

Segmentación de imágenes de ensayo cometa con algoritmos no supervisados

R. Pérez-Cervantes¹, R. Legarda-Sáenz^{1,*}, C. Brito-Loeza¹

¹CLIR en Facultad de Matemáticas, Universidad Autónoma de Yucatán, México

Fecha de recepción: 10 de junio de 2025 — Fecha de aceptación: 10 de marzo de 2026

Resumen

La electroforesis en gel unicelular, también conocida como ensayo cometa, es una técnica que permite cuantificar sobre imágenes digitales el daño del ADN a nivel celular. La segmentación del cuerpo del cometa y la cabeza de este posibilitan extraer características distintivas del cometa, como son el área, la longitud, el nivel de intensidad y la esfericidad, para un posterior análisis y estimación del daño celular. En este artículo se evalúan cuantitativamente tres métodos de segmentación no supervisada para la extracción del cuerpo y la cabeza del cometa, que son fáciles de implementar y no utilizan un gran tiempo computacional en su ejecución.

Palabras Clave: Ensayo cometa, Modelos de Gauss-Markov, K-medias, Mezcla de Gaussinas

Segmentation of comet assay images using unsupervised algorithms

Abstract

Single-cell gel electrophoresis, also known as the comet assay, is a technique that allows the quantification of DNA damage at the cellular level on digital images. Segmentation of the comet body and head makes it possible to extract distinctive comet characteristics, such as area, length, intensity level, and sphericity, for subsequent analysis and estimation of cellular damage. This article quantitatively evaluates three unsupervised segmentation methods for extracting the comet body and head. These methods are easy to implement and do not require extensive computational time.

Keywords: Comet test, Gauss-Markov models, K-medias, Mixture of Gaussians

1. Introducción

El ensayo de electroforesis de gel de célula única (SC-GE, por sus siglas en inglés), también conocido como *ensayo cometa*, es una técnica sencilla utilizada para detectar daños y capacidad regenerativa en el ácido desoxirribonucleico (ADN) en una sola célula eucariótica. El patrón de migración del ADN a través del gel de electroforesis, que con frecuencia se asemeja a un cometa, se denomina *cometa* [1].

La técnica del ensayo cometa tiene tres etapas: encapsulación, lisis y electroforesis. En la primera etapa, encapsulación, una muestra celular se dispersa en célu-

las individuales y se suspende en agarosa; la suspensión es puesta en una lámina de microscopio y se aplica un agente de prueba como luz ultravioleta, radiación ionizante, o una sustancia química genotóxica para inducir el daño del ADN en la célula encapsulada. En la segunda etapa, lisis, las láminas son luego sumergidas en una solución de sal acuosa altamente concentrada y detergente, que causa que las células se muestren. La célula es destruida, quedando solo su ADN formando una estructura llamada nucleoide. En la tercera etapa, las láminas son lavadas en agua destilada y sumergidas en una solución de electroforesis durante aproximadamente 20 minutos y entonces se aplica un campo eléctrico durante

*rlegarda@correo.uady.mx

otros 20 minutos. Finalmente, las láminas son teñidas con colorante fluorescente específico del ADN y analizadas utilizando un microscopio digital [2]. Un ejemplo de las imágenes resultantes del ensayo cometa se muestra en la Figura 1.

Las imágenes del ensayo cometa muestran dos estructuras principales: un núcleo, conocido como la *cabeza del cometa*, y un rastro, conocido como la *cola del cometa*. Para evaluar el daño del ADN, un experto observa la imagen digital y determina con alguna herramienta de software qué porción del cometa corresponde a cada estructura. Esta evaluación, aun cuando se realiza en una computadora, es subjetiva porque es realizada de acuerdo al criterio del experto. Con el fin de disminuir esta subjetividad, en años recientes han surgido diferentes propuestas para automatizar este tipo de procesos, desde la segmentación de las imágenes, cuyo objetivo es identificar qué porciones del cometa corresponden a la cabeza y cuáles a la cola, hasta la medición de parámetros y medidas de las estructuras de los cometas que permiten analizar estadísticamente su posible relación con enfermedades o agentes patógenos [3]. En la Figura 1 se muestra un ejemplo de una imagen del ensayo cometa y su segmentación manual realizada por un experto. En años recientes ha surgido software abierto y comercial para llevar a cabo el proceso de segmentación y mediciones en las imágenes de ensayo cometa. Por ejemplo, en la referencia [4] proponen *HiComet*, un software computacional para segmentar y caracterizar los cometas extrayendo sus perfiles de intensidad y momentos de cola; por su parte, en la referencia [5] los autores proponen *DeepComet*, un método basado en algoritmos de aprendizaje profundo con redes neuronales de máscara R-CNN. Por otro lado, en la referencia [6] se propone la herramienta *CometQ* para la segmentación y cuantificación del daño del ADN por el análisis de imágenes del ensayo cometa. En la referencia [7], el autor y colaboradores proponen un algoritmo de libre acceso de aprendizaje profundo llamado *CometAnalyser*, diseñado para una fácil segmentación y clasificación de cometas en imágenes con teñido fluorescente y plateado. Un último ejemplo puede ser el trabajo publicado en la referencia [8] donde presentan el *OpenComet*, que es una herramienta para la segmentación de imágenes del ensayo cometa mediante la separación de las cabezas de los cometas y la medición de los parámetros de los mismos.

De las técnicas mencionadas previamente, algunas de ellas se basan en algoritmos de aprendizaje supervisado con redes neuronales profundas, que han mostrado una gran precisión en la segmentación. Sin embargo, estas técnicas requieren de grandes bases de datos para el entrenamiento, así como un poder de cómputo considerable. Esto limita su uso extendido en equipos de cómputo estándares. Por otro lado, otras técnicas basan su proceso de segmentación en la umbralización de la intensidad

de los píxeles y la aplicación de operadores morfológicos. Aunque son fáciles de implementar desde el punto de vista computacional, pueden introducir sesgos en los resultados. Al usar estos operadores para separar el ruido del objeto que nos interesa, estos métodos a menudo alteran la forma de los cometas [9], lo que puede llevar a errores en las mediciones de la información.

La contribución de este trabajo radica en la exploración de algoritmos no supervisados para la segmentación automática de imágenes de ensayo cometa, utilizando solo la información de la intensidad de la imagen. A diferencia de los métodos basados en redes neuronales, los métodos que analizamos, K-medias [10], mezcla de Gaussianas [11, 12] y el que, por basarse en dicha teoría, llamamos de campos de medida de Gauss-Markov [13], no requieren de grandes bases de entrenamiento ni poder de cómputo, lo que los hace fáciles de implementar. El manuscrito está organizado de la siguiente forma: la Sección 2 aborda el fundamento teórico de los algoritmos seleccionados. En la Sección 3 se muestran algunos experimentos realizados en la segmentación de imágenes con múltiples cometas, y en la Sección 4 se analizan los resultados obtenidos para cada método a través del estadístico del coeficiente de Sorensen-Dice [14, 15]. Finalmente, en la Sección 5 se presentan las conclusiones generales del trabajo realizado.

2. Metodología

El tipo de segmentación que se utiliza en las imágenes de ensayo cometa es semántica [3]; esta consiste en clasificar cada píxel de la imagen en alguna de tres clases posibles: cabeza, cola y fondo. Para esto, considere una imagen I , la cual es una matriz en $A = X^{n \times m}$, donde X es un conjunto ordenado. Sea $f : \{1, \dots, n\} \times \{1, \dots, m\} \rightarrow A$ una función definida como $f(i, j) = I_{i,j}$, donde $I_{i,j}$ representa la entrada (i, j) de la matriz I . La segmentación semántica binaria de una imagen, basada en la intensidad o valor de cada píxel, consiste en encontrar un umbral T y definir la función $S_T : \{1, \dots, n\} \times \{1, \dots, m\} \rightarrow \{a, b\}$

$$S_T(i, j) = \begin{cases} a & \text{si } f(i, j) \geq T \\ b & \text{si } f(i, j) < T \end{cases} \quad (1)$$

donde $T \in X$ es el umbral de segmentación y $a, b \in X$ son las dos clases posibles. El mismo método aplicado a una imagen a color en el espacio RGB, esto es, de tres canales, consiste en definir una función similar a la anterior, considerando $A = (X^3)^{n \times m}$, donde el orden en X^3 está definido para $(a_1, a_2, a_3), (b_1, b_2, b_3) \in X^3$ como $(a_1, a_2, a_3) > (b_1, b_2, b_3)$ si y solo si $a_i > b_i$ para $i = 1, 2, 3$. De esta forma, mientras que en una imagen en escala de grises, $X = \{0, 1, \dots, 255\}$, en una imagen a color RGB esta X corresponde a cada canal de color.

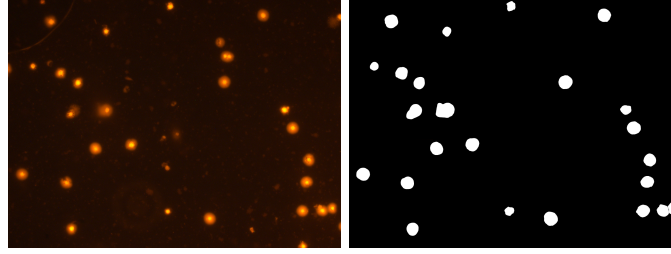


Figura 1: A la izquierda, una imagen a color del ensayo cometa; a la derecha, la segmentación manual correspondiente.

A continuación, se presentan los fundamentos teóricos de los métodos de K-medias, mezcla de Gaussianas y campos de medida de Gauss-Markov. Para ello, considérese un conjunto de muestras $D = \{x_1, \dots, x_N\}$ en un espacio X d -dimensional.

2.1. K-medias

K-medias es un método de agrupamiento que produce una partición de un conjunto en grupos disjuntos con sus centroides contenidos. Para inicializar dicho algoritmo, primero se especifica un conjunto de K tentativos centroides [16] que define el número de grupos que se formarán; existen variaciones de K-medias, que permiten estimar el número de grupos [17]. Usualmente, los centroides $z_1, \dots, z_K \in X$, son seleccionados como las primeras K muestras del conjunto de muestras [10], es decir, del conjunto D .

Luego, se calcula la distancia de los elementos en el conjunto de muestras a los centroides. La mínima distancia entre el elemento y el centroide determinará la pertenencia al grupo, es decir, x en el conjunto de muestras, pertenece al i -ésimo grupo C_i , si

$$\|x - z_i\| < \|x - z_j\| \quad (2)$$

para todo $j \neq i$, $j \in \{1, 2, \dots, K\}$ [10].

Con los resultados del paso anterior, se determinan nuevos centroides para cada uno de los grupos. Para ello, se minimiza, con respecto del nuevo centroide z'_i , la suma de las distancias entre los puntos muestra y el nuevo centroide, esto es,

$$\min_{z'_i} \sum_{x \in C_i} \|x - z'_i\|^2 \quad (3)$$

para $i = 1, 2, \dots, K$. El algoritmo termina si $z'_i = z_i$, para cada i , esto es, si los centroides no se modifican; de otro modo, el ciclo continúa, es decir, se vuelve a calcular las distancias de los elementos a cada centroide y se recalculan nuevamente los centroides [10].

2.2. Mezcla de Gaussianas

Los modelos de mezcla de Gaussianas proveen una solución poderosa para el problema de agrupamiento,

y son muy útiles para sugerir criterios de agrupamiento [11]. En dichos modelos, se asume que las muestras forman un vector aleatorio x con densidad

$$f(x) = \sum_{j=1}^K \pi_j \Phi(x | \theta_j), \quad (4)$$

donde π_j son las proporciones de la mezcla, cumpliendo que $0 < \pi_j < 1$ y $\sum_{j=1}^K \pi_j = 1$; $\Phi(x | \theta)$, con $\theta = (\mu, \Sigma)$, denota la densidad de la distribución Gaussiana con vector de medias μ y matriz de covarianza Σ [11].

Para la estimación de los parámetros óptimos de cada Gaussiana, se utiliza el principio de *Máxima Verosimilitud*; esto es, se intenta encontrar los parámetros para los que la función:

$$P(x_1, \dots, x_N | \pi_1, \dots, \pi_K, \theta_1, \dots, \theta_K) = \prod_{i=1}^N f(x_i) \quad (5)$$

es máxima [12]. En la práctica, la maximización se realiza sobre la función

$$\mathcal{L}(\pi_1, \dots, \pi_K, \theta_1, \dots, \theta_K) = \sum_{i=1}^N \log \left(\sum_{j=1}^K \pi_j \Phi_j(x_i | \theta_j) \right) \quad (6)$$

que provee una forma eficiente de llevar a cabo el proceso de optimización dadas las propiedades de la función logaritmo.

Finalmente, mediante la introducción de la variable latente $P(c | x)$ (que es la probabilidad de que la variable x pertenezca al componente c), se obtiene el logaritmo de verosimilitud completo. Ahora, bajo el supuesto de que $P(c | x)$ es fijo, se efectúa el procedimiento de *Estimación-Maximización*. Este consiste en un proceso iterativo, donde se alterna la estimación de la variable latente, y el cálculo de los parámetros que maximizan el logaritmo de verosimilitud completo [12].

2.3. Campos de medida de Gauss-Markov

En este método, se visualiza un conjunto D de muestras como una imagen I , tal que $I(r) \in D$ para la posición del pixel $r = (i, j)$. Así, en el método de campos de

medida de Gauss-Markov [13], primero se considera una imagen I y una región L en la imagen. Luego, se toma un conjunto finito de puntos $\{r \in L\}$, los cuales toman valores en un conjunto finito $Q = \{q_1, \dots, q_M\}$, donde q_i representa el valor del pixel en la imagen del i -ésimo representante o centroide, y un campo de distribución de probabilidades p sobre Q .

Sea $p(r, k)$ la probabilidad de que la imagen I , en el pixel $r = (x, y)$, tome el valor del pixel representativo de la clase k ; en otras palabras, se puede expresar como:

$$p(r, k) = Pr[I(r) = q_k], \quad (7)$$

donde $k \in \{1, 2, \dots, M\}$.

El estimador óptimo de p se obtiene al minimizar la siguiente función:

$$U(p) = \sum_{k=1}^M \left\{ \sum_{r \in L} [p(r, k) - \hat{p}(r, k)]^2 \alpha + \lambda \sum_{\langle r, s \rangle \in L} \sum_{k'=1}^M [p(r, k) - p(s, k')]^2 \beta \right\}, \quad (8)$$

donde $\alpha = a(r)$ y $\beta = b(r, s, |k - k'|)$ denotan pesos no negativos de variación de espacio.

El término $\hat{p}(r, k)$ denota el campo de distribuciones de probabilidad condicional sobre Q . Un caso particular se obtiene al asumir que la imagen tiene ruido Gaussiano con media 0 y varianza σ^2 . Así, en este caso, $\hat{p}(r, k)$ se estima por

$$\hat{p}(r, k) = \frac{1}{z(r)} e^{-\frac{[I(r) - q_k]^2}{2\sigma^2}} \quad (9)$$

donde $z(r)$ es una constante de normalización tal que $\sum_{k=1}^M \hat{p}(r, k) = 1$ para cada $r \in L$.

Derivando, con respecto a p , la ecuación (8) e igualando a cero, obtenemos la siguiente ecuación:

$$0 = \alpha [p(r, k) - \hat{p}(r, k)] + \lambda \sum_{s \in N_r \cap L} \sum_{k'=1}^M [p(r, k) - p(s, k')] \beta, \quad (10)$$

de donde despejamos $p(r, k)$ para obtener

$$p(r, k) = \frac{\alpha \hat{p}(r, k) + \lambda \sum_{s \in N_r \cap L} \sum_{k'=1}^M p(s, k') \beta}{\alpha + \lambda \sum_{s \in N_r \cap L} \sum_{k'=1}^M \beta} \quad (11)$$

para cada k .

Un caso particular es cuando las restricciones de variación de espacio indican que la solución debe ser suave; así, solo se requiere que la distribución p sea similar a sus vecindades. En este caso, se define $b(r, s, |k - k'|) = 1$ si $k = k'$ y $b(r, s, |k - k'|) = 0$ de otro modo [13]. Así, la ecuación (11) se puede reescribir como

$$p(r, k) = \frac{\alpha \hat{p}(r, k) + \lambda \sum_{s \in N_r \cap L} p(s, k)}{\alpha + \lambda \sum_{s \in N_r \cap L} 1}. \quad (12)$$

¹<https://scikit-learn.org/stable/>

3. Experimentos

Para validar los tres modelos de segmentación descritos con anterioridad, se utilizó un conjunto de 50 imágenes de ensayo cometa, las cuales fueron segmentadas manualmente por dos expertos. Ejemplos de la segmentación manual se muestran en la Figura 2 donde en color blanco se muestran las regiones de la cabeza y de todo el cuerpo del cometa y en negro el fondo de la imagen.

En la implementación computacional de los tres modelos se utilizó el lenguaje de programación *Python* y las bibliotecas de *Scikit-learn*¹ cuando fue posible. En todos los casos se hizo una búsqueda de los mejores parámetros de cada modelo. Las 50 imágenes disponibles se dividieron en dos conjuntos: 35 para el conjunto de entrenamiento y 15 para el conjunto de prueba. A continuación describimos los parámetros utilizados para las técnicas de segmentación utilizadas en los experimentos:

K-medias Este método requiere como parámetro de entrada el número de grupos, el cual se fijó en 3.

Mezcla de Gaussianas Este modelo requiere como parámetro de entrada el número de componentes, el cual se fijó en tres.

Campos de medida de Gauss-Markov Para el método de campos de medida de Gauss-Markov, los parámetros usados son $k = 3$ en la aproximación (9) y los valores de q_1, q_2, q_3 fueron seleccionados para cada uno de los cometas del conjunto de entrenamiento de modo tal que sean representativos para cada clase, a saber, el fondo de la imagen, la cola del cometa y la cabeza del cometa. Los valores de q_1, q_2, q_3 utilizados en los cometas del conjunto de prueba son el promedio de los del conjunto de entrenamiento. El método de Gauss-Seidel se utilizó para encontrar la solución del sistema lineal (12), puesto que, por su definición, utiliza los nuevos valores obtenidos en las estimaciones siguientes. El criterio de parada del algoritmo de Gauss-Seidel utilizado fue detener el algoritmo cuando la diferencia de la norma euclidiana de la solución entre dos iteraciones consecutivas sea menor que 0.1. Finalmente, se usaron $a = 1$ y $\lambda = 1$ ya que dieron los mejores resultados.

Ejemplos de los resultados de segmentación de las imágenes de cometa se muestran en las Figuras 3 y 4. En la Figura 3 se muestran los resultados de la segmentación del cuerpo del cometa, y la Figura 4 presenta la segmentación de la cabeza. En ambas imágenes agregamos las segmentaciones realizadas por los expertos con el fin de comparación. Adicionalmente, en la Tabla 1 se presentan los tiempos promedio de ejecución de cada método, utilizando las imágenes de prueba.

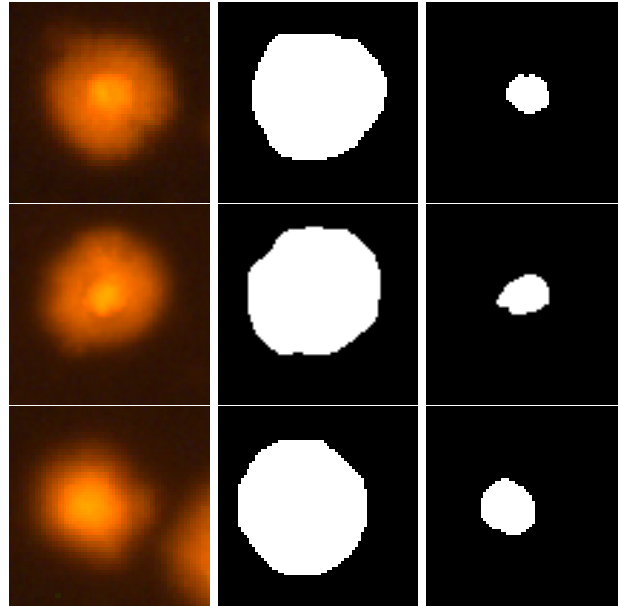


Figura 2: Tres ejemplos de segmentación manual de imágenes de ensayo cometa realizadas por expertos. De izquierda a derecha, en la primera columna se muestra la imagen original, en la segunda columna la segmentación manual del cuerpo del cometa y en la tercera columna la segmentación manual de la cabeza.

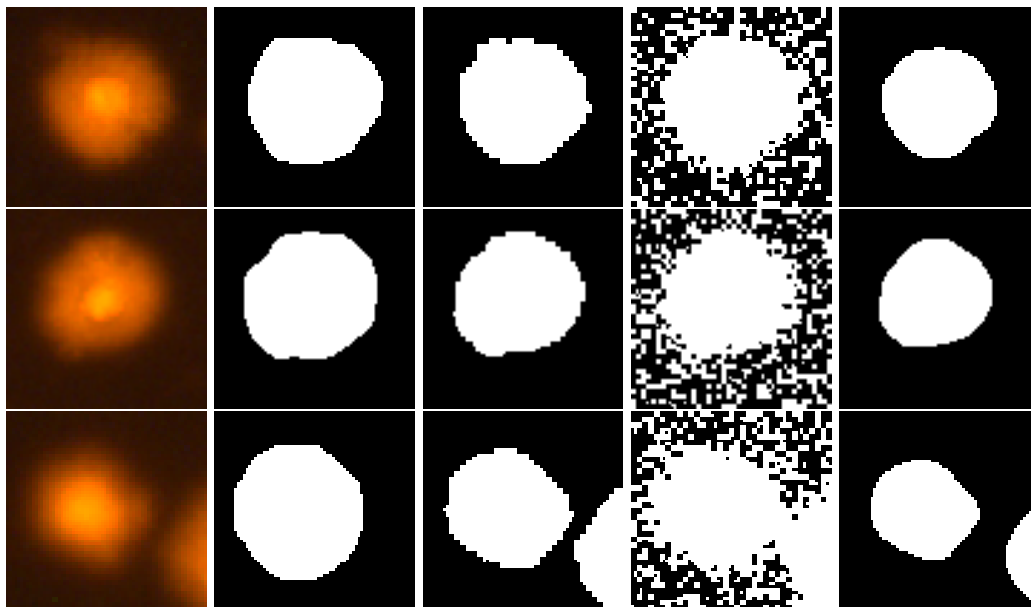


Figura 3: Tres ejemplos de segmentación de imágenes de ensayo cometa: De izquierda a derecha, en la primera columna se muestra la imagen original; en la segunda, la segmentación manual del cuerpo del cometa hecha por los expertos; la tercera es el resultado obtenido por la técnica de K-medias; la cuarta, por la de Mezcla de Gaussianas; y la quinta, por la de Campos de medida de Gauss-Markov.

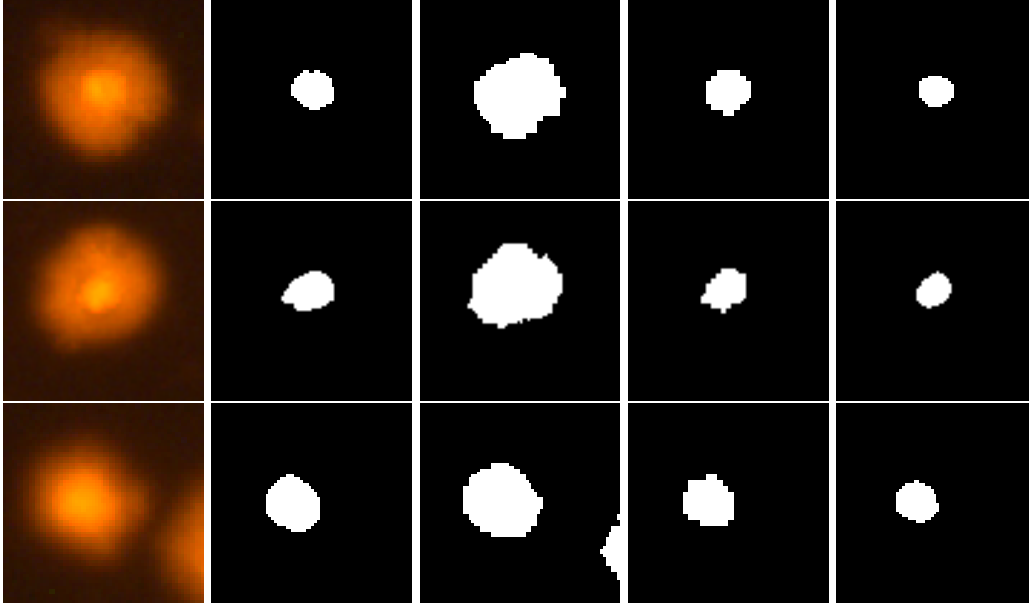


Figura 4: Tres ejemplos de segmentación de imágenes de ensayo cometa: De izquierda a derecha, en la primera columna se muestra la imagen original; en la segunda, la segmentación manual de la cabeza del cometa hecha por los expertos; la tercera es el resultado obtenido por la técnica de K-medias; la cuarta, por la de Mezcla de Gaussianas; y la quinta, por la de Campos de medida de Gauss-Markov.

Tabla 1: Tiempo de ejecución promedio de cada algoritmo, calculado para los 15 cometas del conjunto de prueba.

Métodos	Tiempo (segundos)
K-medias	1.49
Mezcla de Gaussianas	3.26
Campos de medida de Gauss-Markov	1.52

4. Análisis cuantitativo

Para comparar los resultados obtenidos en cada uno de los métodos, se elige el coeficiente de Sorensen-Dice [14, 15]. Considerando dos muestras A y B , el coeficiente de Sorensen-Dice se define como

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|}, \quad (13)$$

donde $|A \cap B|$ representa el número de elementos que pertenecen a ambas muestras y, $|A|$ y $|B|$ son el número de elementos de la muestra A y la muestra B , respectivamente. Así, considerando a A como la segmentación automática de las imágenes por alguno de los tres métodos y B la segmentación manual de las imágenes, entonces $|\cdot|$ se entiende por el número de píxeles activos en la imagen, es decir, aquellos cuyo valor es 1.

En la Figura 5 se presentan los valores obtenidos de los coeficientes de Sorensen-Dice para cada uno de los cometas del conjunto de prueba, tanto en la segmentación de todo el cometa como en la segmentación de únicamente la cabeza.

Para la segmentación del cometa completo, el método de K-medias es consistentemente el mejor, obtenien-

do un coeficiente de 0.8693 ± 0.0672 , mientras que los otros dos obtuvieron 0.7121 ± 0.0614 para el método de mezcla de Gaussianas y 0.7389 ± 0.0934 el método de Gauss-Markov. Es de hacer notar que el método de mezcla de Gaussianas entrega resultados muy malos para la segmentación del cuerpo del cometa, como puede apreciarse visiblemente en la Figura 3. Por otro lado, en la segmentación de la cabeza, el mejor método es mezcla de Gaussianas con un coeficiente Dice de 0.8201 ± 0.1195 , siguiéndole los métodos de campos de medida de Gauss-Markov con 0.7566 ± 0.1024 y K-medias con 0.5475 ± 0.1694 . En la Tabla 2 se muestran los resultados obtenidos para ambos tipos de segmentación.

5. Conclusión

De los métodos examinados, se intuye que el método de K-medias otorga una buena aproximación de los cometas completos. En contrapartida, las segmentaciones de solo las cabezas de los cometas, mediante el método de K-medias, no fueron muy acertadas con respecto a la segmentación manual. A su favor se encuentra que el

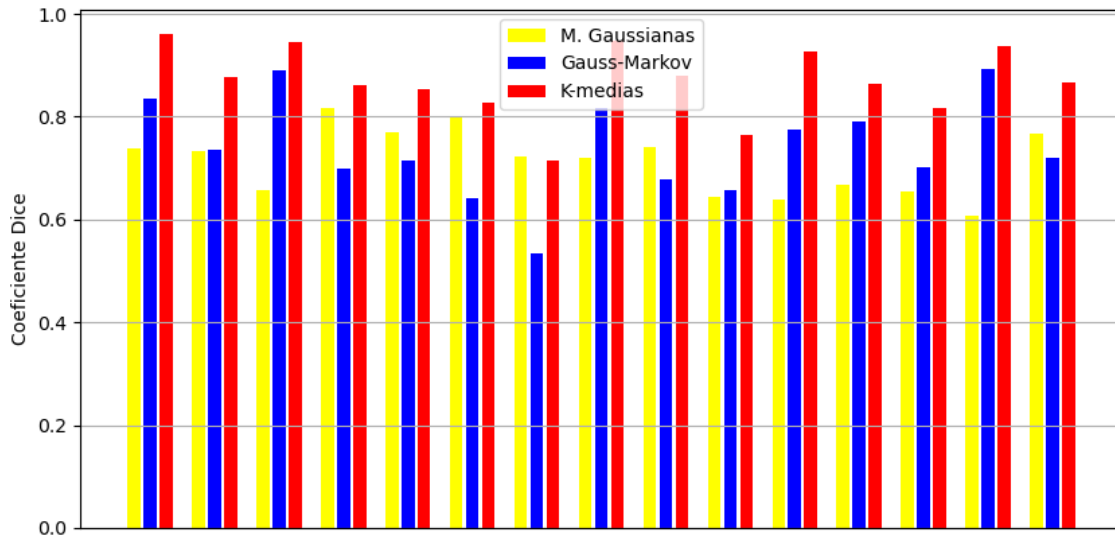


Figura 5: Valores del coeficiente de Sorensen-Dice obtenidos por cada método en la segmentación del cuerpo del cometa en las imágenes del conjunto de prueba.

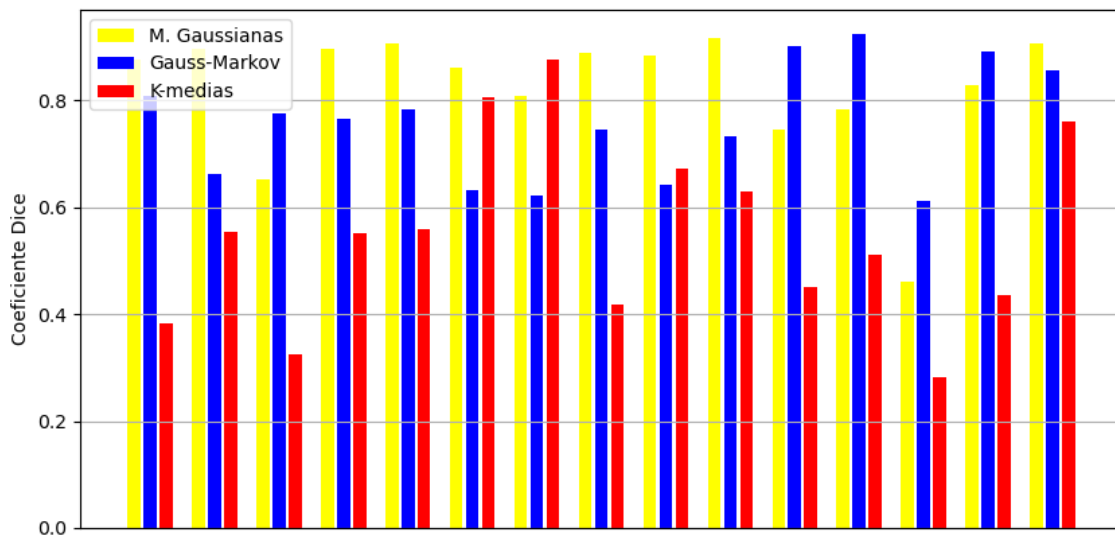


Figura 6: Valores del coeficiente de Sorensen-Dice obtenidos por cada método en la segmentación de la cabeza del cometa en las imágenes del conjunto de prueba.

Tabla 2: Coeficientes de Sorensen-Dice obtenidos por cada método en la segmentación, calculados para los 15 cometas del conjunto de prueba.

Métodos	Cometa completo	Cabeza del cometa
K-medias	0.8693 \pm 0.0672	0.5475 \pm 0.1694
Mezcla de Gaussianas	0.7121 \pm 0.0614	0.8201 \pm 0.1195
Campos de medida de Gauss-Markov	0.7389 \pm 0.0934	0.7566 \pm 0.1024

tiempo de ejecución fue el menor de los tres.

Por otro lado, para el caso de las segmentaciones de las cabezas de los cometas, se obtuvieron coeficientes de Sorensen-Dice muy superiores en el método de mezcla de Gaussianas, esto es, las segmentaciones de dicho método fueron muy cercanas a las segmentaciones manuales realizadas. En el caso de las segmentaciones del cometa completo, el método de mezcla de Gaussianas genera un coeficiente de Sorensen-Dice grande, pero visualmente, sus resultados no son favorables. Además, dicho método obtuvo el mayor tiempo de ejecución, doblando los tiempos de ejecución de K-medias y campos de medida

de Gauss-Markov.

Finalmente, en el método de campos de medida de Gauss-Markov, la segmentación, tanto en el caso de los cometas completos como en el de solo las cabezas, se comportó de forma adecuada con valores cercanos a los mejores en cada caso. De igual forma, el tiempo de ejecución promedio se obtuvo entre los tiempos de ejecución de los otros dos métodos.

Los resultados obtenidos indican que es posible utilizar dos métodos diferentes si se desea segmentar la cabeza y el cuerpo del cometa. Esto es viable dado que los tiempos de ejecución son relativamente pequeños.

Referencias

- [1] A. Rodríguez Rey, E. Noris García, M. T. Fundadora Torres, Principios y relevancia del ensayo cometa, Revista Cubana de Investigaciones Biomédicas. 35 (2016) 184–194.
- [2] F. F. Fierro, Electroforesis de ADN, in: A. Cornejo Romero, A. Serrato Díaz, B. Rendón Aguilar, M. G. Rocha Munive (Eds.), Herramientas moleculares aplicadas en ecología: aspectos teóricos y prácticos, Secretaría de Medio Ambiente y Recursos Naturales, 2014, pp. 27–51.
- [3] J. Luna-González, A. Martin-Gonzalez, C. Brito-Loeza, A. Espinosa-Romero, E. Pacheco-Pantoja, Introducción a técnicas de segmentación de células del ensayo cometa, Ingeniería Revista Académica de la Facultad de Ingeniería Universidad Autónoma de Yucatán 22 (2018) 26–37.
- [4] T. Lee, S. Lee, W. Sim, et al., Hicomet: a high-throughput comet analysis tool for large-scale dna damage assessment, BMC Bioinformatics 19 (2018) 44.
- [5] Y. Hong, H. Han, H. Lee, J. Ko, et al., Deep learning method for comet segmentation and comet assay image analysis, Scientific reports 10 (2020) 18915.
- [6] S. Ganapathy, A. Muraleedharan, P. S. Sathidevi, P. Chand, R. P. Rajkumar, Cometq: An automated tool for the detection and quantification of dna damage using comet assay image analysis, Computer Methods and Programs in Biomedicine 133 (2016) 143–154.
- [7] A. Beleon, S. Pignatta, C. Arienti, A. Carbonaro, P. Horvath, G. Martinelli, G. Castellani, A. Tessei, F. Piccinini, Cometanalyser: A user-friendly, open-source deep-learning microscopy tool for quantitative comet assay analysis, Computational and Structural Biotechnology Journal 20 (2022) 4122–4130.
- [8] B. M. Gyori, G. Venkatachalam, P. Thiagarajan, D. Hsu, M.-V. Clement, Opencomet: An automated tool for comet assay image analysis, Redox Biology 2 (2014) 457–465.
- [9] D. Rakov, Advantages and disadvantages of morphological methods in engineering, MATEC Web of Conferences 329 (2020) 03028.
- [10] J. Tou, R. Gonzalez, Pattern Recognition Principles, Addison-Wesley Publishing Company, 1974.
- [11] G. Gan, C. Ma, J. Wu, Data Clustering: Theory, Algorithms and Applications, SIAM, 2007.

- [12] J. Liu, D. Cai, X. He, Gaussian mixture model with local consistency, Proceedings of the AAAI Conference on Artificial Intelligence 24 (2010) 512–517.
- [13] J. Marroquin, R. Rodriguez-Vera, M. Servin, Local phase from local orientation by solution of a sequence of linear systems, Optical Society of America A 15 (1998) 1536–1544.
- [14] L. R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (1945) 297–302.
- [15] T. Sorensen, A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons, Munksgaard in Komm., 1948.
- [16] B. Mirkin, Clustering for Data Mining. A Data Recovery Approach, Chapman & Hall/CRC, 2005.
- [17] K. Sinaga, M. Yang, Unsupervised k-means clustering algorithm, IEEE Access 8 (2020) 80716–80727.