

# Desarrollo de una Herramienta de Análisis y determinación de Características en imágenes de Citología.

C. Suárez-Varela Pineda<sup>1</sup>, E. Gómez González<sup>1</sup>, Eloy Rivas<sup>2</sup>, J. Márquez Rivas<sup>3</sup>

<sup>1</sup> Grupo de Física Interdisciplinar (GFI), Escuela Superior de Ingenieros, Sevilla, España, cristina.s.pineda@gmail.com

<sup>2</sup> Servicio de Anatomía Patológica, Hospitales Universitarios Virgen del Rocío, Sevilla, España.

<sup>3</sup> Servicio de Neurocirugía, Hospitales Universitarios Virgen del Rocío, Sevilla, España.

## Resumen

*En este trabajo se presenta una herramienta (programa) de procesamiento de imágenes de citología que permite segmentar las células de una muestra y calcular sobre ellas los parámetros necesarios para caracterizarla y poder clasificarla según el criterio de Bethesda, proporcionando un primer cribado de las imágenes y facilitando de este modo el trabajo de los especialistas. Se presentan la interfaz gráfica de la herramienta diseñada y los resultados de las imágenes segmentadas automáticamente así como los parámetros calculados en su aplicación para la automatización de los test de Papanicolaou. Este trabajo se ha realizado en colaboración entre el Grupo de Física Interdisciplinar de la Escuela de Ingenieros de Sevilla y Hospitales Universitarios Virgen del Rocío.*

## 1. Introducción

El cáncer de cuello de útero es el segundo tipo de cáncer más frecuente en la población mundial femenina. Se diagnostica cerca del 80% de los casos en países no desarrollados [1].

Se concluye en [2] que la citología cérvicovaginal satisface ampliamente los requisitos como método ideal de tamizaje del cáncer de cuello uterino en sus etapas más tempranas. Es por ello que se hace necesario un seguimiento de la población femenina en edad de riesgo (entre 25 y 64 años) tendiendo a la medicina preventiva.

En este proyecto se presenta el diseño y elaboración de una herramienta software ([3], [4]) que permita la medición y extracción de características celulares de citologías cervico-uterinas, con objeto de clasificarlas en un paso posterior. Esto contribuiría a agilizar el diagnóstico del cáncer de cuello uterino haciendo un screening previo al análisis del patólogo. Esto supondría un alivio en la carga de trabajo, aunque se ha de tener presente que en ningún caso con esta herramienta se pretende sustituir el criterio del médico, sino ayudarlo.

En un estado normal las estructuras nucleares clave son: núcleo, nucléolo y cromocentro redondeados, un patrón de cromatina y paracromatina uniforme con un grosor regular del anillo cromatínico, asociado internamente a la membrana nuclear. Los cambios debidos al cáncer se relacionan con alteraciones en estas estructuras como: membrana nuclear irregular, cromatina condensada en grumos gruesos, paracromatina más fácilmente

apreciable, nucléolo más grande, irregular y de forma angulada, presencia de mitosis anormales, multinucleación, con nucléolos que exhiben pleomorfismo e incremento en el radio núcleo citoplasma[7].

Se han desarrollado herramientas que permiten contrastar el criterio del patólogo con resultados cuantitativos (como la descrita en [5]), pero en esta que se presenta se propone como un paso previo al análisis del patólogo con objeto de clasificar las células según el estándar de clasificación Bethesda [6].

Se han desarrollado también en los últimos años algoritmos multiespectrales, que requieren de una técnica de adquisición de imagen que registre simultáneamente información espacial y espectral de la muestra[18].

Por otra parte existen dos sistemas comerciales (ThinPrep Imaging System y FocalPoint GS Imaging System) que desempeñan esta tarea; pero estos equipos aun son poco versátiles y caros[19].

## 2. Material y Método

### 2.1. Recolección de las muestras

Se trabaja con un conjunto de imágenes estandarizadas proporcionadas por el Servicio de Anatomía Patológica del Hospital Universitario Virgen de Rocío. Estas imágenes cumplen los siguientes criterios:

- Coloración de Papanicolaou
- Que cumplan las condiciones del Sistema Bethesda (evaluadas por los patólogos)
- Muestras diagnosticadas.

### 2.2. Entorno de trabajo

Se ha utilizado un PC convencional (Pentium 4 2.8 GHz 1GB RAM) trabajando sobre el sistema Operativo Windows XP Home Edition.

El lenguaje de programación elegido para el desarrollo de esta herramienta es el MATLAB®, en su versión R2006a. Nos hemos basado en algunos algoritmos descritos en [8] y [9].

### 2.3. Metodología

A continuación se sintetizan los requisitos del usuario, es decir, las principales funcionalidades que debe tener la aplicación ([10][2]).

- Aplicar técnicas de mejora de la imagen.
- Obtener la región de interés (ROI). En este punto deben eliminarse las zonas de la imagen sin muestras representativas de células, o con excesiva concentración de estas, de modo que no permitan una visualización clara de la patología.
- Sobre la superposición de la imagen original con la máscara ROI, segmentar las células y/o los núcleos.
- Detectar propiedades morfológicas de las células y/o núcleos que nos resulten de interés para detectar patologías (proporción del tamaño del núcleo respecto a la célula, forma de la membrana exterior, etc.). Para este punto necesitamos información de los patólogos acerca de la morfología de cada patología.

Para facilitar el uso de esta herramienta al médico, se ha desarrollado una plataforma de visualización completa, amigable y de fácil manejo.

#### *Preprocesado*

Las técnicas de procesado (Véase [10][1], [12][1], [13][1] y [14][1]) pretenden mejorar o realzar las propiedades de la imagen para facilitar las siguientes operaciones de la Visión Artificial, tales como las etapas de segmentación, extracción de las características y finalmente la interpretación automática de las imágenes.

Para mejorar la calidad de la imagen, se han tenido en cuenta los problemas que puedan afectar a cada una de las características de adquisición de la imagen.

Para solucionar el problema del ruido se da la opción de aplicar técnicas de suavizado. Se ofrecen dos opciones de suavizado: Filtro de la media y filtro de la mediana.

Para corregir los efectos de una mala iluminación se aplica la ecualización adaptativa sobre el histograma de la imagen (tras corregir el ruido). Aunque este proceso no aumenta la información en la imagen, si la mejora para las siguientes etapas.

Se ofrece la posibilidad de realzar los bordes con objeto de obtener mejores resultados a la hora de detectarlos. Esto se hace aplicando un filtro 'unsharp' o realce por filtrado ponderado.

Una expresión genérica de esta función será:

$$f_e(m, n) = [g(m, n) - \mu_g(m, n)] + g(m, n)\alpha$$

Donde el pixel correspondiente a las coordenadas (m,n) en la imagen realzada  $f_e(m, n)$  es dado por una combinación de pesos del pixel de la imagen degradada  $g(m, n)$ , y la diferencia entre el pixel y la media local  $\mu_g(m, n)$ .

El efecto deseado se consigue asignando un valor apropiado para el pixel central de la máscara, que equivale al valor del peso  $\alpha$ .

#### *Elección de ROI*

El objeto de esta etapa es determinar las regiones donde se encuentran células, eliminando el fondo de la imagen para que el procesado sea más rápido. En esta etapa se creará una máscara con las regiones de interés. Se ofrece la opción de realizar esta etapa de forma automática o manual:

Para la detección automática de los bloques de células se aplican propiedades morfológicas sobre la imagen de contornos.

En primer lugar se aplica un filtro para la detección de bordes. Existe la posibilidad de elegir entre el filtro de Sobel o de Prewitt, y aplicar el que ofrezca mejores resultados. Sobre la imagen binarizada de bordes se aplica procesamiento morfológico para crear una máscara de región. Esto permite crear una máscara de regiones cerradas eliminando las regiones de pequeño tamaño.

Se ofrece también la opción de seleccionar la región de interés manualmente. Aplicando la función de MATLAB® 'roipoly', con la que debemos seleccionar un polígono que contenga la región de interés sobre la imagen marcando puntos con el ratón. Se pueden seleccionar tantas regiones de interés como se desee.

#### *Segmentación por Crecimiento de regiones*

La estrategia en este tipo de segmentación (Véase [10][1], [12][1], [13][1] y [14][1]) está basada en los criterios de similitud y continuidad de los píxeles que forman una región. El paso previo radica en hacer una correcta selección de los píxeles semilla. Para este paso se realiza una búsqueda de máximos regionales de imágenes filtradas. Este método tiene la ventaja de que hace la elección de marcadores independiente al valor de un umbral. El filtro de la imagen utilizado ha sido h-máxima. Este filtro elimina toda 'colina' de la imagen cuyo peso es menor que h y decrementa el peso del resto de las 'colinas' en h. De esta forma elimina los máximos locales que puedan generarse por ruido o los máximos locales poco significativos.

#### *Segmentación por Transformada Watershed*

Esta transformada se utiliza en muchos artículos (Véase [15][2] y [16][2]) para hallar la separación entre células. Parece la mejor alternativa para separar objetos solapados, ya que utiliza información de los bordes de la imagen (Laplaciana). Por sí sola no es eficaz, ya que proporciona demasiadas regiones (sobresgmentación), pero si añadimos marcadores de las células limitamos el número de objetos que se van a segmentar. SE han utilizado como marcadores de las células los núcleos de las mismas, ya que una célula va a tener al menos un núcleo.

### Extracción de características

Partiendo de los candidatos a célula segmentados y etiquetados, valoramos en primer lugar que se trate de una célula. Para ello aplicamos dos filtros morfológicos:

- Filtro de tamaño: elimina los candidatos demasiado grandes y los demasiado pequeños;
- Filtro de existencia de núcleos: elimina los candidatos que no tengan núcleo definido, y los que tengan más de tres núcleos (ya que probablemente será un área con artefactos en la imagen).

Tras determinar las células, se segmentan los núcleos. Para ello usamos la detección de bordes sobre la imagen de la célula y aplicamos dos filtrados:

- Filtro de tamaño (número de píxeles),
- Filtro de dispersión (distancia al centroide de todos los puntos del borde).

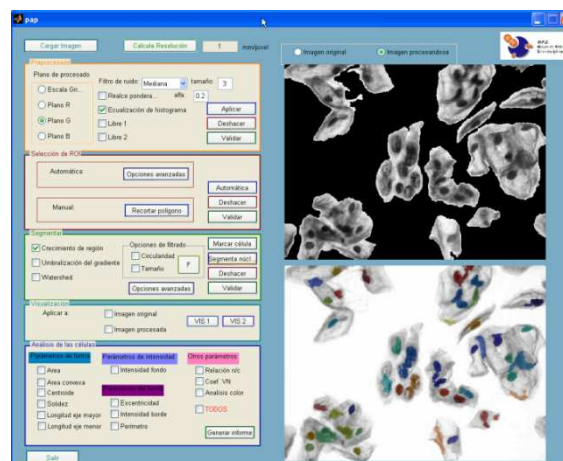
Los valores umbrales de estos filtros se han estimado de manera empírica, aunque ofrecen la posibilidad de ser modificados por el usuario de la aplicación.

Partiendo de las células con su núcleo delimitado, calculamos las características que más nos interesen para evaluar el estado de la célula. En este caso se han calculado las siguientes características [17][2]:

- Morfológicas (tanto para el núcleo como para el citoplasma)
  - Área.
  - Área convexa.
  - Centroides.
  - Excentricidad.
  - Longitud del eje mayor.
  - Longitud del eje menor.
  - Solidez.
- De intensidad
  - Valor medio de intensidad del citoplasma.
  - Valor medio de intensidad del núcleo.
- Que relacionen núcleo y citoplasma (Véase [16])
  - Relación núcleo/citoplasma.
  - Coeficiente de variación del núcleo.
  - Multinucleación.
  - Coeficiente de posición del núcleo.
- Análisis de color.

### 3. Resultados

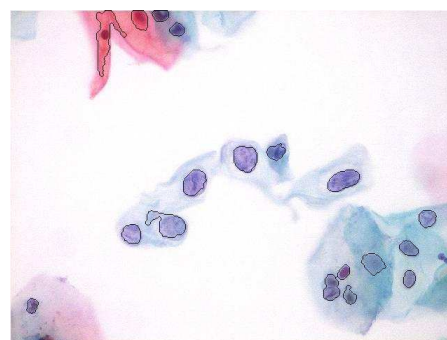
Se ha creado una herramienta eficaz y versátil (*Ilustración 1*). La herramienta permite un ajuste dinámico de los parámetros de procesamiento y análisis, permitiendo la visualización de cada paso. Finalmente genera un fichero con los parámetros usados para el análisis y el resultado se guarda en forma de tabla en una hoja de Excel®.



*Ilustración 1. Aspecto de la herramienta*

La herramienta creada se ha testado con una secuencia de imágenes de citologías. Esto ha servido para evaluar el funcionamiento de la aplicación.

Se han obtenido resultados satisfactorios en la segmentación de los núcleos de casi todas las imágenes. En el caso de la *Ilustración 2* la aplicación detecta todos los núcleos de la imagen, aunque se dan algunos falsos positivos. Se puede perfeccionar la adaptación a los bordes y diseñar otra opción de filtrado para eliminar algunos falsos positivos.



*Ilustración 2. Resultado de una imagen de citología con los núcleos segmentados por el método de la transformada Watershed.*

### 4. Conclusiones

La prueba de Papanicolaou es el primer paso para el diagnóstico de cáncer de cérvix. Dado que es de gran importancia detectar la enfermedad en sus primeras etapas, esta prueba se ha convertido en un análisis de control rutinario. Es por ello que en este proyecto se ha desarrollado una herramienta que podría llegar a ayudar al médico en el cribado de dichas pruebas.

Con esta herramienta, el especialista puede:

- Elegir las distintas opciones de mejora de la imagen para mejorar el análisis.
- Elegir el tipo de segmentación y los parámetros necesarios para la misma entre las opciones que se ofrecen.
- Analizar las imágenes, extrayendo información cuantitativa de ciertos parámetros seleccionados a "priori" basándonos en el análisis diagnóstico

de los especialistas. Sin embargo, es posible que no todos estos parámetros proporcionen información útil. Por ello sería interesante realizar como continuación a este proyecto, un estudio de éstos para conocer cuáles son realmente necesarios en nuestro análisis.

La potencialidad de esta herramienta se basa en el procesamiento de imágenes “por lotes”. Optimizando los parámetros, podemos analizar varias imágenes y extraer parámetros de las células que permitan clasificarlas. De estos pueden extraerse estadísticas que señalen en que grupo de la clasificación Bethesda se encuentran las células y en qué porcentaje.

Finalmente, debemos señalar que, aunque la aplicación de la herramienta en las imágenes de citologías vaginales disponibles ha sido evaluada positivamente por los expertos del Servicio de Anatomía Patológica del Hospital Universitario Virgen de Rocío, es necesario comprobarla en una muestra mayor que permita un análisis estadístico de resultados.

Hay que destacar que no es necesario un tipo de imágenes especial para el procesamiento; se han utilizado imágenes digitalizadas de muestras de Papanicolaou, lo cual hace el método fácilmente implantable.

El método expuesto, a diferencia de otros propuestos en la literatura, ofrece una segmentación automática de las imágenes; sin necesidad de una participación activa por parte del técnico. Y a continuación calcula las propiedades que permiten su diagnóstico.

El siguiente paso que se ha planteado trata de optimizar los parámetros de los que depende la segmentación para alcanzar un compromiso entre sensibilidad y especificidad. Se ha tomado como especificación de diseño dar más peso a la sensibilidad, ya que se trata de un algoritmo de screening. Se pretende agilizar el trabajo del experto, concentrándolo en los casos más dudosos y liberándolo así de la carga de trabajo más tediosa.

## Agradecimientos

Agradecemos al Servicio de Anatomía Patológica del Hospital Universitario Virgen de Rocío la orientación y las imágenes facilitadas para el desarrollo de este trabajo.

## Referencias

- [1] Instituto Nacional de Cáncer. 2008. Available from [www.cancer.gov](http://www.cancer.gov) (Consultada: Mayo 2008).
- [2] Bueno Montaña, M. Tamizaje en Cáncer Ginecológico. *Guías de práctica clínica basada en la evidencia*.
- [3] Oppelt, A. 2005. *Imaging System for Medical Diagnostics*. Primera ed. Siemens.
- [4] Gómez González, E.: Computer-aided diagnosis by Virtual Holographic Recognition, in Brunch, H-P., Köckerling, F., Bouchard, R., Schug-Pass, C. (Eds.): “New aspects of High Technology in Medicine”, Monduzzi Ed., 2000, 331-335.
- [5] Martínez, E., A. Mendoza, C. J. Uribe, and E. García. 2005. Modelo computacional para caracterización de células escamosas de citologías cervico-uterinas. *Revista colombiana de Biotecnología* 7, no. 2:35.
- [6] Lacruz, C. 2003. Nomenclatura de las lesiones cervicales (de Papanicolaou a Bethesda 2001). *Revista Española de Patología* 36, no. 1:5.
- [7] Bibbo, M. *Comprehensive Citopathology*, Second Edition, W.B. Sanders, 1996.
- [8] Jähne, Bernd. 2004. *Practical Handbook on Image Processing for Scientific and Technical Applications*. 2ª ed. USA: CRC Press LLC.
- [9] Semmlow, J. L. 2004. *Biosignal and Biomedical Image Processing. MATLAB-Based Applications*. Primera ed. Signal Processing and Communications Series.
- [10] Romero-Gómez M, Gómez-González E, et al: "Optical analysis of computed tomography images of the liver predicts fibrosis stage and distribution in chronic hepatitis C", *Hepatology*. 2008, 47(3):810-6.
- [11] Rafael, C. G., and Richard, E. W. 2002. *Digital Image Processing*. Segunda ed. USA: Prentice Hall.
- [12] Suetens, P. 2002. *Fundamentals of Medical Imaging*. Primera ed. Cambridge University Press.
- [13] Russ, John C. 2002. *The Image Processing Handbook*. 4ª ed. USA: CRC Press.
- [14] Starck, J. -, Murtagh, F., and Bijaoui, A. 2000. *Image Processing and Data Analysis*. 1ª ed. Cambridge: Cambridge University Press.
- [15] Malpica, N., et al. 1997. Applying Watershed Algorithms to the segmentation of clustered nuclei. *Cytometry* 28, 289.
- [16] Yang, X., H. Li, and X. Zhou. 2006. Nuclei segmentation using marker-controlled Watershed, tracking using mean-shift, and Kalman filter time-lapse Microscopy. *IEEE transactions on circuits and systems-I: Regular papers* 53, no. 11:2405
- [17] Muñoz, K. 2006. Identificación de células displásicas en el epitelio vaginal. *Umbral Científico* 9, 104.
- [18] Qiongshui Wu, *A multispectral imaging analysis system for early detection of cervical cancer*, Physics of Medical Imaging, 2005.
- [19] Mina Desai, *Role of automation in cervical cytology*, DIAGNOSTIC HISTOPATHOLOGY, 2009.