

# A Comparative Study of Clustering Validation Indices and Maximum Entropy for Sintonization of Automatic Segmentation Techniques

J. Hernández, H. Marin and E. Tello

**Abstract**—Automatic image segmentation is a fundamental task in many applications such as video surveillance, image retrieval, medical image analysis, recognition, tracking and objects classification. This task is not easy due to the complexity of image features as well as the number of objects within the images which is unknown in most of the time. However, the correct tuning of the parameters associated with the automatic segmentation algorithms can improve the levels of precision in the object segmentation in the images. This paper proposes the use of clustering validation indices and maximum entropy as cost functions to quantify the quality of segmentation in order to find the optimal parameters of the automatic segmentation techniques. The impact of using clustering validation indices and maximum entropy was evaluated in the automatic segmentation algorithms, K-Means, Watershed and Statistical Region Merging (SRM), using four image databases with images containing different numbers and sizes of objects and different levels of illumination. The results obtained reveal that the use of clustering validation indices and maximum entropy for tuning the automatic segmentation algorithms results in better segmentations with a number of segmented objects close to what really exists in the images, while these results are competitive with the works reported in the literature.

**Index Terms**—Automatic segmentation, Unsupervised metrics, Validation indices, Maximum entropy, Image features.

## I. INTRODUCCIÓN

**H**OY en día, existe una gran cantidad de imágenes de escenas naturales disponibles a partir de diferentes fuentes de información o dispositivos electrónicos, por ejemplo, el teléfono móvil, cámara fotográfica digital, internet, sistemas de visión, sistemas equipados con una cámara digital, entre otros. Las aplicaciones derivadas del procesamiento automático de estas imágenes (video vigilancia, recuperación de imágenes, detección de personas, etc.) demandan entre otras tareas, la segmentación automática, mediante la cual es posible detectar y reconocer a través de programas de computadora (sin la intervención humana) los objetos contenidos en las imágenes digitales de escenas.

La segmentación de imágenes se define como la tarea de clasificación de todos los elementos o píxeles de una imagen en diferentes grupos que exhiben características similares [1], [2].

J. D. Hernández, Unidad Académica Multidisciplinaria Reynosa-RODHE - Universidad Autónoma de Tamaulipas, México, a2183728004@alumnos.uat.edu.mx.

H. M. Marín, Cátedras CONACYT - Universidad Autónoma de Tamaulipas, Cd. Victoria, Tamaulipas, México, hmarin@conacyt.mx.

E. Tello, Universidad Autónoma de Tamaulipas, Cd. Victoria, Tamaulipas, México, etello@docentes.uat.edu.mx.

Dada la necesidad de procesar grandes cantidades de imágenes con alta heterogeneidad, se han propuesto diversas técnicas de segmentación automática [3], [4], [5], [6], las cuales en la mayoría de los casos no logran alcanzar un nivel de precisión equiparable a la segmentación que se realiza manualmente. Esto se debe a que la precisión de los algoritmos de segmentación depende principalmente de la configuración o sintonización de sus parámetros. Esta configuración es diferente para cada imagen con características muy variantes, por ejemplo, la forma de los objetos, las transformaciones geométricas, el ruido, la iluminación de la escena, los colores, las texturas, entre otras. Aunque existen trabajos [7], [8], [9], [10] que intentan emular la capacidad del ser humano para detectar y reconocer objetos, el progreso que se ha logrado hacia el procesamiento automático para reconocer objetos en imágenes de manera similar al ser humano ha sido lento debido a la dificultad de modelar la gran complejidad del sistema visual.

Por otro lado, la sintonización correcta de los parámetros de las técnicas de segmentación automáticas garantiza una segmentación con mayor precisión. Una de las métricas más usadas en la literatura para medir la calidad de la segmentación, es la máxima entropía [6]. Sin embargo, esta métrica no considera la información de dispersión intra/grupo e inter/grupo entre píxeles, es decir, el porcentaje de similitud que existe entre los píxeles de cada objeto segmentado y el nivel de disimilitud entre los píxeles de grupos disjuntos [10]. Debido a lo anterior, los índices de validación de *clustering* utilizados como métricas para medir la calidad de las segmentaciones permite tomar en cuenta la relación inter/intra-grupo entre los píxeles de la imagen, lo cual significa obtener una alta consistencia entre los elementos de los grupos y un número pequeño de grupos. De acuerdo a nuestro conocimiento, consideramos que estos índices de validación no han sido ampliamente aplicados en la tarea de segmentación.

En este trabajo se presenta un estudio y evaluación del proceso de segmentación automática usando tres de las técnicas de segmentación más populares de la literatura, *K-Means* [1], la *Transformada Watershed (Watershed)* [2] y *Statistical Region Merging (SRM)* [11] [12], optimizando los parámetros de las técnicas (*K-Means* y SRM) usando los índices de validación de *clustering*: *Xie Beni (Xie-Beni)*, *Calinski-Harabasz (CH)* y *Davies-Bouldin (DB)* y la máxima entropía (*maxE*) [13], [14]. En el caso de la técnica de segmentación *Watershed*, siempre se usa la misma configuración, para medir su desempeño y compararlo con las técnicas que usan parámetros sintonizados.

Las principales contribuciones de este trabajo son:

- 1) Un estudio del desempeño de técnicas de segmentación automática de imágenes con un número variable de objetos, cuyos parámetros son manualmente configurados por los usuarios.
- 2) Un nuevo enfoque para mejorar la precisión de la tarea de segmentación automática de imágenes, usando índices de validación de *clustering* para encontrar el número óptimo de objetos segmentados en la imagen con una alta consistencia.
- 3) Resultados competitivos que demuestran las ventajas del uso de las métricas de validación de *clustering* y máxima entropía en la tarea de segmentación automática de imágenes.

Una de las aplicaciones más recientes de este trabajo derivada de los resultados obtenidos, es el uso de la segmentación automática de imágenes en áreas como la robótica [15], en donde se requieren de la segmentación de objetos y personas para su identificación o seguimiento. Otra aplicación novedosa es Google Fotos [16], la cual consiste en reconocer formas e interpretar objetos contenidos en imágenes. Esta aplicación tiene la capacidad de separar objetos de cada imagen e identificar aquellos en donde existan objetos similares al objeto seleccionado. Actualmente la aplicación incluye la búsqueda de videos e identificación de objetos en los videos.

El resto de este artículo se encuentra estructurado de la siguiente manera: En la Sección II se presenta el trabajo relacionado. La Sección III provee detalles de cada uno de los algoritmos de segmentación estudiados e implementados. En la Sección IV se detallan las métricas no supervisadas para la sintonización de los parámetros de los algoritmos de segmentación (índices de validación de *clustering* y la máxima entropía) y las métricas supervisadas para la evaluación de la sintonización de los algoritmos de segmentación. En la sección V se describe la metodología para la sintonización de los algoritmos de segmentación automática. La sección VI se presentan los resultados obtenidos a partir de la implementación de los algoritmos y su sintonización. Finalmente, en la sección VII se dan las conclusiones de este trabajo.

## II. TRABAJO RELACIONADO

La segmentación es un concepto clave en el tratamiento de imágenes digitales, esta implica la detección mediante procedimientos de etiquetado deterministas o estocásticos de los contornos o regiones de la imagen, basándose en la información de intensidad y la información espacial [17]. Para las imágenes en escala de grises, una de las técnicas de segmentación clásica y fácil de implementar supone la determinación de agrupamientos (*clusters*) del histograma de la escala de grises, permitiendo la clasificación de los píxeles en regiones homogéneas con la misma, menor o mayor intensidad de gris [6] o bien segmentar la imagen con base a un rango de valores de los píxeles [18]. Estas técnicas de segmentación pueden generalizarse para trabajar con imágenes multiespectrales, tomando en cuenta que la segmentación de las imágenes a color es considerada una tarea difícil, dada la heterogeneidad de las características presentes en las imágenes. Por ello, se

han propuesto varios trabajos para llevar a cabo esta tarea, los cuales son categorizados en tres grupos: a) Espacio ciego: estos algoritmos algoritmos de segmentación usan *clustering* y umbralización por medio del histograma de grises [1], [2], b) Espacialmente guiados: están basados en regiones y en la energía de la imagen [19], [20] y c) Diversos: algoritmos que usan lógica difusa y programación lineal [12].

Dentro del primer grupo, una de las técnicas más utilizadas es la que presentan en el trabajo [1], los autores implementan el algoritmo de *clustering K-Means* para la tarea de segmentación, teniendo como principal desventaja el establecer el número de *clusters*, que es diferente para cada imagen. Debido a esto, resulta necesario usar otra técnica para obtener el número óptimo de *clusters* por cada imagen. Otro trabajo, presentado en [2] se basa en la simulación de la inundación de una zona, donde todas aquellas cordilleras o bordes que se encuentren por encima de un cierto umbral son consideradas como líneas divisorias del objeto. Esta analogía se asume sobre imágenes que contienen regiones con intensidad similar, las cuales son consideradas como inundaciones y los píxeles con mayor intensidad se consideran como las cordilleras o bordes. El problema principal de este tipo de técnicas de segmentación de imágenes es la creación de falsos bordes debido a que los objetos pueden compartir ciertos niveles de gris. Este tipo de técnicas se aplican principalmente sobre imágenes de relieves.

Dentro del segundo grupo, una de las técnicas más utilizadas es *Normalized Cut Graph* [20], la cual hace uso de grafos dirigidos y los segmentos con cierta similitud tienen una relación ponderada. Uno de los principales problemas que enfrenta este tipo de técnica de segmentación es su complejidad computacional, ya que se requiere en la mayoría de los casos de la creación de un número grande de nodos. Esto genera la necesidad de redimensionar el tamaño de la imagen a una escala menor, lo cual puede generar pérdida de información. Otro enfoque es presentado en [19], donde la imagen se divide en pequeños segmentos o bloques, y con la ayuda de una bolsa de palabras cada segmento contiene un concepto con el cual se puede obtener una relación entre segmentos y conceptos. Esto se realiza para agrupar los segmentos pertenecientes a un mismo objeto.

Dentro del tercer grupo, en el trabajo [12] se presenta un estudio sobre once algoritmos de segmentación automática de imágenes de los 3 grupos antes mencionados, seleccionando los algoritmos: *Ultra-metric Contour Map* (UCM), *Gradient Segmentation* (GSEG), *Multiresolution Extension of the GSEG* (MAPGSEG) y *Gibbs Random Field* (GRF) como los algoritmos con el mejor desempeño. Sin embargo, estos algoritmos de segmentación son complejos y computacionalmente costosos.

## III. ALGORITMOS DE SEGMENTACIÓN AUTOMÁTICA

En esta sección se presenta una descripción de los algoritmos de segmentación automática bajo estudio. Los algoritmos *K-Means* y SRM reciben como parámetro un valor entero,  $K$  (número de clusters) y  $Q$  (número de distribuciones de cada píxel) respectivamente, el cual es necesario optimizar en imágenes que contienen un número de objetos variable. Por

otro lado, el algoritmo *Watershed* no recibe ningún parámetro, por lo que siempre usa la misma configuración para todas las imágenes. A través de la evaluación de estos algoritmos se muestra la importancia y dificultades que se pueden llegar a tener al usar técnicas de segmentación automática con o sin sintonización de sus parámetros.

#### A. Algoritmo *K-Means*

El algoritmo *K-Means* [1] es una técnica de segmentación ampliamente utilizada y fácil de implementar, cuya limitante radica en encontrar el número óptimo  $K$  de *clusters* para cada imagen. En el proceso de segmentación de una imagen, los algoritmos de segmentación consideran una imagen  $I(X)$  con  $N$  píxeles, donde  $X = (x, y)$ ,  $X \in R^d$  representa las coordenadas de los píxeles en la imagen y cada píxel se denota como  $x_i$ ,  $i = \{1, 2, \dots, N\}$ ,  $R^d$  es la representación de la imagen en el espacio de color RGB, donde  $d = 3$ .

El objetivo de *K-Means* es dividir el conjunto de datos en subconjuntos disjuntos (*clusters*)  $C_1, \dots, C_K$ , de tal manera que se optimice un criterio de agrupamiento. El criterio más utilizado por el algoritmo *K-Means* es minimizar la suma de las distancias euclidianas al cuadrado entre cada punto de datos  $x_i$  y el centroide  $m_k$  (centro de agrupación) del subconjunto  $C_k$  que contiene a  $x_i$ . Este criterio se le denomina error de agrupación  $E$ , el cual depende de los centros de los *clusters*  $m_1, \dots, m_K$  y se calcula utilizando (1).

$$E(m_1, \dots, m_k) = \sum_{i=1}^N \sum_{j=1}^K F(x_i \in C_j) \|x_i - m_j\|^2 \quad (1)$$

donde  $F(x_i \in C_j) = 1$ , si  $x_i$  existe en el conjunto  $C_j$  y 0 en caso contrario.

#### B. Transformada *Watershed*

*Watershed* [12], [21] es una técnica de segmentación basada en la morfología matemática de la imagen, que permite extraer las fronteras de las regiones que hay en una imagen y clasificar los píxeles de la imagen según su proximidad espacial, el gradiente de sus niveles de gris y la homogeneidad de sus texturas. Por esto, es considerada como una técnica de detección de contornos y crecimiento de regiones al mismo tiempo. Intuitivamente esta técnica, puede considerar una imagen en escala de gris como la imagen topográfica de un relieve terrestre, en donde a cada píxel se le asocia como valor de altura su nivel de gris correspondiente. En este sentido, se puede asumir que las intensidades de gris de mayor amplitud corresponden a llanuras o montañas, mientras que las intensidades de menor amplitud o valor corresponden a valles y ríos. *Watershed* incorpora un proceso de inundación de los valles, desde los niveles más bajos de altura (valores mínimos locales, que constituyen cuencas de inundación rodeadas por cadenas montañosas), hasta los niveles más altos de altura. En el Algoritmo 1 se presenta la descripción de la Transformada *Watershed*. Este algoritmo recibe como entrada una imagen en escala de grises e inicialmente todos los píxeles comienzan con una etiqueta diferente. Para cada etiqueta se obtienen sus

píxeles adyacentes y se agregan a una cola de prioridad con base a su intensidad de escala grises, quedando en primer lugar los píxeles con mayor intensidad. Posteriormente, se extrae el primer elemento de la cola y se verifica si sus píxeles adyacentes tienen la misma intensidad, si es así, los píxeles adyacentes con la misma intensidad se etiquetan con la etiqueta del primer elemento de la cola. Los píxeles adyacentes que no cumplen la similitud son etiquetados como división. Este proceso se repite hasta que la cola esté vacía. Al final, se obtiene como resultado la imagen etiquetada, donde cada objeto tendrá una etiqueta y será separado del resto por los píxeles de división.

---

**Algoritmo 1** Transformada *Watershed* para la segmentación de imágenes.

---

**Entrada:**  $f_i$ : imagen a escala de grises

**Salida:**  $f_0$ : imagen segmentada.

Inicialización: Todos los píxeles se inician con una etiqueta diferente  $L_i$ .

```

1: para i=1 hasta |L| hacer
2:   para j=1 hasta  $p_j \leftarrow |\text{vecinos de } L_i|$  hacer
3:      $p_j$  se agrega a una cola C de prioridad.
4:   fin para
5:   para K=1 hasta |C| hacer
6:     Extraer  $p_{max} \leftarrow$  de C(1)
7:     si intensidad g de  $p_{max} ==$  g,  $g \in L_i$  entonces
8:        $p_{max}$  en  $f_0$  es marcado con la etiqueta de  $L_i$ .
9:     si no
10:       $p_{max}$  en  $f_0$  es marcado como división.
11:    fin si
12:  fin para
13: fin para
14: regresar  $f_0$ 

```

---

#### C. Statistical Region Merging (SRM)

SRM [11] es una técnica de segmentación automática basada en el crecimiento y fusión de regiones utilizando un método estadístico. SRM plantea la segmentación de imágenes como un problema de inferencia, en donde evalúa los valores dentro de una región. Estos valores se agrupan con base a criterios de fusión, donde los píxeles de la imagen tienen las mismas propiedades de homogeneidad en los tres canales de color, resultando una lista menor de regiones.

En la reconstrucción de regiones sobre la imagen observada, SRM depende de dos aspectos: el predicado de fusión y el orden para probar la fusión de regiones.

El predicado de fusión son reglas que debe de cumplir una región para fusionarse con otra región. Inicialmente se tiene una imagen observada  $I$ , que contiene  $N$  píxeles, cada píxel contiene valores *Red-Green-Blue* (RGB), donde cada una de las 3 componentes pertenecen al conjunto  $\{1, 2, 3, \dots, g\}$ , donde  $g$  tiene el valor máximo de 256. El canal de color observado se muestrea a partir de una familia de distribuciones  $Q$  en cada píxel de una escena perfecta  $I^*$ , el rango de  $Q$  es estimado por  $g/Q$ . En  $I^*$ , las regiones óptimas comparten una propiedad de homogeneidad:

- Dentro de una región, los píxeles estadísticos tienen la misma posibilidad para cada canal de color.
- Diferentes regiones tienen posibilidades distintas para al menos un canal de color. Dentro de una región, todas las distribuciones asociadas a cada píxel pueden ser diferentes, siempre que se cumpla la propiedad de homogeneidad. El orden de fusión define un orden a seguir para comprobar el predicado de fusión.

Los pasos principales del algoritmo SRM se muestran en el Algoritmo 2, donde se evalúa si las semillas de los pares de píxeles están en la misma posición y si satisfacen el predicado de fusión. Si sus posiciones no son idénticas y satisfacen el predicado de fusión ( $S < ThI$ ), entonces los dos píxeles se fusionan y el área se actualiza con la suma de los dos píxeles.

---

**Algoritmo 2** Flujo de trabajo del algoritmo SRM.
 

---

**Entrada:**  $f_i$ : imagen RGB

**Salida:**  $f_0$ : imagen segmentada.

- 1: Definir la función de ordenamiento y fusión
  - 2: Inicializar pares-píxeles
  - 3: **mientras**  $S < ThI$  **hacer**
  - 4:  $f_0 \leftarrow fusionar(S)$ .
  - 5: **fin mientras**
  - 6: regresar  $f_0$
- 

Después de exponer las características principales de los tres algoritmos de segmentación seleccionados, a continuación, se presenta la descripción de los índices de validación de *clusters* y la máxima entropía como métricas de sintonización de los parámetros de los algoritmos de segmentación.

#### IV. MEDIDAS DE EVALUACIÓN

##### A. Medidas No Supervisadas

Los índices de validación de *clustering* y la máxima entropía que se presentan en esta sección se consideraron como métricas para sintonizar automáticamente los parámetros que reciben cada uno de los algoritmos de segmentación de imágenes. Estas métricas son también llamadas métricas no supervisadas, cuyo objetivo es evaluar la calidad de un agrupamiento o segmento, así como la cercanía de los píxeles homogéneos. En comparación con las métricas supervisadas, estas métricas no necesitan de una imagen de referencia para cuantificar la calidad de la segmentación [22]. A continuación, se presentan las métricas de validación de *clustering* y la máxima entropía.

Para una mejor comprensión de las variables que se definen en cada índice, se presentan las siguientes definiciones [10]:

- $x_i$ : es un objeto o vector,  $x = [x_1, x_2, \dots, x_d]$  con  $x_i$  características en un espacio  $d$ -dimensional.
- $X$ : es un conjunto con  $N$  objetos denotado como  $X = \{x_1, \dots, x_N\} \in R^d$ .
- $C$ : es un agrupamiento de la forma  $C = \{c_k | \forall k = 1, 2, \dots, K\}$ , donde  $K$  denota el número de grupos disjuntos en  $X$ .
- $n_k$ : es el número de objetos en un grupo  $c_k$ : tal que  $n_k = |c_k|$ .

- $\overline{C}_k$ : es el centroide de un grupo o *cluster*  $c_k$  y es expresado como:  $\overline{C}_k = 1/n_k \sum_{x_i \in c_k} x_i$ .
- $\overline{X}$ : es el centroide del dataset  $X$  y se expresa como:  $\overline{X} = 1/N \sum_{x_i \in X} x_i$ .
- $de(a, b)$ : representa la distancia euclidiana entre dos elementos.

- 1) *Índice Xie-Beni* (Xie-Beni) [23]: es un índice de validez que mide la variación total de un objeto  $x_i$  con respecto a su centroide  $\overline{C}_k$  y la mínima separación entre los centroides de dos agrupamientos  $\overline{C}_k$  y  $\overline{C}_r$ . Este índice está dado por (2).

$$Xie - Beni(C) = \frac{\sum_{c_k \in C} \sum_{x_i \in c_k} de(x_i, \overline{C}_k)}{N \min_{c_k, c_r \in C} \{de(\overline{C}_k, \overline{C}_r)\}} \quad (2)$$

- 2) *Índice Calinski-Harabasz* (CH) [23]: es un índice de validez que se define en términos de matrices de dispersión intra e inter-grupo. Este índice permite estimar la cohesión a través de la suma de las distancias del  $x_i$  a su respectivo centroide  $\overline{C}_k$  y la separación es medida, como la suma de las distancias al centroide global del dataset  $\overline{X}$ , este índice está dado por (3).

$$CH(C) = \frac{N - K}{K - 1} \cdot \frac{\sum_{c_k \in C} n_k de(\overline{C}_k, \overline{X})}{\sum_{c_k \in C} \sum_{x_i \in c_k} de(x_i, \overline{C}_k)} \quad (3)$$

- 3) *Davies-Bouldin* (DB) [23]: fue creado por Davies y Bouldin en 1979, es un índice de validez que mide la dispersión y similaridad entre grupos (distancia inter-/intra grupo). En ese sentido, la cohesión entre los grupos  $c_k \in C$  se estima por la distancia media de los objetos  $x_i$  a su respectivo centroide  $\overline{C}_k$  y la separación es cuantificada como la distancia entre los centroides de los grupos  $\overline{C}_k$  y  $\overline{C}_r$ . El índice DB está dado por (4):

$$DB(C) = \frac{1}{K} \sum_{c_k \in C} \max_{c_r \in C \setminus c_k} \left( \frac{S(c_k) + S(c_r)}{de(\overline{C}_k, \overline{C}_r)} \right) \quad (4)$$

$$\text{donde } S(c_k) = \frac{1}{n_k} \sum_{x_i \in c_k} de(x_i, \overline{C}_k)$$

- 4) *Máxima Entropía* (maxE) [13], [14]: La entropía de una imagen  $I$  en escala de grises con  $K$  regiones, donde  $V_j$  es el conjunto de todos los posibles valores de intensidad en la región  $j$ , mientras que  $L_m$  representa el número de píxeles en la región  $R_j$  que tiene la intensidad  $m$ . La máxima entropía está dada por (5).

$$\text{maxE} = H_{lay}(I) + H_{ere}(I) \quad (5)$$

donde  $H_{lay}(I)$  y  $H_{ere}$  están dadas por (6) y (7) respectivamente,  $S_I$  representa el total de píxeles de la imagen y  $S_j$  es total de píxeles de la región  $j$ .

$$H_{lay}(I) = - \sum_j^K \log\left(\frac{L(R_j)}{S_I}\right) \quad (6)$$

$$H_{ere}(I) = - \sum_j^K \frac{L(R_j)}{S_I} \left( - \sum_{m \in V_j^p} \frac{L_j(m)}{S_j} \cdot \log \frac{L_j(m)}{S_j} \right) \quad (7)$$

Un valor cercano a 0 de los índices Xie-beni y DB se traduce como una buena segmentación, mientras que el índice CH y maxE se busca maximizar su valor lo cual se traduce como una buena segmentación.

Por otro lado, las medidas supervisadas *Normalized Mutual Information* (NMI) [10], *Adjusted Rand Index* (ARI) [24], *Jaccard Index* [25], *Rand Index* (RI) [26], y *Variation of Information* (VI) [27], fueron utilizadas para evaluar la segmentación obtenida por cada algoritmo de segmentación, usando como imagen de referencia una imagen manualmente segmentada.

## V. METODOLOGÍA PARA LA SINTONIZACIÓN DE LOS ALGORITMOS DE SEGMENTACIÓN

Los algoritmos de segmentación *K-Means* y SRM trabajan con un parámetro de entrada  $K$  y  $Q$  respectivamente, donde  $K = \{2, 3, 4, \dots, 10\}$  y  $Q = \{2, 3, 4, \dots, 15\}$ . Estos parámetros fueron seleccionados a partir de las características de las imágenes, tales como, el tamaño de los objetos dentro de las imágenes, el promedio de objetos que contiene cada imagen (aproximado de entre 4.5 objetos por imagen). A partir de estos parámetros utilizados en la sección de experimentación, se percató que mientras el número de objetos segmentados es mayor al promedio de objetos, el desempeño de los algoritmos de segmentación disminuye. En otras palabras, si los parámetros  $K$  y  $Q$  toman un valor muy grande, el desempeño del algoritmo de segmentación se ve afectado y el procesamiento sería innecesario porque al seleccionar la mejor segmentación el valor de  $K$  y  $Q$  estaría cercano al promedio de objetos dentro de las imágenes.

En la Fig. 1 se muestra el flujo de trabajo para la sintonización de los algoritmos *K-Means* y SRM. Inicialmente, se parte de un conjunto de imágenes a partir del cual los algoritmos de segmentación reciben como entrada una imagen y un parámetro  $\sigma$  ( $K$  para *K-Means* y  $Q$  para SRM). Cada algoritmo es ejecutado individualmente para cada valor de  $\sigma$  usando la misma imagen, la salida de los algoritmos es una segmentación propuesta. Una vez obtenido las segmentaciones propuestas, la segmentación de mayor calidad es seleccionada a partir del uso de los índices de validación de *clustering* y máxima entropía. Para cada imagen el valor de  $\sigma$  varía con base a la calidad de la segmentación. Finalmente, la tarea de sintonización de los algoritmos de segmentación es evaluada. En esta evaluación los resultados obtenidos por las métricas supervisadas son promediados. Estos resultados se muestran en la Tabla IV. Las métricas supervisadas miden el grado de similitud que existe entre la mejor segmentación propuesta y la imagen de referencia. Los resultados de la segmentación automática de imágenes realizada por las técnicas *K-Means* y SRM usando las funciones de sintonización Xie-Beni, CH, DB y maxE se muestran en las Tablas I y II. Cada función de sintonización fue evaluada con las cinco métricas supervisadas mencionadas en la Sección IV.

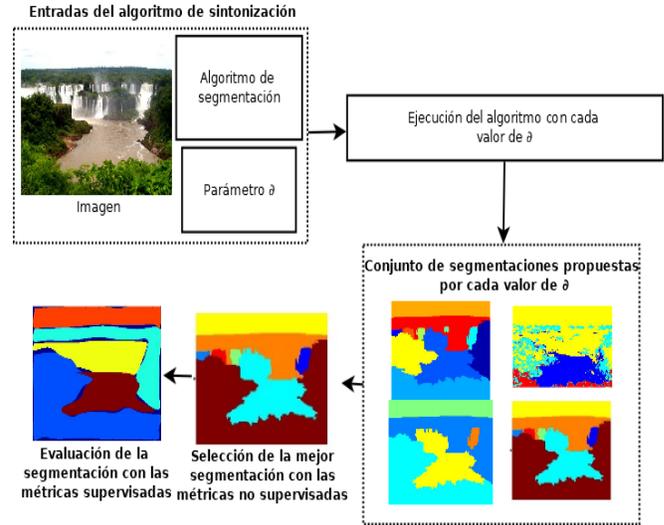


Fig. 1. Flujo de trabajo para la sintonización de los algoritmos de segmentación.

El Algoritmo 3 muestra el proceso de selección del algoritmo de segmentación con el más alto desempeño obtenido, a partir de los algoritmos sintonizados, *K-Means* y SRM, que se representan como *Seg*. Inicialmente, para cada imagen  $I_i$  se ejecuta un algoritmo de segmentación *Seg* usando los valores del parámetro  $\sigma$ . Cada segmentación de *Seg* es almacenada en una lista  $SP_i$ , para posteriormente seleccionar la mejor segmentación de la imagen  $I_i$ , usando una medida no supervisada almacenada en  $MNSS$ . La salida del algoritmo de evaluación es la precisión  $PF$ , la cual mide el desempeño de *Seg*.

**Algoritmo 3** Evaluación de los algoritmos de segmentación automática.

**Entrada:** *Seg*, algoritmo de segmentación,  
 $\sigma = \{1, 2, \dots, n\}$  ( $\sigma = K$  si *Seg*= *K-Means* o  $\sigma = Q$  si *Seg*=SRM),  
 $I = \{I_1, I_2, \dots, I_m\}$ , Imágenes segmentadas manualmente,  
 $IS = \{IS_1, IS_2, \dots, IS_m\}$ , *Seg* Algoritmo de segmentación,  
 $MNS = \{maxE|xie - beni|DB|CH\}$  Métricas no supervisadas,  
 $MS = \{ARI, NMI, VI, JACCARD, RI\}$  Métricas supervisadas.

**Salida:**  $PF$ , precisión del algoritmo de segmentación *Seg*.

- 1: **para**  $i=1$  hasta  $|I|$  **hacer**
- 2:   **para**  $j=2$  hasta  $k$  **hacer**
- 3:      $SP_j \leftarrow sintonizar(Seg, I_i, \sigma_j)$
- 4:      $MNSS_j \leftarrow MNS(SP_j)$ .
- 5:   **fin para**
- 6:    $P_i \leftarrow MS(SP_{MAX(MNSS)}, IS_i)$
- 7: **fin para**
- 8:  $PF \leftarrow \frac{\sum_{i=1}^{|P|} P_i}{|I|}$

## VI. RESULTADOS EXPERIMENTALES

Esta sección presenta los experimentos realizados y los resultados obtenidos usando los dos algoritmos de segmentación automática *K-Means* y SRM sintonizados y el algoritmo *Watershed* a partir de las medidas de validación de *clustering* y máxima entropía. En el proceso de evaluación fueron considerados cuatro datasets de imágenes digitales: 1) *Aerial Image Segmentation Dataset* (ASD) [3], 2) *The Berkeley Segmentation Dataset and Benchmark* (BSR) [5], 3) *SIFT FLOW Dataset* (SFD) [28] y 4) *IAPR-TC12 Benchmark* [29].

La base de datos *Aerial* [3] contiene 80 imágenes aéreas de alta resolución a color. Las imágenes incluyen diferentes escenas como escuelas, residencias, ciudades, almacenes y centrales eléctricas. El tamaño de cada imagen es de 512 x 512 píxeles. La base de datos *The Berkeley Segmentation and Benchmark* [5] contiene 500 imágenes a color con diferentes escenas, incluyendo fauna y flora. En esta base de datos, el tamaño de cada imagen es de 481 x 321 píxeles y en promedio cada imagen contiene 5 objetos. Por otro lado, la base de datos *SIFT FLOW* [28] contiene 2.688 imágenes con 33 tipos de paisajes, que incluye carros, personas, pájaros, etc. El tamaño de las imágenes es de 256 x 256 píxeles y en promedio cada imagen contiene 5 objetos. Finalmente, la base de datos *IAPR-TC12* [29] contiene 20,000 imágenes con 276 tipos de escenas, que incluye animales, fauna, flora, edificios, etc. Para la prueba realizada en esta base de datos solo se tomaron 494 imágenes, que incluyen escenas de flora, fauna, edificios y personas. Las imágenes de *IAPR-TC12* son de tamaño 480 x 360 píxeles y cada imagen contiene en promedio de 4.5 objetos. Estas bases de datos vienen acompañadas con imágenes de referencia (imágenes segmentadas manualmente), las cuales son usadas para evaluar el desempeño de los algoritmos de segmentación usando las medidas supervisadas.

En las Tablas I y II se muestran los resultados obtenidos por los algoritmos de segmentación *K-Means* y SRM. Cada Tabla se compone de 3 filas, que corresponden a las medidas supervisadas ARI, NMI y Jaccard, las cuales miden la similitud de la segmentación entre el algoritmo sintonizado y la imagen segmentada manualmente. Las 4 columnas son los valores obtenidos por cada función de sintonización: maxE y los índices Xie-Beni, CH y DB. Los resultados del algoritmo *Watershed* se muestran en una única tabla (Tabla III), ya que este algoritmo utiliza los mismos parámetros en cada evaluación. De acuerdo con los resultados mostrados en las Tablas I, II y III, el algoritmo SRM, es el algoritmo con el más alto desempeño y estabilidad en todos los conjuntos de datos evaluados de acuerdo con las métricas supervisadas JACCARD, ARI y NMI, usando como función objetivo la máxima entropía. Esto es posible debido a que la técnica de fusión que contempla SRM permite fusionar segmentos a otros segmentos con cierta similitud, evitando segmentaciones detalladas y acercándose al número real de objetos que se encuentran dentro de la imagen. Contrario, al algoritmo *K-Means*, el cual no cuenta con una fusión de objetos a partir de la similitud identificada entre los objetos, por lo que el número de segmentos es igual al parámetro *K* sin importar si existen más o menos objetos en la imagen. En comparación

con *Watershed*, SRM tiene mejores resultados para todas las imágenes. *K-Means* es el segundo algoritmo de segmentación con mejores resultados usando la función de sintonización de máxima entropía.

Para realizar una comparación de nuestro trabajo con trabajos reportados en la literatura, se seleccionó el algoritmo con la mejor segmentación. En este caso, SRM usando como función de sintonización la máxima entropía (**SRM+maxE**) y se comparó con los trabajos [3], [9] y [30], los cuales utilizan las métricas de validación RI y VI, las mismas que se usaron en nuestra evaluación y con los mismos conjuntos de datos. Los resultados de esta comparación se muestran en la Tabla IV.

De acuerdo con los resultados presentados en la Tabla IV, para cada base de datos de imágenes el algoritmo de segmentación seleccionado y sintonizado **SRM+maxE** tiene un desempeño competitivo y en algunos casos superior en comparación a los algoritmos reportados en la literatura. Por ejemplo, en el trabajo presentado en [3], donde se consideran imágenes aéreas de alta resolución, nuestra propuesta presentó mejores resultados, lo que significa que la técnica SRM tiene una precisión competitiva utilizando la máxima entropía como función objetivo para la sintonización de sus parámetros. La medida máxima entropía no penaliza la segmentación con cientos de objetos, es decir, el número de objetos pueden ser el doble de lo que realmente son y la entropía continúa obteniendo resultados aceptables. Por el contrario, las técnicas de validación de *clustering* penalizan una segmentación muy detallada lo cual permite una segmentación más precisa. En este sentido, el número promedio de objetos segmentados usando la máxima entropía es de 8 objetos por imagen en todas las base de datos, mientras que usando los índices de validación de *clustering* se obtuvo un promedio de 4 objetos por imagen en todas las base de datos, siendo los índices de validación de *clustering* más precisos en el número de objetos en promedio que existen en las imágenes.

Por otro lado, los resultados reportados por nuestra propuesta son competitivos en comparación con los obtenidos en [9] y [30], tomando en cuenta que estos trabajos consideran una metodología compleja, la cual requiere de mayor disposición de recursos de cómputo. El trabajo [30] emplea una metodología combinatoria para la segmentación, mientras que el trabajo [9] usa redes neuronales convolucionales, las cuales requieren de un alto tiempo de entrenamiento. En este sentido, la principal contribución de este trabajo de investigación es la sintonización de algoritmos básicos de segmentación automática mediante índices de validación de *clustering* y máxima entropía usando imágenes reales con diferentes características (por ejemplo, color, forma, tamaño, número variable de objetos dentro de la imagen, etc.). Además, de considerar que la sintonización de estos algoritmos permite una segmentación que no requiere de altos recursos de cómputo para obtener resultados prometedores. El tiempo promedio que requiere cada uno de los algoritmos de segmentación: *K-Means*, SRM y *Watershed* en segmentar cada imagen es de 40 Seg, 54 Seg y 1 Seg respectivamente. Mientras que el tiempo total requerido para la segmentación de toda la colección de imágenes de acuerdo con las evaluaciones es de 3:04 Hrs, 4:07

TABLA I  
RESULTADOS OBTENIDOS POR LA TÉCNICA DE SEGMENTACIÓN AUTOMÁTICA K-MEANS

	maxE	Xie-Beni	CH	DB
Base de datos ASD				
ARI	0,03	0,08	0,092	0,057
NMI	0,14	0,12	0,115	0,129
JACCARD	0,53	0,17	0,133	0,296
Base de datos BSR				
ARI	0,223	0,274	0,253	0,263
NMI	0,356	0,340	0,322	0,348
JACCARD	0,547	0,269	0,226	0,373
Base de datos IAPRTC12				
ARI	0,219	0,33	0,326	0,325
NMI	0,363	0,39	0,377	0,394
JACCARD	0,485	0,52	0,485	0,627
Base de datos SIFTFLOW				
ARI	0,168	0,319	0,320	0,282
NMI	0,322	0,352	0,343	0,352
JACCARD	0,351	0,485	0,425	0,571

TABLA II  
RESULTADOS OBTENIDOS POR LA TÉCNICA DE SEGMENTACIÓN AUTOMÁTICA SRM

	maxE	Xie-Beni	CH	DB
Base de datos ASD				
ARI	0,129	0,092	0,087	0,093
NMI	0,248	0,169	0,166	0,176
JACCARD	0,43	0,174	0,160	0,203
Base de datos BSR				
ARI	0,385	0,295	0,279	0,314
NMI	0,509	0,420	0,408	0,440
JACCARD	0,514	0,300	0,263	0,343
Base de datos IAPRTC12				
ARI	0,48	0,445	0,436	0,451
NMI	0,54	0,506	0,498	0,511
JACCARD	0,66	0,630	0,603	0,666
Base de datos SIFTFLOW				
ARI	0,429	0,424	0,423	0,424
NMI	0,484	0,468	0,465	0,472
JACCARD	0,584	0,512	0,470	0,578

TABLA III  
RESULTADOS OBTENIDOS POR LA TÉCNICA DE SEGMENTACIÓN AUTOMÁTICA WATERSHED

Base de datos	ARI	NMI	JACCARD
ASD	0,030	0,168	0,364
BSR	0,137	0,250	0,312
IAPRTC12	0,178	0,252	0,581
SIFTFLOW	0,146	0,170	0,476

Hrs y 55 Min.

En la Fig. 2 se muestra un ejemplo de tres imágenes del conjunto de datos IAPR-TC12 segmentadas automáticamente por el algoritmo SRM utilizando como función objetivo los índices de validación de *clustering* y la máxima entropía.

## VII. CONCLUSIONES

En este trabajo se presentó un estudio del uso de índices de validación de *clustering* y máxima entropía para sintonizar los parámetros de configuración de las técnicas de segmentación

TABLA IV  
COMPARATIVA DE LA TÉCNICA DE SEGMENTACIÓN SELECCIONADA Y TRABAJOS REPORTADOS EN LALITERATURA

Base de datos	ASD		BSR		SIFT FLOW		IAPRT-12	
	ARI	VI	ARI	VI	ARI	VI	ARI	VI
gPb-owt-ucm [3]	0.69	1.0	-	-	-	-	-	-
MCG [30]	-	-	0.86	1.5	-	-	-	-
FCN-16s [9]	-	-	-	-	0.85	-	-	-
<b>SRM+maxE (propuesta)</b>	<b>0.72</b>	<b>1.6</b>	<b>0.82</b>	<b>1.8</b>	<b>0.82</b>	<b>1.1</b>	<b>0.76</b>	<b>1.5</b>

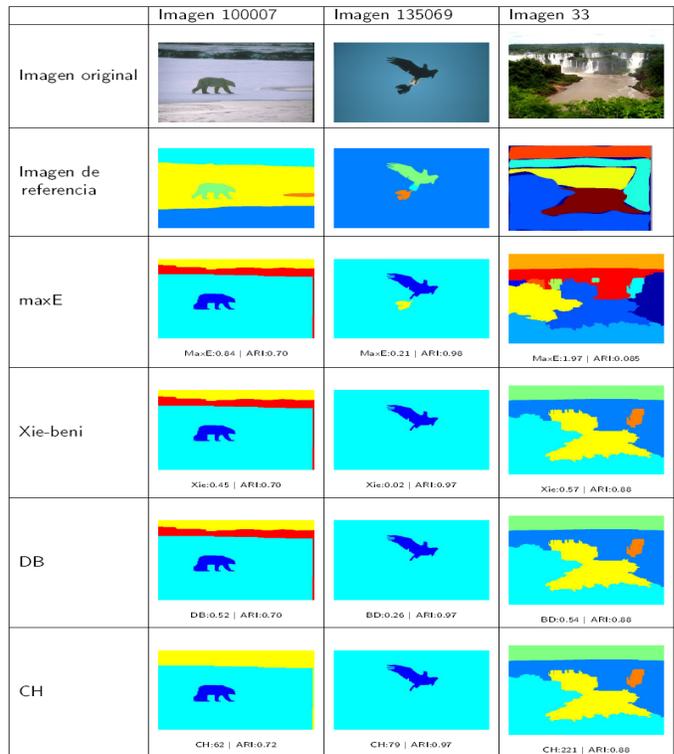


Fig. 2. Ejemplo de imágenes segmentadas, sintonizando el algoritmo SRM con las métricas no supervisadas: Máxima entropía, Xie-beni, CH y DB y la métrica supervisada ARI.

automática de imágenes *K-Means* y SRM. A partir de este estudio se mostró que el algoritmo de segmentación automática SRM usando como métrica de sintonización la máxima entropía, presenta mejores resultados que los algoritmos *K-Means* y *Watershed*, considerando imágenes con diferentes características y con un número variable de objetos dentro de estas. Por otro lado, la sintonización de los parámetros puede verse como una desventaja debido a la dificultad que puede tomar la búsqueda de los parámetros óptimos. Sin embargo, de acuerdo a los resultados obtenidos en este trabajo, el uso de los índices de validación de *clustering* y la máxima entropía permiten mejorar la precisión de los algoritmos de segmentación mediante la identificación del número real de objetos presentes en las imágenes.

## REFERENCIAS

- [1] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern recognition*, vol. 36, no. 2, pp. 451–461, 2003.

- [2] N. Palomino and L. Concepción, "Watershed: un algoritmo eficiente y flexible para segmentación de imágenes de geles 2-de," *Revista de investigación de Sistemas e Informática*, vol. 7, no. 2, pp. 36–41, 2010.
- [3] J. Yuan, S. S. Gleason, and A. M. Cheryadat, "Systematic benchmarking of aerial image segmentation," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 6, pp. 1527–1531, 2013.
- [4] R. S. Marques, A. Conci, M. G. Perez, V. H. Andaluz, and T. M. Mejia, "An approach for automatic segmentation of thermal imaging in computer aided diagnosis," *IEEE Latin America Transactions*, vol. 14, no. 4, pp. 1856–1865, 2016.
- [5] D. Martin, C. Fowlkes, D. Tal, J. Malik *et al.*, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics." *Iccv Vancouver*, 2001.
- [6] J. R. Parker, *Algorithms for image processing and computer vision*. John Wiley & Sons, 2010.
- [7] J. Hernández and W. Gómez, "Automatic tuning of the pulse-coupled neural network using differential evolution for image segmentation," in *Mexican Conference on Pattern Recognition*. Springer, 2016, pp. 157–166.
- [8] T. Lindblad, J. M. Kinser, and J. Taylor, *Image processing using pulse-coupled neural networks: The PCNN and ICM*. Springer, 2005.
- [9] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [10] J. H. López, "Sintonización automática de la red neuronal de pulso acoplado mediante evolución diferencial para la segmentación de imágenes," Master's thesis, Laboratorio de Tecnologías de Información, CINVESTAV-Tamaulipas, Cd. Victoria, Tamaulipas, México., 2016.
- [11] R. Nock and F. Nielsen, "Statistical region merging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1452–1458, 2004.
- [12] S. R. Vantaram and E. Saber, "Survey of contemporary trends in color image segmentation," *J. Electronic Imaging*, vol. 21, p. 040901, 2012.
- [13] W. K. Lai, I. M. Khan, and G. S. Poh, "Weighted entropy-based measure for image segmentation," *Procedia Engineering*, vol. 41, pp. 1261–1267, 2012.
- [14] H. Zhang, J. E. Fritts, and S. A. Goldman, "An entropy-based objective evaluation method for image segmentation," in *Storage and Retrieval Methods and Applications for Multimedia*, 2003.
- [15] K. Lourdes, E. Palacios, and A. Marin, "A fuzzy speed controller for a guide robot using an hri approach," *IEEE Latin America Transactions*, vol. 16, pp. 2102–2107, 2018.
- [16] Google, "Google fotos," [urlhttps://photos.google.com/apps?hl=es](https://photos.google.com/apps?hl=es). EULA, 2015.
- [17] J. Angulo and J. P. F. Serra, "Segmentación de imágenes en color utilizando histogramas bi-variables en espacios color polares luminancia/saturación/matiz," *Computación y Sistemas*, vol. 8, pp. 303–316, 2005.
- [18] M. P. de Albuquerque and I. A. Esquef, "Image segmentation using nonextensive relative entropy," *IEEE Latin America Transactions*, vol. 6, 2008.
- [19] L. YuanLai, "Research on image region classification based on automatic segmentation," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 11, pp. 233–242, 2016.
- [20] and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, Aug 2000.
- [21] M. Gonzalez and V. Ballarin, "Image segmentation using watershed transform: Marker definition based on fuzzy logic," *IEEE Latin America Transactions*, vol. 6, no. 2, pp. 223–228, June 2008.
- [22] J. Fernandez, M. Hurtado, J. Areta, and C. Muravchik, "Polarimetric sar image segmentation using cem algorithm," *IEEE Latin America Transactions*, vol. 12, no. 5, pp. 910–914, Aug 2014.
- [23] B. Balasko, J. Abonyi, and B. Feil, "Fuzzy clustering and data analysis toolbox," *Department of Process Engineering, University of Veszprem, Veszprem*, 2005.
- [24] K. Y. Yeung and W. L. Ruzzo, "An empirical study on principal component analysis for clustering gene expression data," *Bioinformatics*, vol. 17, no. 9, pp. 763–774, 2001.
- [25] M. Levandowsky and D. Winter, "Distance between sets," *Nature*, vol. 234, pp. 34–35, 1971.
- [26] W. Yong and G. Lemaitre, "Evaluation measures for segmentation," *matrix*, vol. 1, p. 2, 2000.
- [27] A. Y. Yang, J. N. Wright, Y. Ma, and S. S. Sastry, "Unsupervised segmentation of natural images via lossy data compression," *Computer Vision and Image Understanding*, vol. 110, pp. 212–225, 2008.
- [28] J. Tighe and S. Lazebnik, "Superparsing: Scalable nonparametric image parsing with superpixels," in *ECCV*, 2010.
- [29] H. J. Escalante, C. A. Hernández, J. A. Gonzalez, A. López-López, M. M. Gómez, E. F. Morales, L. E. Sucar, L. V. Pineda, and M. Grubinger, "The segmented and annotated iapr tc-12 benchmark," *Computer Vision and Image Understanding*, vol. 114, pp. 419–428, 2010.
- [30] P. A. Arbeláez, J. Pont-Tuset, J. T. Barron, F. Marqués, and J. Malik, "Multiscale combinatorial grouping," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 328–335, 2014.



**Jaciel David Hernández Reséndiz** Received the M.E. degree in engineering from Polytechnic University of Victoria, Mexico in 2018, the B.Sc. degree in Information Technology from Polytechnic University of Victoria, Mexico in 2014. Currently, he is doing their studies in the Ph.D in Electrical and Electronic Engineering. His research areas include Machine learning, Computer Vision and Information Retrieval.



**Heidi M. Marin-Castro** Received the Ph.D. degree in Computer Science from Center for Research and Advanced Studies of the National Polytechnic Institute, Mexico (Cinvestav) in 2014 and the M.Sc. degree in Computer Science from the National Institute for Astrophysics, Optics and Electronics (INAOE) in 2008. Currently she is a Conacyt researcher at Information Technology department in the Autonomous University of Tamaulipas, Mexico. Her research areas include Web Data Management, Databases, Data Mining and Information Retrieval.



**Edgar Tello-Leal** Received the PhD degree in Information Systems Engineering from the National Technological University, Santa Fe Faculty, Argentina, in 2012. He is a Full-Time Research-Professor at the Autonomous University of Tamaulipas, Mexico. He is a Researcher at the National System of Researchers of the National Council of Science and Technology, Mexico. His current research interests include business process management, model-driven development, multi-agent systems, Internet of Things, and data mining.