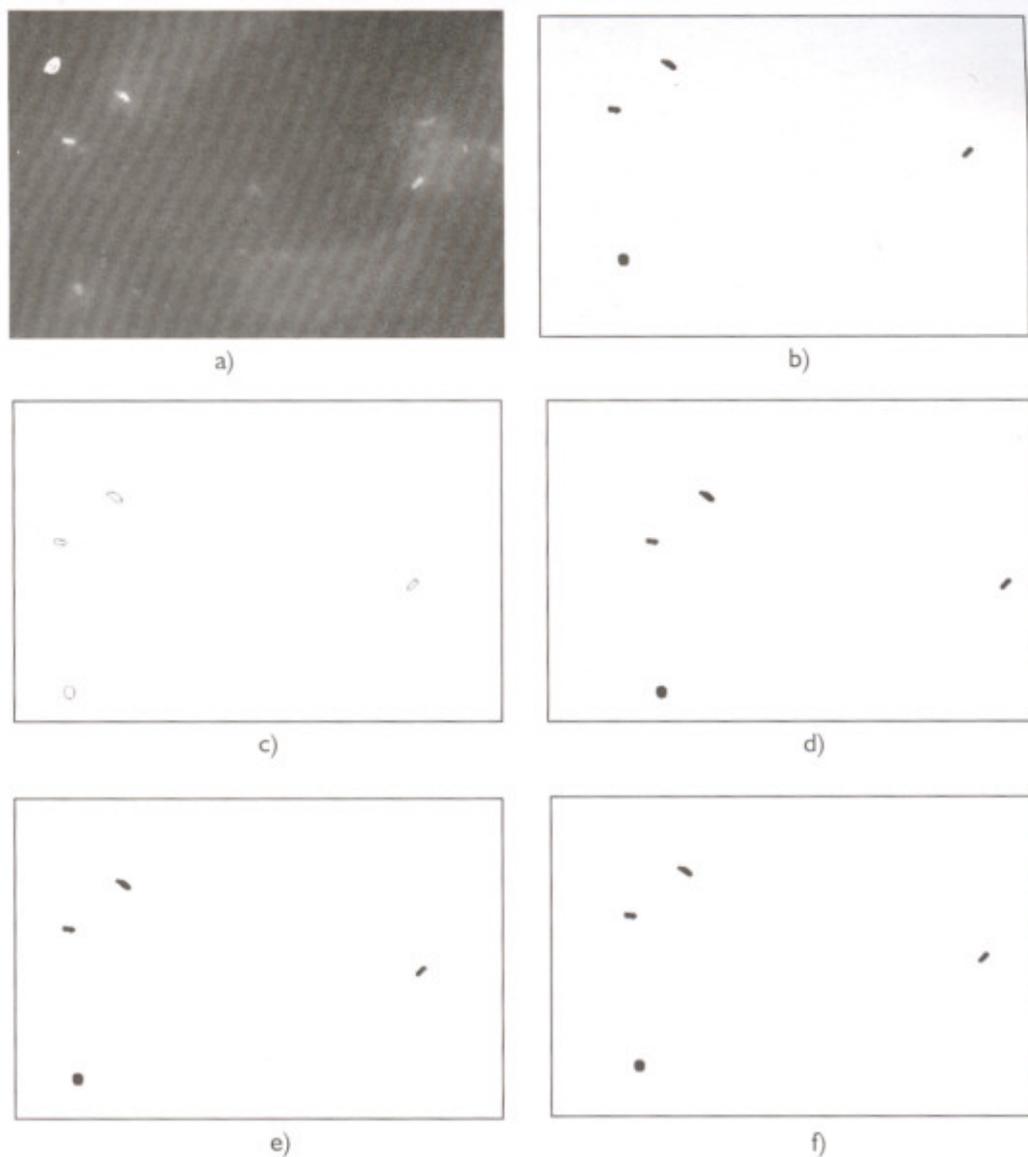


Figura 3. Muestra positiva. a) Imagen original. b) Detección de bordes. c) Cerrado. d) Llenado de regiones cerradas. e) Apertura. f. Segmentación por color.

pleadas por Veropoulos [1,2] y técnicas de agrupamiento.

Para realizar la identificación, los objetos segmentados son representados mediante descriptores, los cuales deben ser preferiblemente invariantes a la traslación, rotación, cambio de escala y punto de inicio del muestreo cuando el descriptor se obtiene a partir del contorno del objeto. Para evitar el cálculo de los descriptores sobre todos los objetos hallados, antes de realizar la descripción, los objetos son filtrados de acuer-

do al número de píxeles de su contorno, eliminando aquellos demasiado pequeños y demasiado grandes con respecto al tamaño de un *M. tuberculosis*.

Se calcularon algunos descriptores: área, compactividad, eje mayor, eje menor, excentricidad, perímetro, solidez, los 7 momentos de Hu y los primeros 16 descriptores de Fourier normalizados. Luego, se realizó un análisis del conjunto de 110 bacilos tomados, mostrados en la figura 4, con el fin de estudiar y sacar conclusiones sobre el conjunto de muestras de refe-

rencia. Este proceso establece la distribución de las muestras, determinante para definir los pasos posteriores en la clasificación, además de establecer si el número de muestras tomadas es adecuado y representativo de forma que decrezca el error de clasificación. Mediante un análisis visual previo se observó que los *M. tuberculosis* no eran uniformes, sino que podían subdividirse en clases de acuerdo a su forma, grosor y longitud. Además, algunos objetos –a pesar de ser bacilos–, tenían forma ovoide, que en general no es característica de un *M. tuberculosis*, que podía ser confundida con otro objeto, como lo ilustran los casos mostrados en la figura 4. Lo anterior coincide con lo expresado por el especialista consultado quien afirmó que si bien algunos bacilos no son fáciles de distinguir por su forma y su color cuando se observan de manera aislada, ellos son determinados como tales si se observa que en la imagen hay bastantes bacilos fácilmente reconocibles. Por tanto, se realizó una edición del conjunto de referencia. Como menciona Cortijo [11], los patrones extraños pueden producir errores en la asignación de los objetos, pues suelen aparecer en zonas cercanas a las regiones de decisión, afectando seriamente el resultado de la clasificación. De esta manera, se redujo la cantidad de objetos de muestra, eliminando aquellos cuya forma no correspondiera claramente a un bacilo y que pudieran afectar la clasificación. Además, se observó estadísticamente, que estos objetos se caracterizan por su baja compacidad y excentricidad. De esta manera, sólo se conservaron 88 bacilos.



Figura 4. Conjunto de *M. tuberculosis* usados para el desarrollo del estudio.



Figura 5. Ejemplos de bacilos descartados debido a su forma.

Para determinar el número de clases presentes que mejor definían las muestras se emplearon 7 técnicas de agrupamiento no supervisado: adaptivo, distancias encadenadas (*chain map*), K-medias, secuencial, máxima distancia también conocido como max-min y Batchelor y Wilkins, ISODATA y matriz de similitud. Estas técnicas resultan útiles en estos casos, en los cuales se trata de agrupar objetos sin información a priori, es decir, a partir de un conjunto de objetos no etiquetados [12].

Simultáneamente, con el fin de retener y reconocer los descriptores más útiles un proceso de eliminación y análisis fue desarrollado. Se busca que los descriptores empleados sean independientes o que su dependencia sea mínima. El empleo de un menor número de descriptores disminuye el tiempo de cálculo y, en muchos casos, permite obtener mejores resultados en la clasificación. Suponiendo que los descriptores tienen una distribución gaussiana, se graficaron todas las distribuciones para un número diferente de clases y se observó qué tanto se solapaban las clases. Como se observa en la figura 6, los 4 primeros momentos de  $H_u$ , la excentricidad y la compacidad presentan una buena separación entre cada clase del bacilo. Por este motivo fueron escogidos para realizar la clasificación.

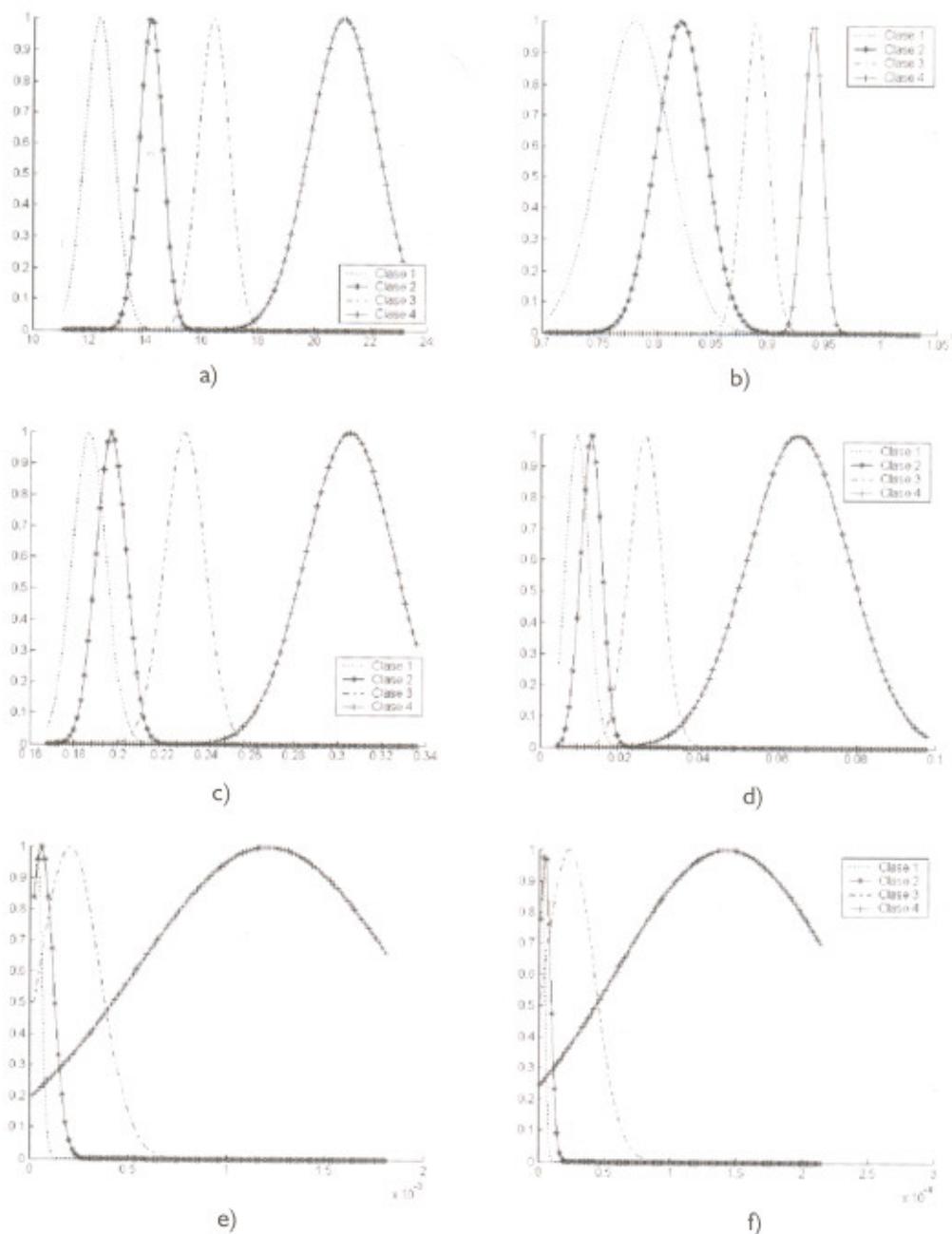


Figura 6. Distribución de los descriptores escogidos. a) Compacidad. b) Excentricidad. c) Hu1. d) Hu2. e) Hu3. f) Hu4.

Además, para el agrupamiento por k-medias, en las cuales el único parámetro definido por el usuario es el número de clases por encontrar, se obtuvo la silueta y el coeficiente de silueta para los agrupamientos resultantes, para un rango de clases entre 3 y 10. Para evitar el sesgo producido en esta técnica, debido a la escogencia de los centroides iniciales, que suele ocasionar, cuando el número de muestras es alto, que los centroides

de las clases encontradas no correspondan al mínimo global sino a un mínimo local, se calcularon los centroides varias veces, seleccionando aleatoriamente distintos objetos como centroides iniciales y comparando los resultados obtenidos hasta encontrar un mínimo que se repitiera de manera consistente. Como resultado de estos análisis, se encontraron 4 clases. La figura 7 presenta la silueta de las clases encontradas, cuyo coefi-

ciente es 0.702, valor que indica el hallazgo de una estructura razonable. El valor de la silueta de un objeto mide su similitud con respecto a los demás objetos de la misma clase comparada con los objetos pertenecientes a las demás clases.

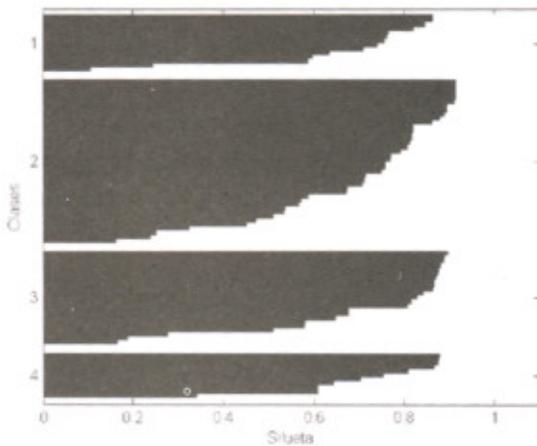


Figura 7. Silueta de las cuatro clases obtenidas.

Encontradas las clases, se intentó clasificar usando el método del vecino más próximo. Los mejores resultados se obtuvieron usando como descriptores los 4 primeros momentos de Hu y la compactividad, y utilizando la distancia de Mahalanobis como criterio de similitud, puesto que la magnitud de los descriptores y su varianza no es similar.

Dada la naturaleza de los objetos por identificar, se tenía una descripción lingüística que podía ser empleada para realizar la clasificación. Esta información se refiere a la longitud y el ancho de los bacilos. Unos ensayos preliminares limitados al empleo de los centroides de las clases obtenidas no permi-

tieron obtener buenos resultados, dado que varios objetos, distintos de los bacilos, poseían descriptores muy similares a los encontrados y, por tanto, eran clasificados como bacilos. La compactividad y la excentricidad son dos descriptores complementarios adecuados para conocer de forma aproximada si un objeto es largo y delgado, forma típica de un bacilo. Empleando estos valores y los descriptores de Hu se construyó un árbol de clasificación [11] mostrado en la figura 8.

En un segundo árbol se ajustó una de las reglas dadas y se fijaron dos reglas adicionales con el fin de aumentar las posibilidades de escoger correctamente un examen como positivo. Estas reglas surgen de la siguiente observación: muy ocasionalmente pueden encontrarse bacilos cuya longitud es mayor de la esperada (no alcanzan a ser clasificados como bacilos ya que su forma no se asemeja a ninguna de las cuatro clases encontradas). De éstos resulta difícil tener un número suficiente de muestras dada su poco frecuente aparición. Por tanto, se simularon algunos bacilos con el fin de establecer los valores de compactividad y excentricidad que podrían tener. El árbol resultante es presentado en la figura 9.

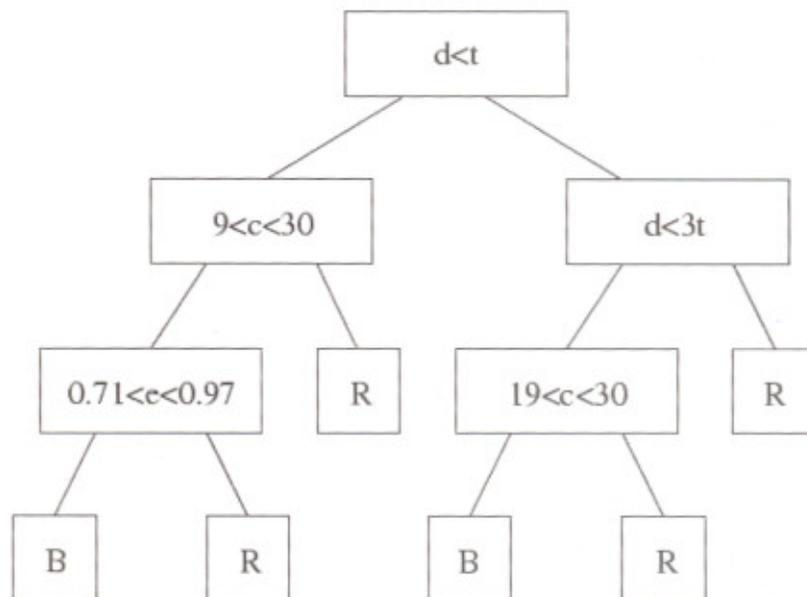


Figura 8. Árbol de clasificación 1. Se toma cada rama izquierda, si el nodo es verdadero; la derecha, si es falso. B significa bacilo y R un objeto rechazado. d es la distancia de Mahalanobis entre el objeto y la clase más cercana, c es la compactividad, e la excentricidad y t un umbral de distancia, fijado en 10.

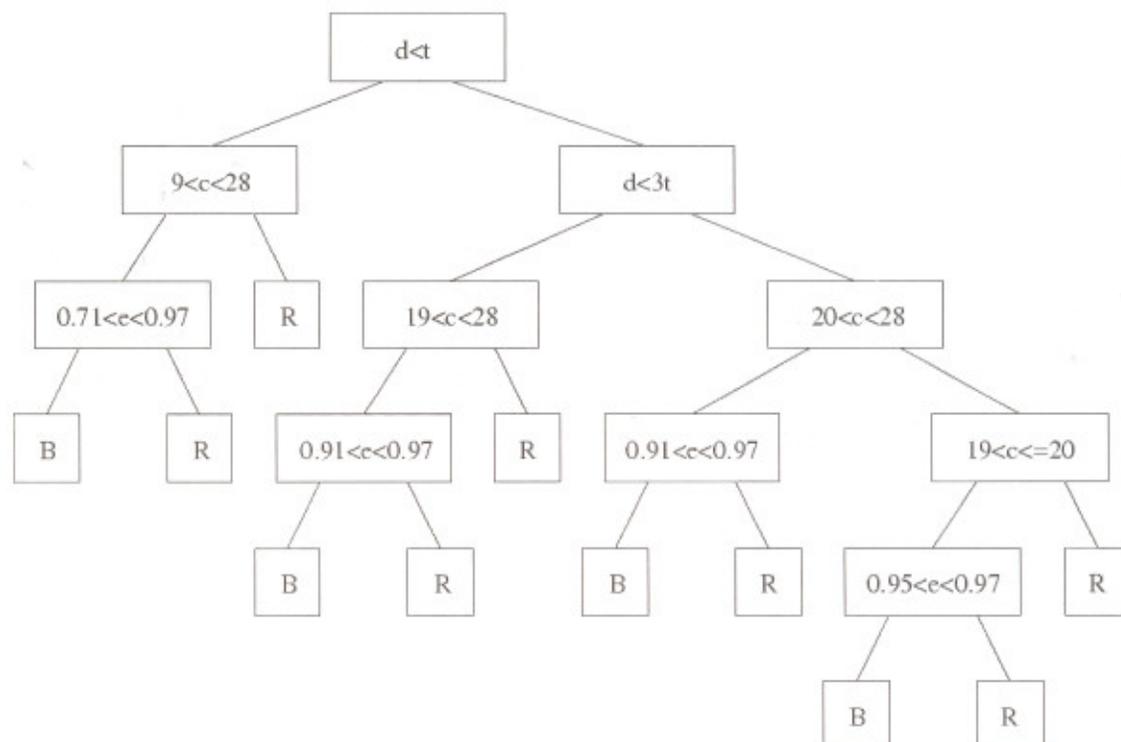


Figura 9. Árbol de clasificación 2.

## RESULTADOS

En las figuras 2 y 3 se observa que la técnica de segmentación desarrollada permite obtener la mayor parte de los bacilos presentes en una imagen y eliminar la mayor parte de los objetos que no son bacilos. En algunas imágenes negativas todos los objetos fueron eliminados y el proceso de identificación fue innecesario para declarar una imagen como negativa.

La tabla 1 presenta los resultados obtenidos. A partir de éstos, se calculan los valores de especificidad y sensibilidad obtenidos, presentados en la tabla 2 incluidos los intervalos de confianza.

Tabla 1. Resultados obtenidos. En la prueba se emplearon 397 imágenes negativas y 60 positivas.

	Falsos positivos	Falsos negativos
Árbol 1	12	2
Árbol 2	12	0

A partir de los valores mostrados en la tabla 2, puede verse que el árbol 2 permite obtener mejores resultados y que la técnica desarrollada aparece como una

solución viable para la identificación de bacilos en muestras de esputo. Se obtuvieron resultados superiores a los encontrados por Veropoulos para el caso de la sensibilidad [4].

Tabla 2. Especificidad y sensibilidad obtenidos.

	Especificidad	Sensibilidad
Árbol 1	91,43±2,75	93,33±6,31%
Árbol 2	91,43±2,75	100%

## CONCLUSIONES

Se ha presentado una nueva técnica para el análisis de imágenes de fluorescencia del esputo. La técnica comprende un nuevo método para la segmentación y otro de identificación.

Los resultados obtenidos de la segmentación permiten verificar la eliminación de una gran cantidad de objetos, de manera que sólo aquellos caracterizados por un color similar al de los bacilos puedan pasar, permitiendo una gran reducción en el cálculo posterior y eli-

minando aquellos objetos o bacilos cuya forma es muy similar a la del *Mycobacterium tuberculosis*, pero que difieren en su color.

Una de las etapas más importantes del sistema desarrollado consistió en el análisis y la edición de las muestras de bacilos tomadas. En este trabajo se emplearon 110 muestras, pero su número puede ampliarse para adquirir mayor conocimiento de los bacilos con el fin de mejorar los resultados encontrados. Igualmente se espera ampliar el número de imágenes con el fin de obtener resultados más significativos y disminuir el intervalo de confianza. Aunque los descriptores utilizados para calcular los centroides de las clases presentaron las características deseables para la elaboración de un sistema de reconocimiento, se encontraron algunos casos en los que fallaba la clasificación. Por este motivo se empleó el conocimiento heurístico acerca de los bacilos y se desarrolló un árbol de clasificación. Se obtuvieron resultados bastante satisfactorios. La siguiente etapa de la investigación será su implantación para prueba experimental en el Hospital General Universitario Gregorio Marañón de Madrid.

Dada la información heurística empleada para construir el árbol de clasificación, se han realizado pruebas utilizando lógica difusa para la identificación. Sin embargo, los resultados obtenidos hasta ahora no han sido tan satisfactorios como los obtenidos con la técnica presentada.

## REFERENCIAS

1. Veropoulos, K.; Learmonth, G.; Campbell, C.; Knight, B.; Simpson, J. (1999). Automatic identification of tubercle bacilli in sputum. A preliminary investigation. *Analytical and Quantitative Cytology and Histology* 21(4), p. 277-281.
2. Veropoulos, K.; Campbell, C.; Learmonth, G.; Knight, B.; Simpson, J. (1998). The automatic identification of tubercle bacilli using image processing and neural computing techniques, en *Proceeding of the 8th International Conference on Artificial Neural Networks*. 2, p. 797-802.
3. Wilkinson, M. (1996). Rapid automatic segmentation of fluorescent and phase-contrast images of bacteria, en *Fluorescence Microscopy and Fluorescent Probes*. New York: Ed. J. Slavik, Plenum Press.
4. Álvarez-Borrego, J.; Mourino, R.; Cristóbal, G.; Pech, J. (2000). Invariant optical color correlation for recognition of vibrio cholerae O1. *Int. Conf. on Pattern Recognition*, 2847, p. 283-286, Barcelona, España.
5. Demantova, P.; Sakamoto, D.; Ioshii, S.; Gamba, H. (2001). Segmentación automática de bacterias para el método DEFT, en *Proceedings of the II Latin American Congress on Biomedical Engineering*. Havana, Cuba.
6. Sammouda, R.; Niki, N.; Nishitani, H.; Nakamura, S.; Mori, S. (1997). Segmentation of sputum color image for lung cancer diagnosis, en *Int. Conf. on Image Processing*. 1, p. 243-246, Washington, USA.
7. Sammouda, R.; Niki, N.; Nishitani, H.; Kyokage, E. (1998). Segmentation of sputum color image for lung cancer diagnosis based on neural network, *IEICE Transactions on Information and Systems*, E-81-D(8), p. 862-871.
8. Forero, M.; Sierra, E.; Álvarez, J.; Pech, J.; Cristóbal, G.; Alcalá, L.; Desco, M. (2001). Automatic sputum color segmentation for tuberculosis diagnosis, en *Algorithms and Systems for Optical Information Processing*, p. 251-261.
9. Forero, M.; Sroubek, F.; Álvarez, J.; Malpica, N.; Cristóbal, G.; Santos, A.; Alcalá, L.; Desco, M.; Cohen, L. (2002). Segmentation, autofocusing and signature extraction of tuberculosis sputum images, en *Photonic Devices and Algorithms for Computing*, p. 171-182.
10. Forero, M. (2002). *Introducción al procesamiento digital de imágenes*. Bogotá: Ed. M.G. Forero.
11. Cortijo, F. (2001). Técnicas supervisadas II: Aproximación no paramétrica. Rep. Tec. URL: [www-etsi2.ugr.es/depar/ccia/ff/www/](http://www-etsi2.ugr.es/depar/ccia/ff/www/), Universidad de Granada.
12. Cortijo, F. (2001). *Técnicas no supervisadas: Métodos de agrupamiento*.