



UNIVERSIDAD CENTRAL "MARTA ABREU" DE LAS VILLAS
VERITATE SOLA NOBIS IMPONETUR VIRILIS TOGA. 1948

Facultad de Ingeniería Eléctrica.

**Centro de Estudios de Electrónica y Tecnologías de la
Información**

TRABAJO DE DIPLOMA

Evaluación de métodos no supervisados de calidad de segmentaciones en imágenes de la prueba de Papanicolaou

Autor: Asiel Hernández Valdés

Tutor: Maykel Orozco Monteagudo

Santa Clara

2014

CON SU ENTRAÑABLE TRANSPARENCIA



Universidad Central “Marta Abreu” de Las Villas

Facultad de Ingeniería Eléctrica

**Centro de Estudios de Electrónica y Tecnologías de la
Información**



TRABAJO DE DIPLOMA

**Evaluación de métodos no supervisados de calidad
de segmentaciones en imágenes de la prueba de
Papanicolaou**

Autor: Asiel Hernández Valdés

Tutor: Mr. C. Maykel Orozco Monteagudo

morozco@uclv.edu.cu

Santa Clara

2014



Hago constar que el presente trabajo de diploma fue realizado en la Universidad Central “Marta Abreu” de Las Villas como parte de la culminación de estudios de la especialidad de Ingeniería Biomédica, autorizando a que el mismo sea utilizado por la Institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos, ni publicados sin autorización de la Universidad.

Firma del Autor

Los abajo firmantes certificamos que el presente trabajo ha sido realizado según acuerdo de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

Firma del Autor

Firma del Jefe de Departamento
donde se defiende el trabajo

Firma del Responsable de
Información Científico-Técnica

PENSAMIENTO

*No podemos resolver problemas usando el mismo tipo
de pensamiento que usamos cuando los creamos.*

Albert Einstein.

DEDICATORIA

*A mi familia, en especial a mi hija por darme con su
existencia fuerza y valor para seguir adelante en este difícil
camino que es la vida.*

AGRADECIMIENTOS

A mi familia por siempre apoyarme.

A mis amigos por estar siempre a mi lado.

*A mi tutor por la ayuda, el sacrificio y los consejos
brindados.*

*A todas las personas que de una forma u otra contribuyeron a
la confección de este trabajo.*

A todos gracias.

TAREA TÉCNICA

- Aprender los fundamentos de:
 - Segmentación de imágenes.
 - Evaluación de la calidad de segmentaciones.
- Realizar un estudio específico sobre las siguientes técnicas de segmentación de imágenes:
 - Marcas de agua (*Watershed*).
 - Medias desplazadas (*Mean-shift*).
 - Clusterización usando *k*-Medias.
- Realizar una revisión bibliográfica relacionada con la segmentación de imágenes de la prueba de Papanicolaou.
- Realizar las segmentaciones de imágenes de la prueba de Papanicolaou usando la transformada *watershed*, el algoritmo *mean-shift* y las *k*-medias.
- Medir la calidad de las segmentaciones usando los métodos que se evaluarán.
- Interpretar los resultados obtenidos.

RESUMEN

El cáncer cervical es una de las enfermedades que más afecta a las mujeres en todo el mundo. La prueba de Papanicolaou es la principal contribuyente en la detección temprana de células pre-malignas y malignas en el cuello del útero. Su efectividad aumenta la posibilidad de vida de las mujeres afectadas. En este trabajo se evaluó el desempeño de cinco métodos no supervisados de evaluación de la calidad de segmentación en segmentaciones de imágenes de la prueba de Papanicolaou. Para ello se utilizó la base de casos anotados de imágenes de la prueba de Papanicolaou del Hospital Universitario de Herlev, Dinamarca. En la evaluación, donde se utilizó la Tasa de Similitud de Comparaciones Correctas, se involucraron los dos métodos de Levine y Nazif, el método de Rosenberger, el método de Borsotti y el método de Homogeneidad de Color. El mejor desempeño se obtuvo para el método de Rosenberger. Un riguroso análisis estadístico demuestra esta afirmación.

ÍNDICE

PENSAMIENTO	ii
DEDICATORIA	iii
AGRADECIMIENTOS	iv
TAREA TÉCNICA.....	v
RESUMEN	vi
ÍNDICE	1
Introducción.....	4
Capítulo 1. Métodos para la evaluación de la calidad de segmentaciones	7
1.1 Segmentación de imágenes	7
1.1.1 El método de las marcas de agua (Watershed).	9
1.1.2 El método de las medias desplazadas.	11
1.1.3 Segmentación de imágenes usando el método de las k medias.	12
1.2 Métodos supervisados para la evaluación de la calidad de segmentación.....	14
1.2.1 Método de Vinet.	14
1.3 Métodos no supervisados para la evaluación de la calidad de segmentación.....	14
1.3.1 Métodos de Levine.	14
1.3.2 Método de Rosenberger para imágenes no texturizadas.	16
1.3.3 Método de Borsotti.	17
1.3.4 Homogeneidad de color.	18
Capítulo 2. Materiales y Métodos.....	19
2.1 El cáncer cérvicouterino y la prueba de Papanicolaou	19

2.2 Base de casos Herlev	21
2.3 Evaluación de los métodos no supervisados de calidad de segmentaciones	23
Capítulo 3. Resultados y Discusión	26
3.1 Resultados para todos los casos	27
3.2 Resultados para casos normales y anomalías	29
3.3 Resultados para las siete clases de imágenes.....	30
Conclusiones y Recomendaciones.....	32
Bibliografía	33
Anexo 1. Pruebas estadísticas para las clases “célula anormal”- “célula normal”	37
A1.1 Prueba de Friedman: Normalidad = Anormal.....	37
A1.2 Prueba de Friedman: Normalidad = Normal.....	37
A1.3 Prueba de Rangos Signados de Wilcoxon: Normalidad = Anormal.....	38
A1.4 Prueba de Rangos Signados de Wilcoxon: Normalidad = Normal.....	39
Anexo 2. Pruebas estadísticas para las siete clases.....	40
A2.1. Prueba de Friedman: Clase = Carcinoma in Situ	40
A2.2. Prueba de Friedman: Clase = Displasia Severa	40
A2.3. Prueba de Friedman: Clase = Displasia Moderada	41
A2.4. Prueba de Friedman: Clase = Displasia Leve	42
A2.5. Prueba de Friedman: Clase = Columnar Normal.....	42
A2.6. Prueba de Friedman: Clase = Intermedia Normal	43
A2.7. Prueba de Friedman: Clase = Superficial Normal	43
A2.8. Prueba de Rangos Signados de Wilcoxon: Clase = Carcinoma in Situ	44
A2.9. Prueba de Rangos Signados de Wilcoxon: Clase = Displasia Severa	45
A2.10. Prueba de Rangos Signados de Wilcoxon: Clase = Displasia Moderada	45

A2.11. Prueba de Rangos Signados de Wilcoxon: Clase = Displasia Leve	46
A2.12. Prueba de Rangos Signados de Wilcoxon: Clase = Columnar Normal.....	46
A2.13. Prueba de Rangos Signados de Wilcoxon: Clase = Intermedia Normal.....	47
A2.14. Prueba de Rangos Signados de Wilcoxon: Clase = Superficial Normal	48

Introducción

En el curso de los últimos años se ha desarrollado un acelerado crecimiento del número de técnicas referentes a imágenes digitales que han sido introducidas en la práctica médica. Este desarrollo viene de la mano con las capacidades y gran factibilidad de ejecución de las computadoras modernas, lo que hace posible que las imágenes en la medicina, que eran grabadas sobre películas tradicionalmente, ahora puedan tratarse en forma digital [ZHO09, PLI11, MAR09, LEB07, KAL10, JUN10, CLO10].

Un papel importante en las tecnologías de la información y de cómputo lo ha adquirido el Procesamiento Digital de Imágenes (PDI). En la actualidad es la base de una creciente variedad de aplicaciones dentro de las que se incluye la diagnosis médica [ZHO09, PLI11, MAR09, LEB07, KAL10, JUN10, CLO10].

El PDI es el conjunto de técnicas y procesos para descubrir o hacer resaltar información contenida en una imagen usando como herramienta principal una computadora. Su aplicación primordial se basa en el estudio y mejoramiento de la calidad de la información contenida en una imagen con el fin de que esta información pueda ser interpretada.

En la actualidad la importancia del procesamiento digital de imágenes para el desarrollo de la medicina moderna es inmensa, pues con el mismo se logran con éxito detecciones de patologías que pueden ser tratadas posteriormente por los especialistas. Quizás uno de los ejemplos de mayor importancia y uso del PDI resulta su utilización en las imágenes de las pruebas de Papanicolaou para detectar anomalías, llamada así en honor de Georgios Papanicolaou, médico griego que fue pionero en citología y detección temprana de cáncer [PAP42]. Si la prueba de Papanicolaou se realiza regularmente, el cáncer cérvico-uterino es una enfermedad prevenible en el 90% de los casos. Con el estudio preciso, realizado por expertos, de estas imágenes del útero de la mujer podríamos obtener un mejor y más temprano diagnóstico y no solo se remitiría a una prueba de Papanicolaou, sino al estudio de una imagen en la que incluso se separan las regiones más importantes de la misma para descubrir otras posibles patologías.

Una rama importante del PDI es la segmentación [FRI10, GEE09a, PAR10]. Existen numerosos métodos de segmentación de imágenes dentro de los cuales se encuentran las medias desplazadas, el de las k -medias y el de las marcas de agua (*watersheds*).

La calidad de una segmentación depende de las características de la propia imagen y del método en sí. Es por esto que existen mecanismos o métodos para evaluar la calidad de una segmentación. Esta evaluación puede realizarse de dos formas: supervisada cuando contamos con un estándar dorado (*ground truth*) con el cual comparar, o de forma no supervisada cuando la evaluación depende de estadígrafos calculados a partir de la imagen y su segmentación (homogeneidad de las regiones, heterogeneidad entre regiones, grado de sobresegmentación, etc). En la literatura científica dentro del PDI se han reportado innumerables métodos no supervisados para evaluar la calidad de las segmentaciones [CHA06]. Utilizar uno u otro depende fuertemente del tipo de imagen que con la que se está trabajando.

La hipótesis de esta investigación es la siguiente:

Hipótesis de investigación:

- Existen uno o varios métodos no supervisados de evaluación de la calidad de las segmentaciones que superen en desempeño a otros, para imágenes de la prueba de Papanicolaou.

El presente trabajo consta de los siguientes objetivos:

Objetivo General:

- Proponer uno o varios métodos no supervisados de evaluación de la calidad de las segmentaciones para la evaluación de la calidad de segmentaciones de imágenes de la prueba de Papanicolaou.

Objetivos Específicos:

- Evaluar el desempeño de varios métodos no supervisados de evaluación de la calidad de segmentaciones de imágenes de la prueba de Papanicolaou.
- Proponer uno o varios métodos como superiores a la hora de evaluar la calidad de segmentaciones de imágenes de la prueba de Papanicolaou.

- Demostrar estadísticamente que el método propuesto (o los métodos propuestos) es superior (son superiores) para la evaluación de la calidad de segmentaciones de imágenes de la prueba de Papanicolaou.

El presente informe se encuentra estructurado como sigue: tres capítulos, conclusiones y recomendaciones, referencias bibliográficas y anexos. En el capítulo 1 se describen algunos de los métodos de evaluación de la calidad de segmentación. En el segundo capítulo se describe la base de casos de imágenes de la prueba de Papanicolaou utilizada y los métodos realizados para la evaluación. En el capítulo 3 se realiza un análisis de los resultados obtenidos para cada método utilizado. Los anexos presentan en profundidad las pruebas estadísticas realizadas.

Capítulo 1. Métodos para la evaluación de la calidad de segmentaciones

1.1 Segmentación de imágenes

La segmentación es el proceso de partición de una imagen en regiones no intersectadas, de manera tal que cada región r_i es homogénea y la unión de dos regiones adyacentes no es homogénea [FRI10, GEE09a, PAR10]. Formalmente, la segmentación de una imagen f en l regiones se define como el mapeado de la imagen en el dominio Ω a los índices I_{s_l} :

$$S_l f : \Omega \mapsto I_{s_l} = \{1, 2, \dots, l\}. \quad (1.1)$$

La segmentación debe ser completa, es decir cada punto del dominio de la imagen tiene que pertenecer a exactamente una región:

$$\bigcup_{i \in I_{s_l}} r_i = \Omega \quad (1.2)$$

$$r_i \cap r_j = \emptyset, \quad \forall i, j \in I_{s_l}, i \neq j$$

donde r_i representa una región (segmento) que es a su vez un número de píxeles conectados.

La homogeneidad se define a través de un predicado de uniformidad H . La segmentación $S_l f$ debe cumplir:

$$H r_i = \text{verdadero} \quad \forall i \in I_{s_l} \quad (1.3)$$

$$H r_i \cup r_j = \text{falso} \quad \forall i, j \in I_{s_l}; r_i \text{ y } r_j \text{ son adyacentes} \quad (1.4)$$

Resulta común que ocurra una segmentación alta (*oversegmentation*) o segmentación baja (*undersegmentation*) según se cumplan las condiciones expuestas en los predicados de las ecuaciones (1.3) y (1.4) (Tabla 1.1). La figura 1.1 muestra un ejemplo de segmentación de una imagen de microscopía celular.

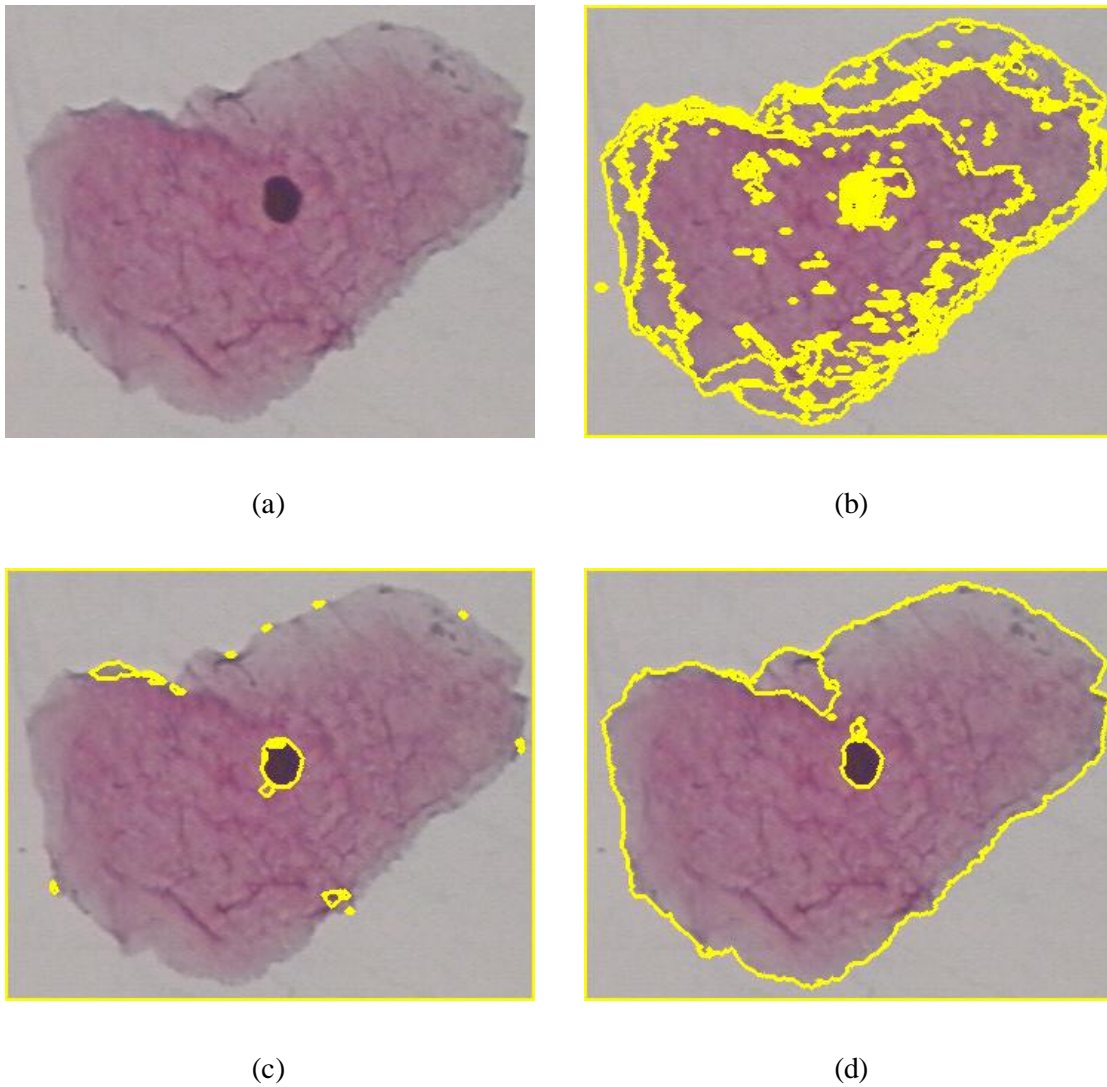


Figura 1.1. Segmentación de imágenes. (a) Imagen original. (b) Segmentación alta (*Oversegmentation*). (c) Segmentación baja (*Undersegmentation*). (d) Segmentación “adecuada”.

Tabla 1.1. Segmentación alta o baja de acuerdo al cumplimiento de los predicados de las ecuaciones (1.3) y (1.4).

	Predicado de la Ec. (1.3)	Predicado de la Ec. (1.4)
Segmentación alta	Se cumple	No se cumple
Segmentación baja	No se cumple	Se cumple
Segmentación “adecuada”	Se cumple	Se cumple

1.1.1 El método de las marcas de agua (Watershed).

El concepto de marcas de agua (*watersheds*, en inglés) es basado en visualizar una imagen en tres dimensiones: dos coordenadas espaciales *vs* niveles de gris. En tal interpretación "topográfica" (figura 1.2) consideramos tres tipos de puntos: (a) los puntos pertenecientes a un mínimo regional; (b) puntos tales que, si ponemos sobre ellos una gota de agua, esta se deslizaría hasta un mínimo bien definido; y (c) puntos tales que, si ponemos sobre ellos una gota de agua, esta se deslizaría hasta más de un mínimo. Para un mínimo regional particular, el conjunto de puntos que cumplen con la condición (b) reciben el nombre de “cuenca de captación” (*catchment basin* o *watershed*) de ese mínimo. Los puntos que satisfacen la condición (c) forman “crestas” en la superficie topográfica, y se denominan “líneas de división” o “líneas de *watershed*”.

El objetivo principal de los algoritmos de segmentación basados en estos conceptos es encontrar las líneas de *watershed*. Para entender esta transformada se considera una imagen en escala de gris como un relieve terrestre. Las intensidades de gris de mayor amplitud se corresponden con llanuras o montañas, mientras que las de menor intensidad corresponden a valles y ríos. La idea básica es simple: Suponga que un agujero es perforado en cada mínimo regional y que la topografía entera es inundada desde abajo para dejar al agua elevarse por los agujeros con una proporción uniforme (algoritmo de inundación). Cuando el agua ascendente en las distintas cuencas de captación está a punto de combinarse, un dique es fabricado para impedir la fusión. La inundación alcanzará finalmente un paso cuando sólo las partes superiores de los diques sean visibles sobre la línea de agua. Estos límites de dique corresponden a las líneas divisoras de los *watersheds*. Por lo tanto, son los límites (continuos) los extraídos por el algoritmo de segmentación de *watershed*.

Existen diferentes definiciones formales de las *watersheds* morfológicas, entre ellas:

- Para casos continuos.
- Para casos discretos.
- Definición algorítmica mediante un proceso de inmersión.
- Definición mediante distancias topográficas

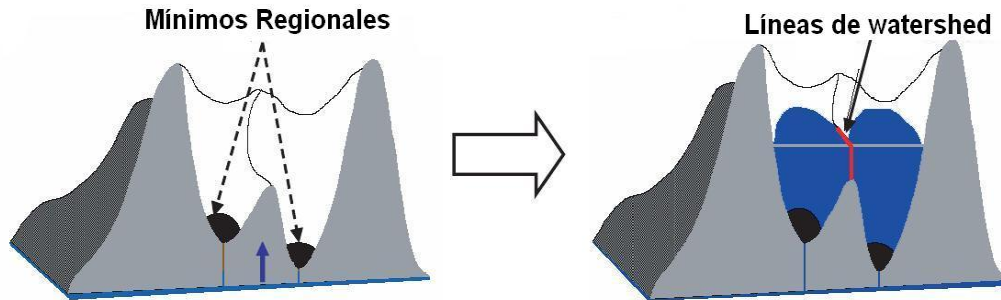


Figura 1.2. Transformada *watershed*.

La base de la transformada *watershed* consiste en modelar las imágenes en escala de grises como superficies topológicas, donde los valores de $f(x, y)$ sean considerados como alturas. Existen varios enfoques teóricos y algoritmos de computación para el cálculo de la transformada en imágenes, los cuales tienen en común como objetivo final encontrar los píxeles que componen las crestas.

Los algoritmos denominados *waterfall* consisten en suponer que cae agua desde arriba proveniente desde la posición de cada píxel de la imagen. El agua de lluvia recorre la montaña pendiente abajo hasta llegar a los ríos inundando los valles, quedando denominados inicialmente los mínimos regionales de la topografía. En estos puntos el agua no tiene dirección posible hacia abajo y queda estancada. El camino tridimensional descendiente más corto es el que tomaría el agua naturalmente [FRI10, GEE09a, PAL99, VAN06, LIU04, HOU12].

Este procedimiento se realiza una sola vez, convirtiendo a este algoritmo en no iterativo. Sin embargo, para cada píxel es necesario determinar todos los caminos posibles hacia los mínimos regionales dentro del relieve tridimensional, lo que incrementa significativamente

el costo computacional de la transformada. Si bien en la actualidad la implementación del algoritmo *waterfall* se ha simplificado, este tipo de algoritmos no constituyen una herramienta eficiente para ser aplicado a imágenes médicas, ya que estas poseen una gran resolución espacial.

La transformada *watershed* se aplica por lo general a tres tipos de imágenes: la imagen original en escala de grises, la imagen del gradiente y la imagen correspondiente al negativo de la transformada de distancia.

1.1.2 El método de las medias desplazadas.

El método de las medias desplazadas (*mean shift*, en inglés) es muy popular para segmentar imágenes y videos. Los pixeles son representados por puntos de rasgos, y la segmentación es manejada por la densidad del punto en el espacio de rasgos: las medias desplazadas se basa en un esquema iterativo para determinar las modas en una función de densidad probabilística [COM02, FRI10, GEE09a, VAN06].

La estimación del núcleo (*kernel*) de densidad (*Kernel Density Estimation*) es el método de estimación de densidad más usado. Dados n puntos x_i , $i = 1, \dots, n$ en el espacio d -dimensional \mathbb{R}^d , el estimador del núcleo de densidad multivariado $K(x)$ y la matriz de ancho de banda H simétrica y definida positiva, vienen definidos por:

$$f(x) = \frac{1}{n} \sum_{i=1}^n K_H(x - x_i) \quad (1.5)$$

donde:

$$K_H(x) = H^{-\frac{1}{2}} K(H^{-\frac{1}{2}} x) \quad (1.6)$$

El núcleo $K(x)$ es una función que satisface las siguientes cuatro propiedades:

$$\int_{\mathbb{R}^d} K(x) dx = 1 \quad (1.7)$$

$$\int_{\mathbb{R}^d} x K(x) dx = 0 \quad (1.8)$$

$$\lim_{x \rightarrow \infty} x^d K(x) = 0 \quad (1.9)$$

$$\int_{\mathbb{R}^d} x x^T K(x) dx = c_K \mathbf{I} \quad (1.10)$$

Así que el estimador del núcleo de densidad puede escribirse como:

$$f_{K_S}(x) = \frac{1}{nh^d} \sum_{i=1}^n K_S\left(\frac{x - x_i}{h}\right) \quad (1.11)$$

Donde el parámetro h es el radio de la ventana usada por el núcleo K_S . El primer paso del análisis de un espacio que tiene una densidad $f(x)$ es encontrar las modas de esta densidad probabilística, que se localizan en los ceros del gradiente $\nabla f(x) = 0$: el procedimiento de medias desplazadas es una manera elegante de localizar dichos ceros sin estimar la densidad.

Al finalizar el proceso, el vector de medias desplazadas multivariado en el punto x viene dado por:

$$m_K(x) = \frac{\sum_{i=1}^n x_i K\left(\frac{x - x_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)} - x \quad (1.12)$$

1.1.3 Segmentación de imágenes usando el método de las k medias.

El algoritmo de las k-medias es un método de formación de conglomeración o clusterización [FOR65, DUD00]. Estos métodos toman un grupo de objetos y crean grupos, conglomerados o clústeres de forma tal que aquellos objetos que pertenezcan al mismo clúster son similares y aquellos que no se encuentran en el mismo clúster presentan marcadas disimilitudes.

Supongamos que tenemos un conjunto \mathbf{S} de n objetos descritos por n vectores reales d -dimensionales $\{X_1, X_2, \dots, X_n\}$, $X_i = (X_{i1}, X_{i2}, \dots, X_{id})$, $1 \leq i \leq n$. El algoritmo de las k -medias trata de encontrar una partición de \mathbf{S} en $k < n$ conjuntos $S = \{S_1, S_2, \dots, S_k\}$ de forma tal que minimice las distancias de cada objeto al centro del clúster al que pertenece.

Esto se logra mediante la resolución del siguiente problema de optimización:

$$WCSS = \min_S \sum_{i=1}^k \sum_{X_j \in S_i} D(X_j, \mu_i) \quad (1.13)$$

donde D es una medida de distancia entre los vectores X_j y μ_i , y μ_i es el centro del conjunto S_i .

El algoritmo de las k -medias consta de dos fases [FOR65]. La fase 1 (*Batch Update*) reasigna todos los objetos a los centros de clúster más cercanos. Al final, los centros de clúster se recalculan. La fase 2 (*Online Update*) reasigna todos los objetos a los centros de clúster más cercanos sólo si reducen el valor de la función que se está minimizando. Al final los centros de clúster se recalculan.

La fase 1 es rápida pero en ocasiones no converge ni tan siquiera a un mínimo local. Se usa ella sola siempre que estemos trabajando con pocos datos. La fase 2 es más lenta pero siempre converge a un mínimo local. La convergencia al mínimo global sólo se logra usando una búsqueda exhaustiva.

El algoritmo de las k -medias se puede aplicar a la segmentación de imágenes a color simplemente asumiendo que cada pixel es el equivalente a un objeto que se describe mediante su color. Estos descriptores pudieran ser las componentes R , G y B del espacio de color RGB , las componentes L , a y b del espacio de color Lab , entre otras.

1.2 Métodos supervisados para la evaluación de la calidad de segmentación

1.2.1 Método de Vinet.

El método de Vinet [VIN91] es una medida supervisada ampliamente usada para cuantificar la distancia entre dos segmentaciones A y B, con m y n regiones, respectivamente, de una imagen con N pixeles. Primero, se computa una tabla de superposición: $T_{ij} = A_i \cap B_j$ con $0 \leq i \leq m$, y $0 \leq j \leq n$. El máximo de esta matriz da las dos regiones más similares extraídas de A y B, respectivamente. El criterio de similitud es definido por: $C_0 = \max(T_{ij})$ con $0 \leq i \leq m$, y $0 \leq j \leq n$. La búsqueda del segundo máximo (sin tener en cuenta las dos últimas regiones) da el criterio de similitud C_1 y así sucesivamente hasta C_{k-1} , donde $k = \min(m, n)$. La medida de no similitud entre las dos segmentaciones A y B es dada por:

$$D_{A,B} = 1 - \frac{1}{N} \sum_{i=0}^{k-1} C_i \quad (1.14)$$

Cuando una de las segmentaciones (B, por ejemplo) es la hecha a mano entonces bajos valores de la medida de Vinet significan una segmentación buena (cerca de la referencia).

El algoritmo húngaro [MUN57] se puede usar en lugar del enfoque ávido para calcular los valores de C_i , $i = 1 \dots \min(n, m)$. El algoritmo húngaro aumenta el costo computacional pero también la exactitud de la medida.

La figura 1.3 muestra dos segmentaciones y la matriz de superposición generada por ellas.

1.3 Métodos no supervisados para la evaluación de la calidad de segmentación

1.3.1 Métodos de Levine.

Uno de los criterios más intuitivos, capaz de cuantificar la calidad de una segmentación, es la uniformidad de intrarregiones. Basado en esta idea de uniformidad de intrarregiones, Levine y Nazif [LEV85] definieron un criterio que calcula la uniformidad de una región:

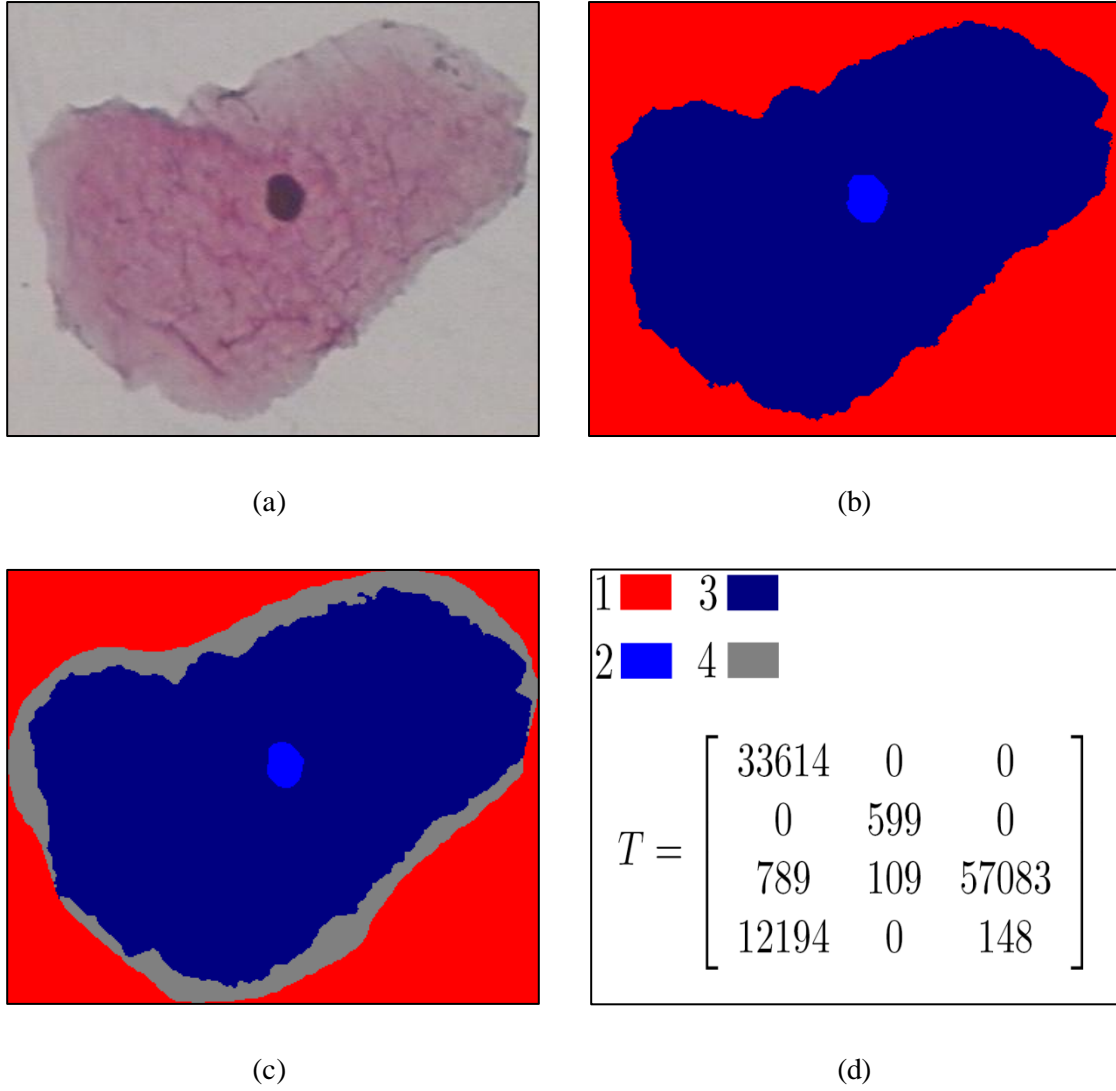


Figura 1.3. Distancia de Vinet. (a) Imagen original. (b) Segmentación. (c) Segmentación de referencia (*Ground truth*). (d) Tabla de superposición.

$$LEV1 \ I_r = 1 - \frac{1}{Card \ I} \sum_{k=1}^{N_R} \frac{\sum_{s \in R_k} g_I(s) - \sum_{t \in R_k} g_I(t)}{(\max_{s \in R_k} g_I(s) - \min_{s \in R_k} g_I(s))^2} \quad (1.15)$$

donde I_r corresponde al resultado de segmentación de la imagen I en un juego de regiones

$R = \{R_1, \dots, R_{N_R}\}$ que tiene N_R regiones, $Card(I)$ corresponde al número de píxeles de la imagen I , y $g_I(s)$ corresponde a la intensidad de niveles de gris del píxel s de la imagen I y se puede generalizar a cualquier otra característica (color, textura).

Otra medida de uniformidad estandarizada es la propuesta por Sezgin y Sankur [SEZ04]. Basada en el mismo principio, la medida de homogeneidad de Cochran [COC54] nos brinda una medida confiable de la homogeneidad de una región. Sin embargo, este método requiere una selección de umbral que es a menudo arbitrariamente hecha, limitando así el método propuesto. Otro criterio para medir la uniformidad de las intrarregiones desarrollado por Pal y Pal [PAL89], basado en un umbral que maximiza la entropía local de segundo orden de las regiones en la segmentación.

Complementario a la uniformidad de intrarregiones, Levine y Nazif [LEV85] definen una medición de disparidad entre dos regiones para evaluar la disimilitud de regiones en una segmentación. La fórmula de disparidad total de interregiones es definida por:

$$LEV2 \ I_r = \frac{\sum_{k=1}^{N_R} (w_{R_k} \sum_{j=1/R_j \in W(R_k)} [p_{R_k \setminus R_j} (g_I R_k - g_I R_j / (g_I R_k + g_I(R_j)))])}{\sum_{k=1}^{N_R} w_{R_k}} \quad (1.16)$$

donde w_{R_k} es un peso asociado a R_k que puede ser dependiente de su área, por ejemplo, g_k es el promedio de los niveles de gris de R_k . $g_I R_k$ se puede generalizar a un vector del rasgo que calcula los valores de los píxeles de la región R_k tal como por LEV1. El término $p_{R_k \setminus R_j}$ corresponde a la longitud del perímetro de la región R_k común al perímetro de la región R_j . Este tipo de criterio tiene la ventaja de penalizar la segmentación alta.

1.3.2 Método de Rosenberger para imágenes no texturizadas.

El método de Rosenberger para imágenes no texturizadas (ROS1) [ROS99], combina las disparidades de las intra e interregiones. La disparidad de la intrarregión es calculada por la desviación estándar normalizada de los niveles de gris en cada región. La disparidad de la interregión calcula la disimilitud del nivel de gris promedio de dos regiones vecinas en la segmentación.

Según Haralick y Shapiro [HAR85], para que una segmentación sea considerada como “buena” debe cumplir que:

- (i) las regiones deben ser uniformes y homogéneas,
- (ii) el interior de las regiones debe ser simple sin demasiados agujeros pequeños,
- (iii) las regiones adyacentes deben presentar significativas diferencias de valores para las características uniformes,
- (iv) se deben suavizar los límites correctamente.

De ahí que el criterio de Rosenberger para imágenes no texturizadas (ROS1) venga dado por:

$$ROS1 I_R = \frac{1 - \frac{4}{255^2} N_R \sum_{i=1}^{N_R} \sigma^2 R_i}{2} + \frac{\frac{2}{N_R(N_R-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^{N_R} \frac{g_I R_k - g_I R_j}{512}}{2} \quad (1.17)$$

donde $\frac{2}{N_R(N_R-1)}$ es el número de combinaciones de 2 regiones, $g_I R_j$ es el nivel de gris promedio de R_j , y $\sigma^2 R_i$ es la variación del nivel de gris de R_i .

Valores altos del criterio de Rosenberger1 indican una buena segmentación.

1.3.3 Método de Borsotti.

En algunos métodos [LIU94], en el caso de imágenes muy ruidosas, se debe penalizar el exceso de regiones. No obstante, el error generado por cada región pequeña (uniformidad intrarregión) está cerca de 0. Por consiguiente, el criterio global está también cerca de 0, indicando de manera errónea que la segmentación es muy buena. Borsotti *et al.* [BOR98] identificaron esta limitación del criterio de evaluación de Liu y Yang y lo modificaron, de forma que la penalización a las segmentaciones que presentan muchas regiones pequeñas es más estricta. Estas modificaciones permiten hacer el criterio más sensible a variaciones pequeñas de la segmentación:

$$BOR(I_R) = \frac{\overline{N_R}}{10^4 \times Card(I)} \sum_{k=1}^{N_R} \frac{E_k^2}{1 + \log Card R_k} + \frac{\chi(Card R_k)}{Card R_k} \quad (1.18)$$

El término $Card()$ es el tamaño (área) o de la región R_k o de la imagen I ; $\chi(Card R_k)$ es el número de regiones que tienen el mismo tamaño (área) que la región R_k ; y E_k es la suma

de las distancias euclidianas entre el vector de color RGB de los píxeles de R_k y los atributos del vector de color para la región R_k en la segmentación.

Cuando estamos en presencia de valores pequeños del criterio de Borsotti significa que obtuvimos una buena segmentación. Para armonizar con los otros criterios, usamos el valor de $1 - BOR(I_R)$ como criterio de la calidad de segmentación.

1.3.4 Homogeneidad de color.

El método Homogeneidad de Color, al cual nos referiremos como (*CH*) [GEE09b] por sus siglas en inglés (*Color Homogeneity*) combina la medida de homogeneidad de la región con una medida de contraste de la interregión:

$$CH(I_r) = \frac{\sum_{k=1}^{N_R} C(R_k) / (N_R - 1)}{\sum_{k=1}^{N_R} H(R_k) / (Card I - N_R)} \quad (1.19)$$

Para una imagen I , $Card I$ denota el área de la imagen completa (número de píxeles), $H(R_k)$ es la medida de homogeneidad de la región (error local de color) de la región R_k mientras, $C(R_k)$ es una medida de separación de interregiones (diferencia de color entre regiones adyacentes), que expresa el cambio entre la precisión de la frontera de una región y la diferencia entre la región y sus vecinos.

CH penaliza esas regiones que tienen una diferencia de color grande y un contraste de frontera bajo. Si la diferencia de color dentro de la región es bajo y las regiones adyacentes son significativamente diferentes (alto contraste de frontera), su valor es alto, mientras más alto sea el valor de *CH* mejor será la segmentación.

La tabla 1.2 muestra la evaluación de la calidad de segmentación para las segmentaciones de la figura 1.1.

Tabla 1.2. Métodos no supervisados para la evaluación de la calidad de segmentación para las segmentaciones de la figura 1.1.

	LEV1	LEV2	ROS1	BOR	CH
Fig. 1.1 (b)	0.9889	0.1264	0.5380	0.9884	95001
Fig. 1.1 (c)	0.9865	0.0576	0.5277	0.9603	4165.9
Fig. 1.1 (d)	0.9651	0.1501	0.5336	0.8986	541.6

Capítulo 2. Materiales y Métodos

2.1 El cáncer cérvicouterino y la prueba de Papanicolaou

El cáncer cérvicouterino es uno de las enfermedades que afecta a miles de mujeres cada año. El cáncer de cérvix es el segundo tipo de cáncer más frecuente en las mujeres en todo el mundo y es causado por una infección de transmisión sexual conocida como el virus del papiloma humano [JEM06].

A través del uso de la prueba de Papanicolaou [MEI98] se ha reducido considerablemente el número de mujeres que mueren a causa de esta enfermedad. Aproximadamente 1 de cada 4 mujeres contraerá los tipos de este virus relacionados con el cáncer cervical en algún momento de su vida. No obstante, hoy en día, sólo 1 de cada 1000 mujeres que tienen el virus desarrollan cáncer cervical. Esto se debe a que muchas mujeres saben cómo prevenir el cáncer cervical realizándose estas pruebas periódicamente [JEM06].

Esta prueba se usa para descubrir cambios malignos y pre-malignos en la cérvix uterina. La prueba es un examen citológico en el que se toman muestras de células epiteliales en la zona de transición del cuello uterino, en busca de anormalidades celulares que orienten a (y no que diagnostiquen) la presencia de una posible neoplasia (proceso de proliferación anormal de células) del cuello uterino.

La cérvix es la porción más baja del útero y está cubierta por una capa delgada de células llamada epitelio. Las células del epitelio tienen dos formas diferentes: escamosa y columnar. Estas están localizadas en áreas separadas: área escamosa y área columnar (figura 2.1). El área escamosa se localiza al fondo del canal de la cérvix (figura 2.1) y aquí las células se dividen en 4 capas: basal, parabasal, intermedia y superficial. Cuando las células maduran se mueven a través de las capas y van cambiando de forma, color y otras características, el citoplasma aumenta su tamaño y el núcleo lo disminuye.

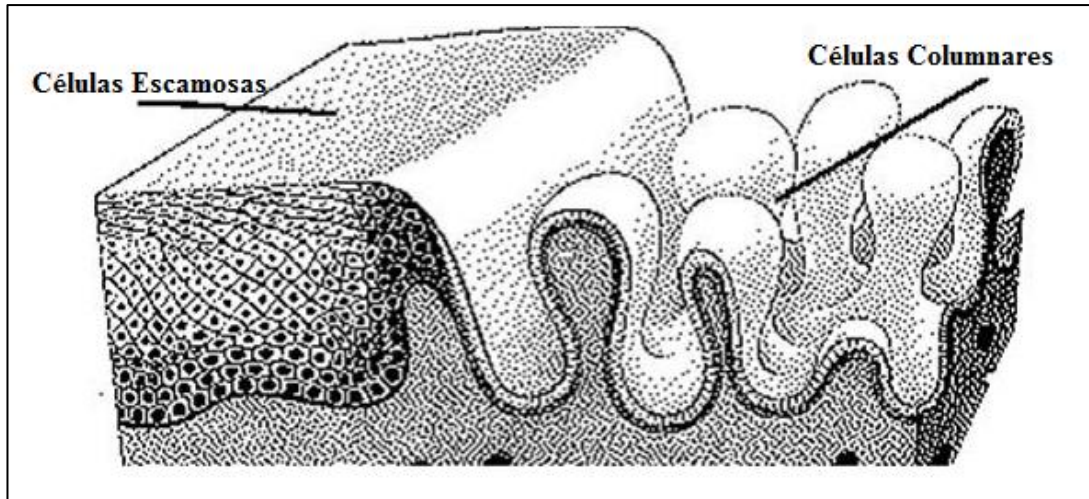


Figura 2.1 Células escamosas y columnares.

El área columnar [MEI98, BJE01] se localiza más arriba, en el canal de la cérvix (figura 2.2). Las células columnares solo existen en la capa basal, tienen forma de columnas como su propio nombre lo indica con un citoplasma oblongo y un núcleo largo localizado en un extremo. La unión entre estas dos áreas se denomina unión escamocolumnar, y puede estar localizada dentro o fuera de la cérvix. La unión escamocolumnar es llamada también como zona de transición porque las células se transforman de una forma a otra.

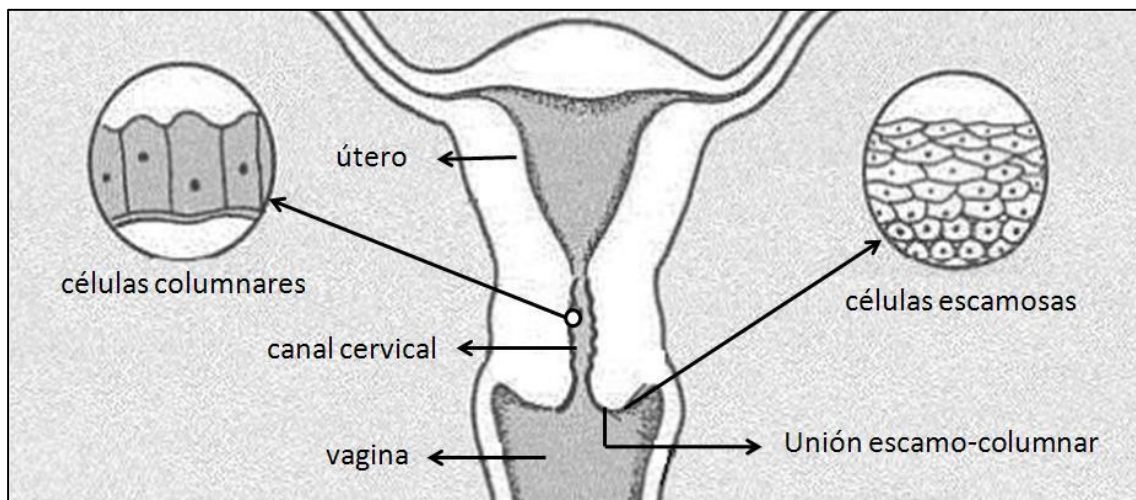


Figura 2.2 Detalles del útero y localización de las células escamosas y columnares.

Para la recolección de la muestra citológica se coloca a la paciente en posición ginecológica y se expone correctamente el cuello del útero utilizando un espéculo sin emplear lubricantes, se retira el exceso de secreción o mucus, si fuera necesario, sin tocar la superficie del cuello. El raspado generalmente se realiza en la línea de la unión escamo-

columnar donde se unen los dos epitelios. Cuando esta línea coincide con el orificio externo del cuello, se introduce el extremo saliente de la espátula de Ayre en el mismo y se hace girar 360 grados en el sentido de las manecillas del reloj, con cierta presión.

Las muestras utilizadas para esta prueba se toman de tres sitios:

1. Endocérvix, que es el orificio que comunica con el útero;
2. Cérvix, que es la parte más externa del útero, y que comunica directamente con la vagina
3. Tercio superior de la vagina, que es la región que rodea el cuello del útero.

Esta prueba deben realizársela todas las mujeres que hayan iniciado su vida sexual una vez al año durante dos o tres años consecutivos, y si los resultados son negativos se repetirá cada tres a cinco años en caso de no haber factores de riesgo y hasta los 65 años.

Aunque se trata de una prueba con baja sensibilidad (50-60%) y una alta tasa de falsos negativos (30%), en muchos casos se trata de errores en la toma de muestra o del laboratorio. Puede mejorarse la técnica usando la citología en medio líquido que consiste en diluir la muestra en una solución fijada para su procesamiento [PAY00].

La prueba de Papanicolaou es un examen barato, fácil y simple que cualquier médico o enfermera puede realizar y que contribuye eficazmente al diagnóstico precoz del cáncer del cuello uterino. La muestra obtenida debe ser analizada por médicos especializados en anatomía patológica con entrenamiento específico. Su uso ha reducido las muertes por cáncer de cérvix en más del 50% [JEM06].


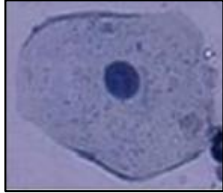

2.2 Base de casos Herlev

La base de casos utilizada fue desarrollada por el Hospital Universitario Herlev, el departamento de Patología y el departamento de automatización en tecnología universitaria de Dinamarca. La misma contiene un total de 917 muestras, distribuidas en 7 clases diferentes, 3 normales y 4 anormales (Tabla 2.1) [JAN05, MAR03, NOR05]. La figura 2.3 muestra la forma general y las características de las células en cada una de las clases. El propósito de la prueba es diagnosticar los cambios de las células pre-malignas antes de que estas se conviertan en cáncer.

Tabla 2.1. Células utilizadas en la construcción de la base de casos.

Tipo de células	No. de casos	Clasificación
Epitelial escamoso (Superficial)	74	Normal
Epitelial escamoso (Intermedio)	70	Normal
Epitelial columnar	98	Normal
Displasia leve escamosa no queratinizante	182	Anormal
Displasia moderada escamosa no queratinizante	146	Anormal
Displasia severa escamosa no queratinizante	197	Anormal
Carcinoma escamoso <i>in situ</i> (Intermedio)	150	Anormal
Total	917	

Cada célula de la base de casos se describe por 20 rasgos, según sus creadores. De los 20 rasgos, 6 son rasgos de intensidad/color y los 14 restantes son rasgos morfométricos o relacionados con la forma.

CÉLULAS NORMALES	
<p>1. Escamoso Superficial</p> <ul style="list-style-type: none"> • Forma: llana u ovalada. • Núcleo muy pequeño. • Relación núcleo-citoplasma muy pequeña. 	
<p>2. Escamoso Intermedio</p> <ul style="list-style-type: none"> • Forma: redonda. • Núcleo largo. • Relación núcleo-citoplasma pequeña. 	
<p>3. Columnar</p> <ul style="list-style-type: none"> • Forma: redonda. • Núcleo largo. • Relación núcleo-citoplasma media. 	


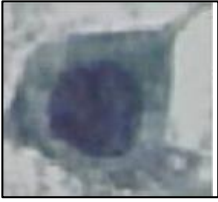


CÉLULAS ANORMALES	
<p>4. Displasia Leve</p> <ul style="list-style-type: none"> • Núcleo claro o largo. • Relación núcleo-citoplasma media. 	
<p>5. Displasia Moderada</p> <ul style="list-style-type: none"> • Núcleo largo u oscuro. • Citoplasma oscuro. • Relación núcleo-citoplasma grande. 	
<p>6. Displasia Severa</p> <ul style="list-style-type: none"> • Núcleo largo, oscuro o deforme. • Citoplasma oscuro. • Relación núcleo-citoplasma muy grande. 	
<p>7. Carcinoma in situ</p> <ul style="list-style-type: none"> • Núcleo largo, oscuro o deforme. • Relación núcleo-citoplasma muy grande. 	

Figura 2.3 Tipos de células y sus características obtenidas a través de la prueba de Papanicolaou.

2.3 Evaluación de los métodos no supervisados de calidad de segmentaciones

La evaluación de los métodos no supervisados de calidad de segmentaciones se realizó usando el método propuesto en [CHA06]. En este método se buscan los criterios de evaluación que tienen un comportamiento similar a la distancia de Vinet.

Consideramos los resultados de comparación de las segmentaciones para todos los criterios de evaluación. Como tenemos cuatro segmentaciones de cada imagen (3 métodos de

segmentación más el *ground truth*), obtenemos 6 posibles comparaciones. Las 6 posibles comparaciones de cuatro resultados de segmentación A, B, C, y D son:

1. A con B,
2. A con C,
3. A con D,
4. B con C,
5. B con D,
6. C con D.

A partir de estas comparaciones se definen las siguientes funciones:

$A_{i,k}$: es 1 si, para la k -ésima imagen y de acuerdo al método no supervisado de evaluación de la calidad de la segmentación, la primera segmentación de la comparación i es mejor que la segunda. En otro caso es 0.

$B_{i,k}$: es 1 si, para la k -ésima imagen y de acuerdo a la distancia de Vinet, la primera segmentación de la comparación i es mejor que la segunda. En otro caso es 0.

Por ejemplo, supongamos que analizamos al método de Rosenberger. Si $A_{4,25} = 1$ esto quiere decir que de acuerdo al método de Rosenberger la segmentación B de la imagen 25 es mejor que la segmentación C de esta misma imagen. Del mismo modo si $B_{4,25} = 1$ esto quiere decir que de acuerdo a la distancia de Vinet, la segmentación B de la imagen 25 es mejor que la segmentación C de esta misma imagen.

A partir de estas funciones, para cada método no supervisado de evaluación de la calidad de las segmentaciones, se define la Similitud acumulada de Comparaciones Correctas (SCC) de la siguiente manera:

$$SCC = \sum_{k=1}^N CC_k = \sum_{k=1}^N \sum_{i=1}^6 A_{i,k} - B_{i,k} \quad (2.1)$$

donde N es la cantidad de imágenes y CC_k son las comparaciones correctas para la imagen k , $CC_k = \sum_{i=1}^6 A_{i,k} - B_{i,k}$.

Con el objetivo de cuantificar la eficiencia de los criterios de evaluación, definimos la tasa de similitud de comparaciones correctas (SRCC) para cada método no supervisado de evaluación de la calidad de las segmentaciones:

$$SRCC = 1 - \frac{SCC}{SCC_{max}} \cdot 100 \% \quad (2.2)$$

donde $SCC_{max} = 6 \times N$ resultados de comparación.

Valores de $SRCC$ cercanos al 100% indican que el método no supervisado de evaluación de la calidad de las segmentaciones y la distancia de Vinet tienen un desempeño similar.

Capítulo 3. Resultados y Discusión

En el presente trabajo se estudiaron las siguientes medidas no supervisadas para evaluar la calidad de las segmentaciones:

- Método de Borsotti (Bor).
- Homogeneidad de Color (CH).
- Primer método de Levine (Lev1).
- Segundo método de Levine (Lev2).
- Método de Rosenberger (Ros1).

Para cada una de ellas se calculó el *SRCC* utilizando 70 imágenes de la base de imágenes de la prueba de Papanicolaou Herlev. Se seleccionaron 10 imágenes por clase. Estas imágenes se segmentaron usando los métodos que se describen en la sección 1.1:

- Transformada *Watershed* con mezcla de regiones usando el algoritmo *Waterfall*.
- Medias desplazadas (*Mean Shift*).
- *K*-medias.

El *SRCC* para cada medida no supervisada para evaluar la calidad de las segmentaciones se calculó por separado en las siguientes tres situaciones:

1. Para todos los casos (Sección 3.1).
2. Para los casos normales y anormales (Sección 3.2).
3. Para cada una de las clases (Sección 3.3).

Las pruebas estadísticas fueron realizadas utilizando el programa SPSS para Windows, versión 16.0.

3.1 Resultados para todos los casos

La figura 3.1 muestra el *SRCC* calculado para los cinco métodos usando todos los casos. El mejor resultado se obtuvo usando el método de Rosenberger (69.52%).

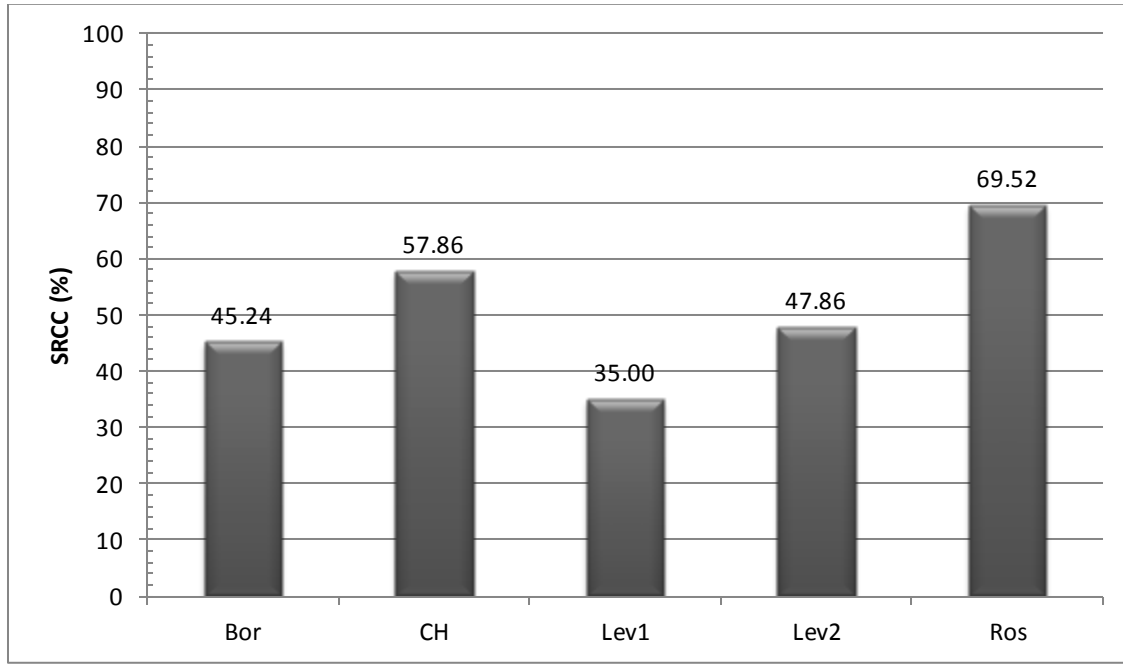


Figura 3.1 Resultados generales para la evaluación de la calidad de segmentación.

Para determinar si estas diferencias del método de Rosenberger con respecto a los demás son significativas se realizó un experimento no paramétrico usando la prueba de Friedman. En ella se compararon las comparaciones correctas (Ecuación 2.1) para los cinco métodos de estudio para las 70 imágenes del experimento.

Los resultados se muestran en la Tabla 3.1 y 3.2.

Tabla 3.1. Rangos

	Media de Rangos
Bor	3.41
CH	2.46
Lev1	4.01
Lev2	3.12
Ros1	2.00

Tabla 3.2. Prueba de Friedman

N	70.000
Chi-Cuadrado	83.653
gl	4.000
Significación Asintótica	.000

La prueba de Friedman nos asegura con un nivel de confianza del 95% que hay diferencias significativas en las comparaciones correctas para los 5 métodos no supervisados de evaluación de la calidad de las segmentaciones. Los mejores resultados se obtuvieron para Ros1 y CH. Las tablas 3.3 y 3.4 muestran los resultados de la prueba de los rangos signados de Wilcoxon para estos dos métodos.

Tabla 3.3. Rangos

	N	Media de Rangos	Suma de Rangos
Ros1 - CH Rangos Negativos	33 ^a	29.50	973.50
Rangos Positivos	18 ^b	19.58	352.50
Empates	19 ^c		
Total	70		

a. Ros1 < CH

b. Ros1 > CH

c. Ros1 = CH

Tabla 3.4. Prueba de los rangos signados de Wilcoxon

	Ros1 - CH
Z	-2.956 ^a
Significación Asintótica. (2-colas)	.003

a. Basado en los rangos positivos.

Los resultados concluyen con un nivel de significación de 0.05 que las comparaciones correctas para el método de Rosenberger son superiores que para los otros métodos.

3.2 Resultados para casos normales y anomalías

En la figura 3.2 se muestra el *SRCC* calculado por separado para células normales y anómalas. Todos los criterios (excepto Lev1) coinciden en que las segmentaciones de células normales tienen más calidad que las de las anomalías. Nuevamente resulta superior el criterio de Rosenberger para ambos casos (casos normales y anómalos). Resulta interesante constatar la marcada diferencia entre los resultados para células normales (81.67%) y anormales (60.42%) para este criterio.

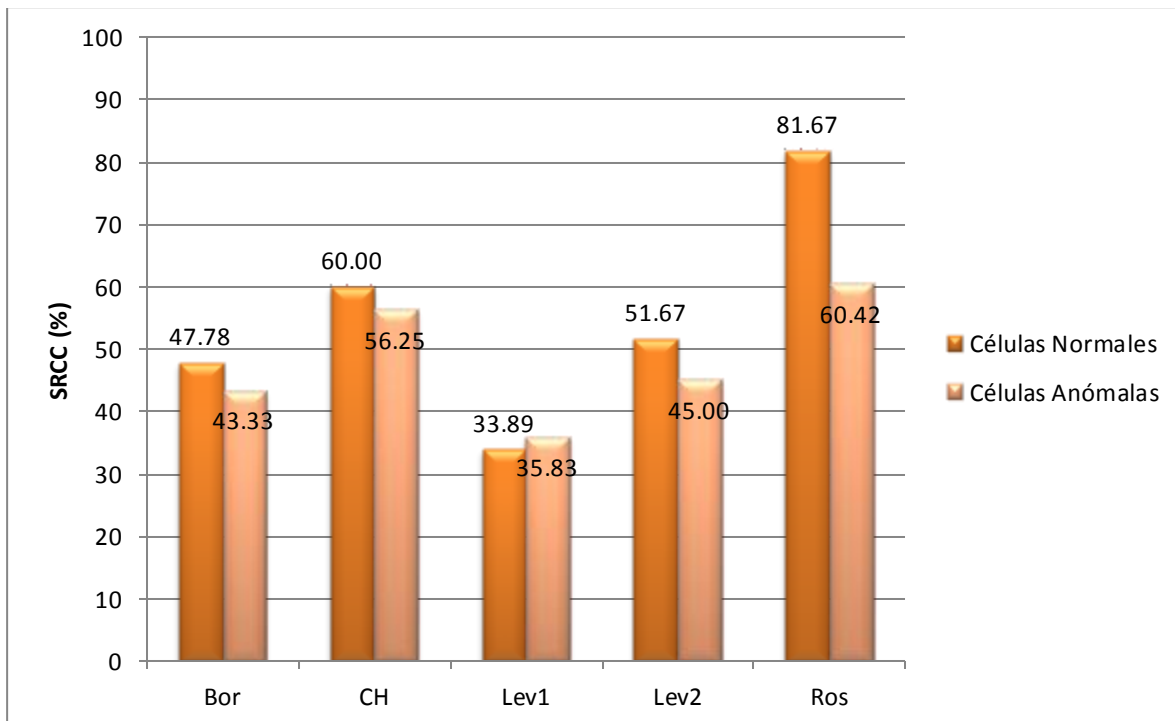


Figura 3.2 Resultados generales de evaluación de la calidad de segmentación para células normales y anomalías.

Las pruebas estadísticas para estos grupos se muestran en el Anexo 1. Para las células normales el mejor método es el método de Rosenberger. Sin embargo, no hay diferencias significativas para las células anómalas entre el método de Rosenberger y la homogeneidad de color. Todas estas pruebas se realizaron con nivel de significación de 0.05.

3.3 Resultados para las siete clases de imágenes

Como sucede en los resultados anteriores, cuando calculamos el *SRCC* para las clases de imágenes por separadas (Figura 3.3) destaca igualmente el criterio de Rosenberger. El método de Homogeneidad de Color es ligeramente superior a Rosenberger en las clases de displasia media y displasia severa.

Las pruebas estadísticas que avalan estos resultados se muestran en el Anexo 2. La significación asintótica de las pruebas de Friedman para cada una de las siete clases se muestra en la tabla 3.5.

Tabla 3.5. Significaciones asintóticas de las pruebas de Friedman y rangos signados de Wilcoxon para las siete clases de células.

	Significación Asintótica para la Prueba de Friedman	Significación Asintótica para la Prueba de los rangos signados de Wilcoxon entre Ros1 y CH
Normal Superficial	0.001	0.084
Normal Intermedia	0.000	0.007
Normal Columnar	0.029	0.572
Displasia Leve	0.173	-
Displasia Moderada	0.067	-
Displasia Severa	0.002	0.453
Carcinoma in Situ	0.026	0.098

El método de Rosenberger es estadísticamente superior en las clases:

- Normal Intermedia.

Sin embargo, en las restantes clases no hay diferencias significativas entre el método de Rosenberger y el de homogeneidad de color. Independientemente de lo interesante de estos resultados, no se deberían tomar al pie de la letra debido al tamaño reducido de las muestras usadas en este experimento.

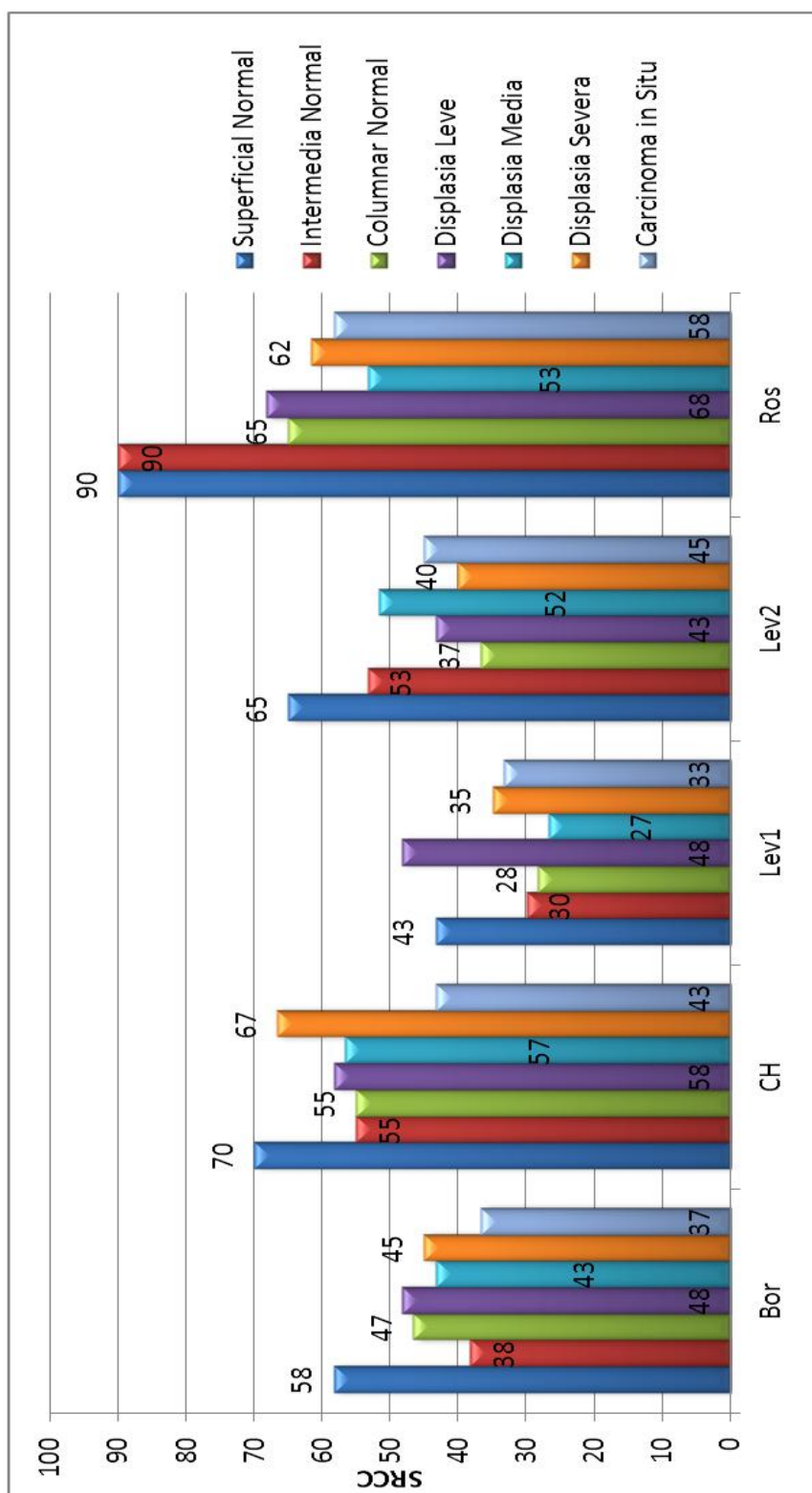


Figura 3.3 Resultados de evaluación de la calidad de segmentación para las siete clases de imágenes.

Conclusiones y Recomendaciones

A partir de los resultados alcanzados en esta investigación se llegó a las siguientes conclusiones:

1. Se demostró el mejor desempeño del Método de Rosenberger sobre los demás métodos no supervisados para evaluar la calidad de segmentación en imágenes de la prueba de Papanicolaou aplicados.
2. Se demostró estadísticamente (con un nivel de significación de 0.05) la superioridad en el desempeño del método de Rosenberger sobre los demás métodos.

Al presente trabajo le realizamos las siguientes recomendaciones con el objetivo de darle continuidad:

1. Analizar otros métodos no supervisados de la calidad de segmentación.
2. Realizar los mismos experimentos con muestras de mayor tamaño.

Bibliografía

- [BJE01] Bjerregaard, B., J. Jantzen, D. Landwehr, Foldberg AG. *Web-based pap-smear classification. A step towards auto-screening*. In Proc. 14 th International Congress of Cytology, Amsterdam. 2001.
- [COM02] Comaniciu, D., P. Meer. *Mean shift: A robust approach toward feature space analysis*. IEEE Transactions on pattern analysis and machine intelligence 2002. (456): 603-619.
- [DUD00] Duda, R. O., P. E. Hart, D. G. Stork, *Pattern Classification*, 2nd Edition, Wiley, 2000, ISBN 978-0-471-05669-0
- [FOR65] Forgy, E. W. *Cluster analysis of multivariate data: efficiency versus interpretability of classifications*. *Biometrics* 1965. **21**: 768–769.
- [BOR98] Borsotti, M., P. Campadelli, R. Schettini, *Quantitative evaluation of color image segmentation results*. Pattern Recognition Letters, 1998. **19**(8): p. 741-747.
- [CLO10] Cloppet, F., A. Boucher, *Segmentation of complex nucleus configurations in biological images*. Pattern Recognition Letters, 2010. **31**(8): p. 755-761.
- [COC54] Cochran, W.G., *Some methods for strengthening the common X2 tests*. *Biometrics*, 1954.
- [CHA06] Chabrier, S., B. Emile, C. Rosenberger., H. Laurent, *Unsupervised performance evaluation of image segmentation*. EURASIP Journal on Applied Signal Processing, 2006. **2006**: p. 217-217.

- [FRI10] Frintrop, S., A. Konigs, F. Hoeller, D. Schulz. *A component-based approach to visual person tracking from a mobile platform*. International Journal of Social Robotics, 2010. **2**(1): p. 53-62.
- [GEE09a] Geerinck, T., V. Enescu, I. Ravyse, H. Sahli. *Rule-based video interpretation framework: Application to automated surveillance*. In: Image and Graphics, 2009. ICIG'09. Fifth International Conference on. 2009: IEEE.
- [GEE09b] Geerinck, T., C. Mihai, M. Jabloun, I. Vanhamel, H. Sahli. *Multi-scale image analysis of satellite data using perceptual grouping*. In: IEEE International Geoscience and Remote Sensing Symposium. 2009.
- [HAR85] Haralick, R.M., L.G. Shapiro, *Image segmentation techniques*. Computer vision, graphics, and image processing, 1985. **29**(1): p. 100-132.
- [HOU12] Hou, X., J. Harel, C. Koch. *Image signature: Highlighting sparse salient regions*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2012. **34**(1): p. 194-201.
- [JAN05] Jantzen, J., J. Norup, G. Dounias, B. Bejerregaard. *Pap-smear benchmark data for pattern classification*. Nature inspired Smart Information Systems (NiSIS 2005), 2005: p. 1-9.
- [JEM06] Jemal, A., R. Sigel, E. Ward, T. Murray, J. Xu, C. Smigal, M. J. Thun. *Cancer statistics, 2006*. CA: a cancer journal for clinicians, 2006. **56**(2): p. 106-130.
- [JUN10] Jung, C., C. Kim, S. W. Chae, S. Oh. *Unsupervised segmentation of overlapped nuclei using Bayesian classification*. Biomedical Engineering, IEEE Transactions on, 2010. **57**(12): p. 2825-2832.
- [KAL10] Kale, A., S. Aksoy. *Segmentation of cervical cell images*. in *Pattern Recognition (ICPR)*, 2010 20th International Conference on. 2010: IEEE.
- [LEB07] Lebrun, G., C. Charrier, O. Lezoray, C. Meurie, H. Cardot. *A fast and efficient*

- segmentation scheme for cell microscopic image*. Cell Mol Biol (Noisy-le-grand), 2007. **53**(2): p. 51-61.
- [LEV85] Levine, M.D., A.M. Nazif, *Dynamic measurement of computer generated image segmentations*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1985(2): p. 155-164.
- [LIU04] Liu, H., X. Xie, X. Tang, Z.-W. Li, W.-Y. Ma. *Effective browsing of web image search results*. In: Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval. 2004: ACM.
- [LIU94] Liu, J., Y.-H. Yang, *Multiresolution color image segmentation*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1994. **16**(7): p. 689-700.
- [MAR03] Martin, L., M. Exbrayat, *Pap-smear classification*. 2003.
- [MAR09] Marcuzzo, M., P. Quelhas, A. Campilho. *Automated Arabidopsis plant root cell segmentation based on SVM classification and region merging*. Computers in biology and medicine, 2009. **39**(9): p. 785-793.
- [MEI98] Meisels, A., C. Morin, D. Giri, R. S. Hoda. *Cytopathology of the uterus*. International Journal of Gynecologic Pathology, 1998. **17**(3): p. 286.
- [MUN57] Munkres, J., *Algorithms for the assignment and transportation problems*. Journal of the Society for Industrial & Applied Mathematics, 1957. **5**(1): p. 32-38.
- [NOR05] Norup, J., *Classification of Pap-smear data by transduction neuro-fuzzy methods*. 2005, Technical University of Denmark, DTU, DK-2800 Kgs. Lyngby, Denmark.
- [PAL89] Pal, N.R., S.K. Pal, *Entropic thresholding*. Signal processing, 1989. **16**(2): p. 97-108.

- [PAL99] Palmer, S.E., *Vision science: Photons to phenomenology*. 1999: The MIT press.
- [PAP42] Papanicolaou, G.N., *A new procedure for staining vaginal smears*. Science, 1942. **95**(2469): p. 438-439.
- [PAR10] Parikh, N., L. Itti, J. Weiland, *Saliency-based image processing for retinal prostheses*. Journal of neural engineering, 2010. **7**(1): p. 016006.
- [PAY00] Payne, N., J. Chilcott, E. McGoogan, *Liquid-based cytology in cervical screening: a rapid and systematic review*. 2000.
- [PLI11] Plissiti, M.E., C. Nikou, A. Charchanti, *Combining shape, texture and intensity features for cell nuclei extraction in Pap smear images*. Pattern Recognition Letters, 2011. **32**(6): p. 838-853.
- [ROS99] Rosenberger, C., *Mise en oeuvre d'un système adaptatif de segmentation d'images*. 1999, Université Rennes 1.
- [SEZ04] Sezgin, M., *Survey over image thresholding techniques and quantitative performance evaluation*. Journal of Electronic imaging, 2004. **13**(1): p. 146-168.
- [VAN06] Vanhamel, I., *Vector valued nonlinear diffusion and its application to image segmentation*. 2006, PhD thesis, ETRO/IRIS, Vrije Universiteit Brussel, Brussels-Belgium.
- [VIN91] Vinet, L., *Segmentation et mise en correspondance de régions de paires d'images stéréoscopiques*. 1991.
- [ZHO09] Zhou, X., F. Li, J. Yan, S. T. Wong. *A novel cell segmentation method and cell phase identification using Markov model*. Information Technology in Biomedicine, IEEE Transactions on, 2009. **13**(2): p. 152-157.

Anexo 1. Pruebas estadísticas para las clases “célula anormal”-“célula normal”

A1.1 Prueba de Friedman: Normalidad = Anormal

Rangos^a

	Media de Rango
Bor	3.36
CH	2.40
Lev1	3.84
Lev2	3.18
Ros1	2.22

a. Normalidad = Anormal

Prueba de Friedman^a

N	40.000
Chi-Cuadrado	34.952
gl	4.000
Significación	.000
Asintótica	

a. Normalidad = Anormal

A1.2 Prueba de Friedman: Normalidad = Normal

Rangos^a

	Media de Rango
Bor	3.47
CH	2.55

Lev1	4.23
Lev2	3.05
Ros1	1.70

a. Normalidad = Normal

Prueba de Friedman^a

N	30.000
Chi-Cuadrado	52.857
gl	4.000
Significación Asintótica	.000

a. Normalidad = Normal

A1.3 Prueba de Rangos Signados de Wilcoxon: Normalidad = Anormal

Rangos^d

	N	Media de Rango	Suma de Rangos
Ros1 - CH Rangos Negativos	15 ^a	16.40	246.00
Rangos Positivos	13 ^b	12.31	160.00
Empates	12 ^c		
Total	40		

a. Ros1 < CH

b. Ros1 > CH

c. Ros1 = CH

d. Normalidad = Anormal

Prueba de Rangos Signados de Wilcoxon^b

	Ros1 - CH
Z	-1.001 ^a
Significación Asintótica (2-colas)	.317

a. Basado en los rangos positivos.

b. Normalidad = Anormal

A1.4 Prueba de Rangos Signados de Wilcoxon: Normalidad = Normal

Rangos ^d		N	Media de Rango	Suma de Rangos
Ros1 - CH	Rangos Negativos	18 ^a	13.17	237.00
	Rangos Positivos	5 ^b	7.80	39.00
	Empates	7 ^c		
	Total	30		

a. Ros1 < CH

b. Ros1 > CH

c. Ros1 = CH

d. Normalidad = Normal

Prueba de Rangos Signados de Wilcoxon^b

	Ros1 - CH
Z	-3.042 ^a
Significación Asintótica (2-colas)	.002

a. Basado en los rangos positivos.

b. Normalidad = Normal

Anexo 2. Pruebas estadísticas para las siete clases

A2.1. Prueba de Friedman: Clase = Carcinoma in Situ

Rangos^a

	Media de Rangos
Bor	3.55
CH	2.85
Lev1	3.75
Lev2	2.90
Ros1	1.95

a. Clase = Carcinoma in Situ

Prueba de Friedman^a

N	10.000
Chi-Cuadrado	11.034
gl	4.000
Significación	
Asintótica.	.026

a. Clase = Carcinoma in Situ

A2.2. Prueba de Friedman: Clase = Displasia Severa

Rangos^a

	Media de Rango
Bor	3.60
CH	1.85
Lev1	4.10
Lev2	3.20

Ros1	2.25
------	------

a. Clase = Displasia Severa

Prueba de Friedman^a

N	10.000
Chi-Cuadrado	16.945
gl	4.000
Significación	
Asintótica.	.002

a. Clase = Displasia Severa

A2.3. Prueba de Friedman: Clase = Displasia Moderada

Rangos^a

	Media de Rango
Bor	3.05
CH	2.30
Lev1	4.10
Lev2	3.00
Ros1	2.55

a. Clase = Displasia

Moderada

Prueba de Friedman^a

N	10.000
Chi-Cuadrado	8.759
gl	4.000
Significación	
Asintótica.	.067

a. Clase = Displasia Moderada

A2.4. Prueba de Friedman: Clase = Displasia Leve

Rangos^a

	Media de Rango
Bor	3.25
CH	2.60
Lev1	3.40
Lev2	3.60
Ros1	2.15

a. Clase = Displasia Leve

Prueba de Friedman^a

N	10.000
Chi-Cuadrado	6.370
gl	4.000
Significación	
Asintótica.	.173

a. Clase = Displasia Leve

A2.5. Prueba de Friedman: Clase = Columnar Normal

Rangos^a

	Media de Rango
Bor	2.95
CH	2.35
Lev1	4.05
Lev2	3.45
Ros1	2.20

a. Clase = Columnar
Normal

Prueba de Friedman^a

N	10.000
Chi-Cuadrado	10.773
gl	4.000

Significación Asintótica.	.029
------------------------------	------

a. Clase = Columnar Normal

A2.6. Prueba de Friedman: Clase = Intermedia Normal

Rangos ^a	
	Media de Rango
Bor	3.80
CH	2.65
Lev1	4.50
Lev2	2.90
Ros1	1.15

a. Clase = Intermedia

Normal

Prueba de Friedman ^a	
N	10.000
Chi-Cuadrado	29.977
gl	4.000
Significación Asintótica.	.000

a. Clase = Intermedia Normal

A2.7. Prueba de Friedman: Clase = Superficial Normal

Rangos ^a	
	Media de Rango
Bor	3.65
CH	2.65
Lev1	4.15
Lev2	2.80

Ros1	1.75
------	------

a. Clase = Superficial Normal

Prueba de Friedman^a

N	10.000
Chi-Cuadrado	18.884
gl	4.000
Significación Asintótica.	.001

a. Clase = Superficial Normal

A2.8. Prueba de Rangos Signados de Wilcoxon: Clase = Carcinoma in Situ

Rangos^d

	N	Media de Rango	Suma de Rangos
Ros1 - CH Rangos Negativos	6 ^a	4.92	29.50
Rangos Positivos	2 ^b	3.25	6.50
Empates	2 ^c		
Total	10		

a. Ros1 < CH

b. Ros1 > CH

c. Ros1 = CH

d. Clase = Carcinoma in Situ

Prueba de Rangos Signados de Wilcoxon^b

	Ros1 - CH
Z	-1.653 ^a
Significación Asintótica (2-colas)	.098

a. Basado en los rangos positivos.

b. Clase = Carcinoma in Situ

A2.9. Prueba de Rangos Signados de Wilcoxon: Clase = Displasia Severa

Rangos ^d		N	Media de Rango	Suma de Rangos
Ros1 - CH	Rangos Negativos	2 ^a	3.50	7.00
	Rangos Positivos	4 ^b	3.50	14.00
	Empates	4 ^c		
	Total	10		

a. Ros1 < CH

b. Ros1 > CH

c. Ros1 = CH

d. Clase = Displasia Severa

Prueba de Rangos Signados de Wilcoxon^b

	Ros1 - CH
Z	-.750 ^a
Significación Asintótica (2-colas)	.453

a. Basado en los rangos positivos.

b. Clase = Displasia Severa

A2.10. Prueba de Rangos Signados de Wilcoxon: Clase = Displasia Moderada

Rangos ^d		N	Media de Rango	Suma de Rangos
Ros1 - CH	Rangos Negativos	2 ^a	4.00	8.00
	Rangos Positivos	4 ^b	3.25	13.00
	Empates	4 ^c		
	Total	10		

a. Ros1 < CH

b. Ros1 > CH

c. Ros1 = CH

d. Clase = Displasia Moderada

Prueba de Rangos Signados de Wilcoxon^b

	Ros1 - CH
Z	-.531 ^a
Significación Asintótica (2-colas)	.595

a. Basado en los rangos positivos.

b. Clase = Displasia Moderada

A2.11. Prueba de Rangos Signados de Wilcoxon: Clase = Displasia Leve

Rangos^d

	N	Media de Rango	Suma de Rangos
Ros1 - CH Rangos Negativos	5 ^a	5.10	25.50
Rangos Positivos	3 ^b	3.50	10.50
Empates	2 ^c		
Total	10		

a. Ros1 < CH

b. Ros1 > CH

c. Ros1 = CH

d. Clase = Displasia Leve

Prueba de Rangos Signados de Wilcoxon^b

	Ros1 - CH
Z	-1.098 ^a
Significación Asintótica (2-colas)	.272

a. Basado en los rangos positivos.

b. Clase = Displasia Leve

A2.12. Prueba de Rangos Signados de Wilcoxon: Clase = Columnar Normal

Rangos^d

	N	Media de Rango	Suma de Rangos
--	---	----------------	----------------

Ros1 - CH	Rangos Negativos	4 ^a	5.50	22.00
	Rangos Positivos	4 ^b	3.50	14.00
	Empates	2 ^c		
	Total	10		

a. Ros1 < CH

b. Ros1 > CH

c. Ros1 = CH

d. Clase = Columnar Normal

Prueba de Rangos Signados de Wilcoxon^b

	Ros1 - CH
Z	-.566 ^a
Significación Asintótica (2-colas)	.572

a. Basado en los rangos positivos.

b. Clase = Columnar Normal

A2.13. Prueba de Rangos Signados de Wilcoxon: Clase = Intermedia Normal

Rangos^d

	N	Media de Rango	Suma de Rangos
Ros1 - CH			
Rangos Negativos	9 ^a	5.00	45.00
Rangos Positivos	0 ^b	.00	.00
Empates	1 ^c		
Total	10		

a. Ros1 < CH

b. Ros1 > CH

c. Ros1 = CH

d. Clase = Intermedia Normal

Prueba de Rangos Signados de Wilcoxon^b

	Ros1 - CH
Z	-2.719 ^a
Significación Asintótica (2-colas)	.007

a. Basado en los rangos positivos.

b. Clase = Intermedia Normal

A2.14. Prueba de Rangos Signados de Wilcoxon: Clase = Superficial Normal

		Rangos ^d		
		N	Media de Rango	Suma de Rangos
Ros1 - CH	Rangos Negativos	5 ^a	3.70	18.50
	Rangos Positivos	1 ^b	2.50	2.50
	Empates	4 ^c		
	Total	10		

a. Ros1 < CH

b. Ros1 > CH

c. Ros1 = CH

d. Clase = Superficial Normal

Prueba de Rangos Signados de Wilcoxon^b

	Ros1 - CH
Z	-1.725 ^a
Significación Asintótica (2-colas)	.084

a. Basado en los rangos positivos.

b. Clase = Superficial Normal