

Imágenes Biomédicas

Tareas II y III

Joel Chacón Castillo

Guanajuato, México

1. *Thresholding Based on Attribute Similarity - Umbralización Basado en la Similitud de Atributos*

Estos algoritmos realizan la selección de un valor de umbralización en base a algunos atributos de calidad o mediciones de similaridad entre la imagen original y la versión binaria de la imagen[1]. Estos atributos pueden ser basados en la coincidencia de los bordes “Edge Matching”, momentos en escala de grises, conectividad, textura, o estabilidad de los objetos segmentados. Algunos otros algoritmos evalúan directamente la semejanza de las distribuciones de probabilidad acumulativa o la semejanza de una imagen original en escala de grises con una imagen binaria por medio de una medida difusa.

Particularmente, en el trabajo Huang and Wang [2] se propone un método donde se minimizó la una medida difusa de una imagen de entrada. La función de pertenencia en este método de umbralización es utilizada para denotar la relación característica entre un pixel y su región de pertenencia (i.e. el objeto o el fondo). Entre mayor sea la diferencia absoluta entre el objeto y el fondo menor es el valor de pertenencia. Además, este método puede ser extendido fácilmente a umbralización multi-nivel.

1.1. *The fuzzy set and membership function - Conjunto difuso y función de pertenencia*

Sea X un conjunto imagen de tamaño $M \times N$ con L niveles, y X_{mn} es el nivel de gris de un pixel (m, n) en X . Sea $\mu_x(X_{mn})$ un valor de pertenencia el cual representa el grado de poseer una cierta propiedad por el pixel (m, n) en X ; esto es, un sub-conjunto difuso del conjunto imagen X puede ser escrito de la siguiente forma.

$$X = \{(X_{mn}, \mu_x(X_{mn}))\} \quad (1)$$

donde $0 \leq \mu_x(X_{mn}) \leq 1$, $m = 0, 1, \dots, M - 1$ y $n = 0, 1, \dots, N - 1$. Particularmente, la función de pertenencia $\mu_X(X_{mn})$ puede comprenderse como una función característica que representa lo difuso de un pixel (m, n) en X . Para el propósito de umbralización de imágenes, cada pixel en la imagen debería poseer una relación cercana con su región de pertenencia, es decir, el objeto o el fondo. Por lo tanto, el valor de pertenencia de un pixel en X puede ser definido utilizando la relación entre el pixel y su región de pertenencia.

Así, siendo $h(g)$ el número de ocurrencias en un nivel de grises g en una imagen de entrada. Además, dado un valor de umbralización t , los niveles de gris promedio del fondo μ_0 y el objeto μ_1 pueden ser obtenidos como se indica a continuación.

$$\mu_0 = \sum_{g=0}^t gh(g) / \sum_{g=0}^t h(g) \quad (2)$$

$$\mu_1 = \sum_{g=t+1}^{L-1} gh(g) / \sum_{g=t+1}^{L-1} h(g) \quad (3)$$

Los promedios de cada nivel de gris μ_0 y μ_1 , pueden ser considerados como los valores objetivo del fondo y el objeto dado un umbral t . Así, la relación entre un pixel en X y su región de pertenencia debería depender intuitivamente en la diferencia de su nivel de gris y el valor objetivo de su región de pertenencia. Por lo tanto, dejando la relación posea la propiedad de que cuanto menor sea la diferencia absoluta entre el nivel de gris de un pixel y su valor objetivo correspondiente, mayor será el valor de pertenencia del pixel. Entonces, la función de pertenencia que realiza la evaluación de lo previamente mencionado para un pixel (m, n) se define a continuación.

$$\mu_x(X_{mn}) = \begin{cases} \frac{1}{1+|x_{mn}-\mu_0|/C} & \text{si } X_{mn} \leq t \\ \frac{1}{1+|x_{mn}-\mu_1|/C} & \text{si } X_{mn} > t, \end{cases} \quad (4)$$

donde C es un valor constante (usualmente $C = g_{max} - g_{min}$) tal que $1/2 \leq \mu_x(X_{mn}) \leq 1$. Para un umbral dado t , cualquier pixel en la imagen de entrada debería pertenecer ya sea al objeto o al fondo. Por lo tanto, se espera que el valor de pertenencia para un pixel debería no ser menor que $1/2$. La función de pertenencia propuesta refleja esta relación entre un pixel y su región de pertenencia.

1.2. Measures of Fuzziness - Mediciones difusas

La medición difusa indica un grado en el conjunto difuso, es decir es una función $f : A \rightarrow \mathfrak{R}$. Se han propuesto varios enfoques para medir lo difuso. En este caso se introducen dos métodos popularmente utilizados. El primero es la medición de la entropía por medio de la función de Shannon y otro es la medición de Yager's por medio de la distancia entre el conjunto difuso y su complemento.

1.2.1. Entropía

La entropía utilizada como una medición difusa se puede ver como una analogía con la entropía en teoría de la información, sin embargo una diferencia muy sutil en la definición. Basado en la función de Shannon, se define la entropía de un conjunto difuso A como se indica a continuación.

$$E(A) = \frac{1}{n \ln 2} \sum_i S(\mu_A(x_i)), \quad i = 1, 2, \dots, n, \quad (5)$$

con la siguiente función de Shannon.

$$S(\mu_A(x_i)) = -\mu_A(x_i) \ln[\mu_A(x_i)] - [1 - \mu_A(x_i)] \ln[1 - \mu_A(x_i)] \quad (6)$$

Extendiendo al caso de dos dimensiones de una imagen plana, la entropía de una imagen X se expresa a continuación.

$$E(X) = \frac{1}{MN \ln 2} \sum_m \sum_n S(\mu_x(X_{mn})) \quad (7)$$

con $m = 0, 1, \dots, M - 1$ y $n = 0, 1, \dots, N - 1$. Finalmente, utilizando la información del histograma, la medición puede ser vista de la siguiente forma.

$$E(X) = \frac{1}{MN \ln 2} \sum_g S(\mu_x(g)) h(g) \quad g = 0, 1, \dots, L - 1 \quad (8)$$

Es importante notar que la función de Shannon es monótonicamente creciente en el intervalo $[0, 0,5]$ y decreciente en el intervalo $[0,5, 1]$. Así, cuando $\mu_x(X_{mn}) = 0,5$ para todos los pixeles, la entropía E será máxima.

1.2.2. Medición de Yager

La mayor distinción entre el conjunto difuso y la tendencia tradicional es que el enfoque de conjuntos difusos no siempre satisface la ley de excluir el medio. Así, Yager argumenta que la medición difusa debería depender en la relación que existe entre un conjunto difuso A y su complemento \bar{A} . La distancia entre el conjunto difuso de una imágen X y su complemento \bar{X} es definido como se indica a continuación.

$$D_p(X, \bar{X}) = \left[\sum_m \sum_n |\mu_x(X_{mn}) - \mu_{\bar{x}}(X_{mn})|^p \right]^{\frac{1}{p}} \quad (9)$$

donde $\mu_x(X_{mn}) = 1 - \mu_x(\bar{X}_{mn})$. Por lo tanto, la medición difusa de X puede ser denotada como:

$$\eta_p(X) = 1 - \frac{D_p(X, \bar{X})}{|X|^{\frac{1}{p}}} = 1 - \frac{D_p(X, \bar{X})}{(MN)^{\frac{1}{p}}} \quad (10)$$

Para simplificar los cálculos, se puede utilizar el histograma de la siguiente forma:

$$D_p(X, \bar{X}) = \left[\sum_g |\mu_x(g) - \mu_{\bar{x}}(g)|^p \right]^{\frac{1}{p}} h(g) \quad (11)$$

Es importante notar que para $p = 1$ se utiliza la métrica de hamming, y para $p = 2$ es considerada la métrica Euclidieana. Así, para una imágen X , se espera que la medición difusa debería ser tan pequeña como sea posible.

2. A Threshold Selection Method from Gray-Level Histograms - Un Método para la Selección de Umbrales Basado en Histogramas de Escalas de Grises (Otsu)

En este trabajo se presenta un método automático y no-paramétrico para la selección de un umbral cuyo propósito es la segmentación de imágenes propuesto por Otsu [3]. El procedimiento automatizado para la selección del umbral óptimo es seleccionado en base a el criterio discriminante, donde se desea maximizar la separabilidad de las clases resultantes en escala de grises. El procedimiento es muy simple, y además utiliza el primer momento acumulativo (media acumulativa) del histograma en escala de grises. Dados los pixels de una imágen y representados en escalas de grises $[1, 2, \dots, L]$.

El número de pixeles en el nivel i es denotado por n_i y el número total de pixeles por $N = n_1 + n_2 + \dots + n_L$. Además, el histograma en escala de grises es normalizado y representado como una distribución de probabilidad:

$$p_i = n_i/N \quad p_i \geq 0, \quad \sum_{i=1}^L p_i = 1 \quad (12)$$

Así, en base a un umbral k se separan los pixeles en dos clases C_0 y C_1 (fondo y objetos de interés). Donde C_0 denota los pixeles con niveles $[1, \dots, k]$, y C_1 denota los pixeles con niveles $[k+1, \dots, L]$. Entonces, la probabilidad de ocurrencia de cada clase y las medias de cada clase son indicados de la siguiente forma:

$$\omega_0 = Pr(C_0) = \sum_{i=1}^k p_i = \omega(k) \quad (13)$$

$$\omega_1 = Pr(C_1) = \sum_{i=1}^k p_i = 1 - \omega(k) \quad (14)$$

$$\mu_0 = \sum_{i=1}^k iPr(i|C_0) = \sum_{i=1}^k ip_i/\omega_0 = \mu(k)/\omega(k) \quad (15)$$

$$\mu_1 = \sum_{i=k+1}^L iPr(i|C_1) = \sum_{i=k+1}^L ip_i/\omega_1 = \frac{\mu_T - \mu(k)}{1 - \omega(k)} \quad (16)$$

donde

$$\begin{aligned} \omega(k) &= \sum_{i=1}^k p_i \\ \mu(k) &= \sum_{i=1}^k ip_i \end{aligned} \quad (17)$$

corresponden al acumulativo del cero y primer momento respectivamente del histograma generado por el umbral k . Además, $\mu_T = \mu(L) = \sum_{i=1}^L ip_i$ es la media total de la escala de grises que pertenece a la imagen. Así, se puede verificar la siguiente relación para cualquier orden k

$$\omega_0\mu_0 + \omega_1 + \mu_1 = \mu_T, \quad \omega_0 + \omega_1 = 1 \quad (18)$$

Las varianzas o momentos acumulativos de segundo orden de cada clase son determinados de la siguiente forma:

$$\sigma_0^2 = \sum_{i=1}^k (i - \mu_0)^2 Pr(i|C_0) = \sum_{i=1}^k (i - \mu_0)^2 p_i / \omega_0 \quad (19)$$

$$\sigma_1^2 = \sum_{i=1}^k (i - \mu_1)^2 Pr(i|C_1) = \sum_{i=k+1}^L (i - \mu_1)^2 p_i / \omega_1 \quad (20)$$

Así, para cualificar un determinado umbral se utiliza un criterio de medición del discriminante (i.e. mide la separabilidad de las clases):

$$\lambda = \sigma_B^2 / \sigma_W^2, \quad k = \sigma_T^2 / \sigma_W^2, \quad \eta = \sigma_B^2 / \sigma_T^2, \quad (21)$$

donde

$$\begin{aligned} \sigma_w^2 &= \omega_0 \sigma_0^2 + \omega_1 \sigma_1^2 \\ \sigma_B^2 &= \omega_0 (\mu_0 - \mu_T)^2 + \omega_1 (\mu_1 - \mu_T)^2 \\ &= \omega_0 \omega_1 (\mu_1 - \mu_0)^2 \end{aligned} \quad (22)$$

Además se tiene que $\sigma_T^2 = \sum_{i=1}^L (i - \mu_T)^2 p_i$, es decir σ_w es la varianza intra-clase y σ_T es la varianza entre-clases. De esta forma el problema se convierte en uno de optimización donde se obtenga el umbral k el cual maximiza cualquier criterio de medición λ, K, η . El criterio para discriminar basado en la maximización de λ, K, η para un k son equivalentes, es decir, basta con maximizar uno dada la siguiente relación básica:

$$\sigma_w^2 + \sigma_B^2 = \sigma_T^2 \quad (23)$$

Se puede observar que σ_w^2 y σ_B^2 son función del umbral k , pero σ_T^2 es independiente de k . Además, σ_w^2 se basa en estadísticas de segundo orden (varianza de las clases), mientras σ_B^2 se basa en la estadística de primer orden (media de las clases). Por lo tanto, η es la medición más simple con respecto al umbral. Entonces, se utiliza η como un criterio de medición para medir la calidad de separación dado un umbral k .

El umbral óptimo k^* el cual maximiza η , o equivalentemente, el que maximiza σ_B^2 es seleccionado en base a una búsqueda secuencial utilizando $\eta(k) = \sigma_B^2(k) / \sigma_T^2$. Por lo tanto el umbral óptimo k^* es determinado en base a la siguiente función objetivo:

$$\sigma_B^2(k^*) = \max_{1 \leq k < L} \sigma_B^2(k) \quad (24)$$

Así, la función objetivo que corresponde a la forma generalizada o multi-clase se puede representar como sigue:

$$\sigma_B^2(k_1^*, \dots, k_{NC}^*) = \arg \max_{1 \leq k_1, \dots, k_{NC}} \sum_{i=1}^{NC} \omega_i (\mu_i - \mu_T)^2 \quad (25)$$

donde NC es el número de clases que se desean separar en la imagen.

3. A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram - Un Nuevo Método para el Umbral de Imágenes de Nivel de Gris Mediante la Entropía del Histograma

Los métodos más comunes para la extracción de objetos de una imagen son basados en un umbral. Si el objeto se distingue de forma clara del fondo, sin embargo el histograma que pertenece a la imagen en escala de grises no siempre posee el comportamiento de bimodalidad. En base a este inconveniente Pun [4] propone utilizar el ratio entre la entropía a posteriori H'_n y la entropía H_n que es acotada por un funcional $f(s)$, definidos de la siguiente forma.

$$H'_n = -P_s \ln P_s - (1 - P_s) \ln (1 - P_s) \quad (26)$$

donde $P_s = \sum_{i=1}^s p_i$, $1 - P_s = \sum_{i=s+1}^n p_i$ por su parte la entropía se define de la siguiente forma

$$H_n = - \sum_{i=1}^n p_i \ln p_i \quad (27)$$

El resultado más importante de ese trabajo es establecer la siguiente relación

$$\left(\phi(s) = \frac{H'_n}{H_n} \right) \leq f(s) \quad (28)$$

donde argumentan que $f(s)$ mayoriza $\phi(s)$ asumiendo un error como se observa en la siguiente imagen.

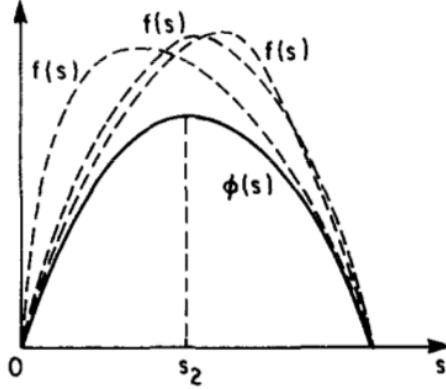


Figura 1: Relación entre el funcional $f(s)$ y el ratio de entroías $\phi(s)$.

Por otra parte varios años después Kapur et al. [5] demostraron que el funcional $f(s)$ tenía inconvenientes importantes:

- La maximización de $f(s)$ no logra una maximización a priori de una entropía a posteriori H'_n .
- Desde que H'_n es mayorizada por otra función, el algoritmo no utiliza las propiedades estadísticas del histograma de escalas de grises.

Por lo tanto Kapur et al. [5] proporcionan un enfoque el cual se basa en la sumatoria de las entropías obtenidas por cada clase como se indica a continuación. Dadas las distribución en escala de gris p_1, p_2, \dots, p_n y dado un umbral s (se utiliza la misma notación que en el trabajo propuesto), se derivan dos distribuciones de probabilidad de la siguiente forma.

$$A : \frac{p_1}{p_s}, \frac{p_2}{p_s}, \dots, \frac{p_s}{p_s} \quad (29)$$

$$B : \frac{p_{s+1}}{1-p_s}, \frac{p_{s+2}}{1-p_s}, \dots, \frac{p_n}{1-p_s} \quad (30)$$

Además, la entropías asociadas con cada distribución se definen a continua-

ción:

$$\begin{aligned}
H(A) &= - \sum_{i=1}^s \frac{p_i}{P_s} \ln \frac{p_i}{P_s} \\
&= -\frac{1}{P_s} \left[\sum_{i=1}^s p_i \ln p_i - p_i \ln P_s \right] \\
&= \ln P_s + \frac{H_s}{P_s}
\end{aligned} \tag{31}$$

similarmente:

$$\begin{aligned}
H(B) &= - \sum_{i=1+s}^n \frac{p_i}{1-P_s} \ln \frac{p_i}{1-P_s} \\
&= -\frac{1}{1-P_s} \left[\sum_{i=1+s}^n p_i \ln p_i - (1-p_i) \ln (1-P_s) \right] \\
&= \ln (1-P_s) + \frac{H_n - H_s}{1-P_s}
\end{aligned} \tag{32}$$

Entonces definiendo la suma de $H(A)$ y $H(B)$ como $\phi(s)$:

$$\phi(s) = \ln P_s (1-P_s) + \frac{H_s}{P_s} + \frac{H_n - H_s}{1-P_s} \tag{33}$$

Por lo tanto se desea maximizar $\phi(s)$ para obtener la máxima información entre la distribución del objeto y la distribución del fondo en la imagen. La forma multi-nivel se obtiene maximizando la siguiente función objetivo

$$\phi(s_1^*, \dots, s_k^*) = \max_{s_1, \dots, s_k} \sum_{i=1}^k H(X_i) \tag{34}$$

donde k es el número de niveles que se desean identificar.

4. Indicadores de Calidad

En el análisis de imágenes médicas es necesario identificar o segmentar estructuras de su fondo. Es importante entender que no existe una técnica universal de segmentación para lidar con todo tipo de imágenes médicas. Por lo tanto, encontrar una algoritmo de segmentación para una específica tarea como elegir los parámetros óptimos aún es un reto. Por otra parte, escoger una medición de la efectividad dado un algoritmo aún es una tarea difícil. Sin embargo, se presentan cinco métricas popularmente consideradas [6]

4.1. Accuracy

La exactitud de una técnica de segmentación se refiere al grado en el cual los resultados dado un método de segmentación concuerdan con el resultado de segmentación real. Así, esta métrica se encarga de medir la proporción de positivos y negativos que fueron correctamente identificados por un clasificador. El dominio de esta métrica es $[0, 1]$, donde 0 es la peor clasificación y 1 es la mejor clasificación.

$$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)} \quad (35)$$

4.1.1. Sensitivity - True Positive Rate or True Positive Fraction

Se encarga de medir la proporción de positivos que fueron correctamente identificados por algún clasificador. Se encuentra definida en el intervalo $[0, 1]$, donde 0 es la peor clasificación y 1 es la clasificación perfecta.

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (36)$$

4.2. Specificity - True Negative Rate or True Negative Fraction

Se encarga de medir la proporción de negativos que fueron correctamente identificados por algún clasificador. El dominio de esta métrica es $[0, 1]$ donde 0 es la peor clasificación y 1 es la clasificación perfecta.

$$Specificity = \frac{TN}{(TN + FP)} \quad (37)$$

4.3. Positive Predictive Value - PPV

Mide la proporción de positivos que fueron correctamente e incorrectamente identificados por un clasificador. Está definido en el dominio $[0, 1]$, donde 0 es la peor clasificación y 1 es la clasificación perfecta.

$$PPV = \frac{TP}{(TP + FP)} \quad (38)$$

4.4. Negative Predictive Value - NPV

Mide la proporción de negativos que fueron correctamente identificados por algún clasificador. Está definido en el intervalo $[0, 1]$, donde 0 es la peor clasificación y 1 es la clasificación perfecta.

$$NPV = \frac{TN}{(TN + FN)} \quad (39)$$

5. Evolución Diferencial en problemas de imágenes

Evolución Diferencial (ED) fue inicialmente propuesto por Price et al. [7], sin embargo en este mismo trabajo se explica que este algoritmo es una variante aleatorizada del algoritmo “Nelder Mead” [8]. Esta sección está dedicada a revisar la variante clásica de DE y a introducir varios términos importantes que son utilizados en el campo de DE. DE es un algoritmo estocástico basado en una población, por lo que en cada instante maneja un conjunto de soluciones candidatas que van evolucionando de forma iterativa. En DE dichas soluciones candidatas son usualmente conocidas como vectores. En la variante básica de DE, para cada miembro de la población (conocidos como *vectores objetivo*) se genera un nuevo vector que es conocido como el *vector mutado*. A continuación, el vector mutado se combina con el vector objetivo para generar al *vector de prueba* y finalmente se procede con la fase de selección para elegir a los vectores sobrevivientes. De esta forma, las generaciones transcurren de forma iterativa hasta cumplir el criterio de paro. Así, el i -ésimo vector de la población en la generación G se denota como $\vec{X}_{i,G} = [x_{1,i,G}, x_{2,i,G}, \dots, x_{D,i,G}]$. A continuación se explica en más detalle cada componente de DE.

5.0.1. Inicialización

Usualmente, DE inicia el proceso de optimización con una población de NP vectores que son creados de forma aleatoria. Habitualmente los vectores de la población inicial son generados en base a una distribución uniforme, ya que usualmente no se posee información sobre cuáles son las zonas más promisorias del espacio de búsqueda. Por lo tanto, el j -ésimo componente del i -ésimo vector es inicializado de la forma $x_{j,i,0} = x_j^{(L)} + rand_{i,j}[0, 1](x_j^{(U)} - x_j^{(L)})$, donde $rand_{i,j}[0, 1]$ es un número aleatorio uniformemente distribuido entre 0 y 1.

5.0.2. Operador de mutación

Por cada vector objetivo se genera un vector mutado para lo cual se han propuesto múltiples estrategias con la particularidad de que en cierta forma se usen las diferencias entre vectores. La variante clásica de DE aplica la estrategia conocida como *rand/1*, en la cual se crea un vector mutado $V_{i,G}$ de la siguiente forma:

$$\vec{V}_{i,G} = \vec{X}_{r1,G} + F \times (\vec{X}_{r2,G} - \vec{X}_{r3,G}) \quad r1 \neq r2 \neq r3 \quad (40)$$

En la Ecuación (40) los índices $r1, r2, r3 \in [1, NP]$ deben ser enteros distintos y son generados de forma aleatoria en el rango $[1, NP]$. Además, estos índices son distintos al índice i . Es importante hacer notar que la diferencia entre los vectores es escalada por medio del parámetro F , el cual usualmente se define en el intervalo $[0,4, 1]$. Posteriormente, el vector de diferencia (escalado) es agregado a un tercer vector, lo que significa que los vectores mutados son similares a los vectores objetivo si el grado de diversidad es bajo, ya que los vectores de diferencias tienen norma pequeña. Como consecuencia de esto es claro que es crítico mantener un grado mínimo de diversidad en DE.

5.0.3. Operador de cruce

El operador de cruce se aplica con el objetivo de combinar la información de distintas soluciones candidatas y de incrementar la diversidad de los vectores. Específicamente, cada vector objetivo $\vec{X}_{i,G}$ se mezcla con su correspondiente vector mutado $\vec{V}_{i,G}$ para generar un vector de prueba $\vec{U}_{i,G} = [u_{1,i,G}, u_{2,i,G}, \dots, u_{D,i,G}]$. La estrategia de cruce más típica es conocida como *cruce binomial* y actúa de la siguiente forma:

$$\vec{U}_{j,i,G} = \begin{cases} \vec{V}_{j,i,G}, & \text{si } (rand_{i,j}[0, 1] \leq CR \text{ o } j = j_{rand}) \\ \vec{X}_{j,i,G}, & \text{de otra forma} \end{cases} \quad (41)$$

En la Ecuación (41) $rand_{i,j}[0, 1]$ es un número uniformemente distribuido, j_{rand} es un índice seleccionado aleatoriamente que asegura que $\vec{U}_{i,G}$ genera al menos un componente de $\vec{V}_{i,G}$ y $CR \in [0, 1]$ es la razón de cruce.

5.0.4. Operador de selección

Finalmente, se aplica una selección elitista para determinar los vectores sobrevivientes que participarán en la siguiente generación. Específicamente, cada vector de prueba se compara con su correspondiente vector objetivo y sobrevive el que tiene la mejor aptitud:

$$\vec{X}_{j,i,G+1} = \begin{cases} \vec{U}_{i,G}, & \text{si } f(\vec{U}_{i,G}) \leq f(\vec{X}_{i,G}) \\ \vec{X}_{i,G}, & \text{de otra forma} \end{cases} \quad (42)$$

Debido a la forma de seleccionar a los sobrevivientes en cada generación los miembros de la población permanecen iguales o mejoran. Por ello, se considera que DE hace uso de una selección muy elitista y es una de las razones por la que en ciertos casos puede sufrir de una convergencia demasiado acelerada.

5.1. Adaptación de Evolución Diferencial en un problema discreto

Particularmente, se desea utilizar DE en un problema de optimización discreto donde el conjunto factible está comprendido de únicamente números enteros y los límites inferior y superior son $x^{(L)} = 0$ y $x^{(U)} = 254$ respectivamente. Inicialmente, se podría considerar la variante estándar de Evolución Diferencial donde el resultado (contínuo) es redondeado a un entero. Sin embargo, esto puede crear problemas de convergencia, e inclusive –dependiendo del problema– podría generar platos *platus*, es decir, cambios en el espacio de las variables no tienen efectos en el espacio objetivo por lo tanto el proceso de búsqueda podría estancarse.

En la versión de Evolución Diferencial propuesta en este trabajo se utilizan una modificación tanto del operador de mutación como el operador de cruce. **La principal ventaja de esta propuesta es que no requiere tanto el parámetro de escala F como la probabilidad de mutación CR .**

5.2. Operador de mutación propuesto para este problema discreto

En las variantes más recientes de Evolución Diferencial se prefiere utilizar una distribución de probabilidad en $F = Normal(0,5,0,1)$, esto en base a que ED tiene un comportamiento de búsqueda más adecuado. En este caso, en lugar de utilizar un factor de escala se considera una distribución uniforme discreta en el intervalo $[1, \vec{X}_{r2,G} - \vec{X}_{r3,G}]$.

$$\vec{V}_{i,G} = \vec{X}_{r1,G} + signo(Diff) \times randint(1, abs(Diff)) \quad r1 \neq r2 \neq r3 \quad (43)$$

donde $Diff = \vec{X}_{r2,G} - \vec{X}_{r3,G}$. En este problema discreto, este procedimiento es válido a pesar de que el dominio es discreto, el espacio factible está comprendido por una especie de cono, es decir, todos los vecinos de un pixel (dentro del hyper-cubo del espacio factible) poseen un mapeo directo al espacio objetivo.

5.3. Operador de cruce propuesto para este problema discreto

En la literatura se argumenta que Evolución Diferencial resuelve un problema continuo discretizándolo. Una de las principales razones por las cuales se considera ED como un problema discreto, es que selecciona los componentes de un vector y el tamaño en base a una distribución binomial la cual modifica su comportamiento en base a la probabilidad de cruce. Entre mayor

sea la probabilidad de cruce, mas elementos serán modificados. Una forma para evitar que asignar este parámetro es mediante un esquema adaptativo como lo hacen los algoritmos Shade [9]. En este caso se opta por utilizar una probabilidad de cruce en base a una distribución uniforme $CR = U[0, 1]$, es decir, en algunos casos se cambian más elementos que en otros. Es importante notar que siempre se va a modificar al menos un valor en cada cruce, por lo tanto si $CR = 0$ no causara vectores iguales.

$$\vec{U}_{j,i,G} = \begin{cases} \vec{V}_{j,i,G}, & \text{si } (rand_{i,j}[0, 1] \leq CR \text{ o } j = j_{rand}) \\ \vec{X}_{j,i,G}, & \text{de otra forma} \end{cases} \quad (44)$$

6. Problema de Combinatoria para Umbralizar a varios niveles

Los métodos multi-nivel considerados en este trabajo fueron programados para calcular cualquier número de clases. La manera de realizar esto fue por medio de un método recursivo el cual resuelve un problema bien conocido en la literatura del ámbito de programación, es decir encontrar todas las combinaciones dado un conjunto factible. Para mayor información se puede visitar el libro Cormen et al. [10].

7. Validación Experimental

La parte experimental se enfoca en medir distintas propiedades de los siguientes variantes previamente explicadas:

- El método de Otsu.
- El método de Kapur.
- El método de difuso con entropía.
- Variante de Otsu optimizada con Evolución Diferencial (ED).
- Variante de Otsu optimizada con descenso de gradiente y diferencias finitas.

7.1. Eficiencia en relación de los métodos propuestos

Es importante notar que todos los métodos propuestos (incluyendo Otsu y Kapur) utilizan el mismo paradigma considerado en imágenes integrales [11]. Este principio consiste en pre-calcular una tabla acumulativa como se muestra en la siguiente tabla:

1	2	3	4	5	6	7
x.i	$\sum_{i=1}^2 x_i$	$\sum_{i=1}^3 x_i$	$\sum_{i=1}^4 x_i$	$\sum_{i=1}^5 x_i$	$\sum_{i=1}^6 x_i$	$\sum_{i=1}^7 x_i$

Así, si se desea la sumatoria entre x_4 y x_6 solo se tendría que calcular la diferencia el valor de la sumatoria que pertenece a la casilla 6 - el valor de la sumatoria que pertenece a la casilla 3, proporcionando consultas en tiempo constante $O(c)$ en lugar de tiempo lineal $O(n)$.

7.2. Umbralización de Imágenes considerando dos clases

En esta sección se presentan los resultados obtenidos de aplicar los métodos previamente explicados. Para llevar esto, se consideran las métricas explicadas en la sección 4, donde el proceso de medición consiste en contabilizar los valores TP , TN , FP y FN de todas las imágenes y tomar las métricas en base a estos valores. Esto sería equivalente a generar una imagen grande la cual está construida de todo el conjunto de imágenes de prueba. Además, se calcular la media obtenida con cada algoritmo, es decir, se consideran las métricas de forma individual en cada imagen. La media sirve para verificar si la calidad de un experimento (una imagen) es muy malo, entonces se podrían comparar los resultados al considerar los resultados de la imagen grande (consiste en juntar todas las imágenes) con la media. Este conjunto consta de 30 imágenes, las cuales aunque no cuentan con su *grand-truth*, sin embargo posee un las coordenadas de los objetos de interés. En base a esto, para obtener el *grand-truth* se utiliza un algoritmo recursivo (backtracking) el cual dada la coordenada de un pixel visita a los pixeles vecinos y si poseen la misma intensidad en escala de grises que el pixel principal son considerados como pixeles de interés. Para todos los experimentos, el algoritmo basado en evolución diferencial utiliza una población de tamaño 10 y un número de 50 generaciones. Principalmente, se puede observar que la función objetivo (enfoque de Otsu) no proporciona una fórmula directa para calcular el gradiente respecto a un umbral determinado. Por lo tanto para realizar el cálculo del gradiente se utilizan diferencias finitas centrales (se tuvo que probar con distintos tamaños de paso se que con $\alpha > 0,5$). Adicionalmente, se considera

la variante de Otsu proporcionada por la librería *skimage*. En la tabla 1 se presentan los resultados de los seis algoritmos, donde en las primeras cuatro columnas se pueden observar los resultados al considerar una imagen grande (generada por todas las imágenes) y en las últimas cuatro columnas se presenta se presenta la medie de considerar una métrica en cada imagen de forma individual. Se puede observar que la sensibilidad es mejor con el método de Kapur, la razón de esto es que el método de Kapur resalta una mayor región de interés. Particularmente, la sensibilidad únicamente mide el grado en que se obtuvo el objeto de interés, la desventaja de esta métrica es el caso extremo en que toda la imagen sea identificada como objeto de interés, en resultado se tiene una métrica no fiable. Este inconveniente, también surge con la especificidad. La variante de Otsu que considera el gradiente obtiene mejores resultados en la especificidad, exactitud e índice de Jaccard. Esto indica algo muy importante, es decir, a pesar de que la función a optimizar se basa la misma función objetivo que la del Otsu estándar esté algoritmo obtuvo mejores resultados. La principal razón de esto es que el método de gradiente no siempre llegaba al óptimo global, es decir, depende totalmente del tamaño de paso. Esto indica que la función objetivo representada por el método de Otsu puede tener problemas de sobre-ajuste. Además es bien conocido en la literatura que la estimación de la varianza tiene problemas de sobre-ajuste, en resultado se puede ver que aunque los optimizadores puedan alcanzar el óptimo la calidad del resultado dependerá totalmente de la función objetivo. Un punto interesante, es que al considerar el gradiente de la imagen el método de gradiente considera una versión simplificada del filtro de Sobel.

Por otra parte se observa que tanto el método de Otsu estándar como el método de Otsu con evolución diferencial proporcionan los mismos resultados en todos los indicadores, por lo tanto se puede establecer que la versión de evolución diferencial propuesto alcanza el óptimo en todas las imágenes.

En la imagen 2 se observa que el método del gradiente (fila 2 derecha) y el que considera la función difusa con entropía *fuzzy-entropy* (fila 3 izquierda) se aproximan más al objeto de interés. Por otra parte en la imagen 3 se observa que el método del gradiente (fila 2 derecha) proporciona los peores resultados. Sin embargo el algoritmo que proporciona la función difusa con entropía representa de una mejor forma al objeto de interés.

	Sensitivity	Specificity	Accuracy	Jaccard-Index	Mean-Sensitivity	Mean-Specificity	Mean-Accuracy	Mean-Jaccard-Index
Fuzzy-Entropy	0.9141733212	0.9706934158	0.9692271815	0.4352382467	0.9507412794	0.9708051953	0.9692271815	0.436446802
Otsu	0.9966280877	0.8716180531	0.8748670474	0.1714980051	0.9989194673	0.8730104869	0.8745987932	0.3286984941
Otsu-DE	0.9966280877	0.8716180531	0.8748670474	0.1714980051	0.9989194673	0.8730104869	0.8745987932	0.3286984941
Otsu-Gradient	0.9557967515	0.9736502483	0.9731870015	0.4805008691	0.8654891558	0.973310246	0.9731825563	0.4123111584
Otsu-skimage	0.9969660354	0.8576350713	0.8612584233	0.1574468336	0.9990275792	0.8591676934	0.8609691471	0.2899804435
Kapur	1	0.8362011413	0.8404741323	0.1405443855	1	0.837969414	0.8405240775	0.2177993445

Cuadro 1: Resultados al considerar únicamente dos clases.

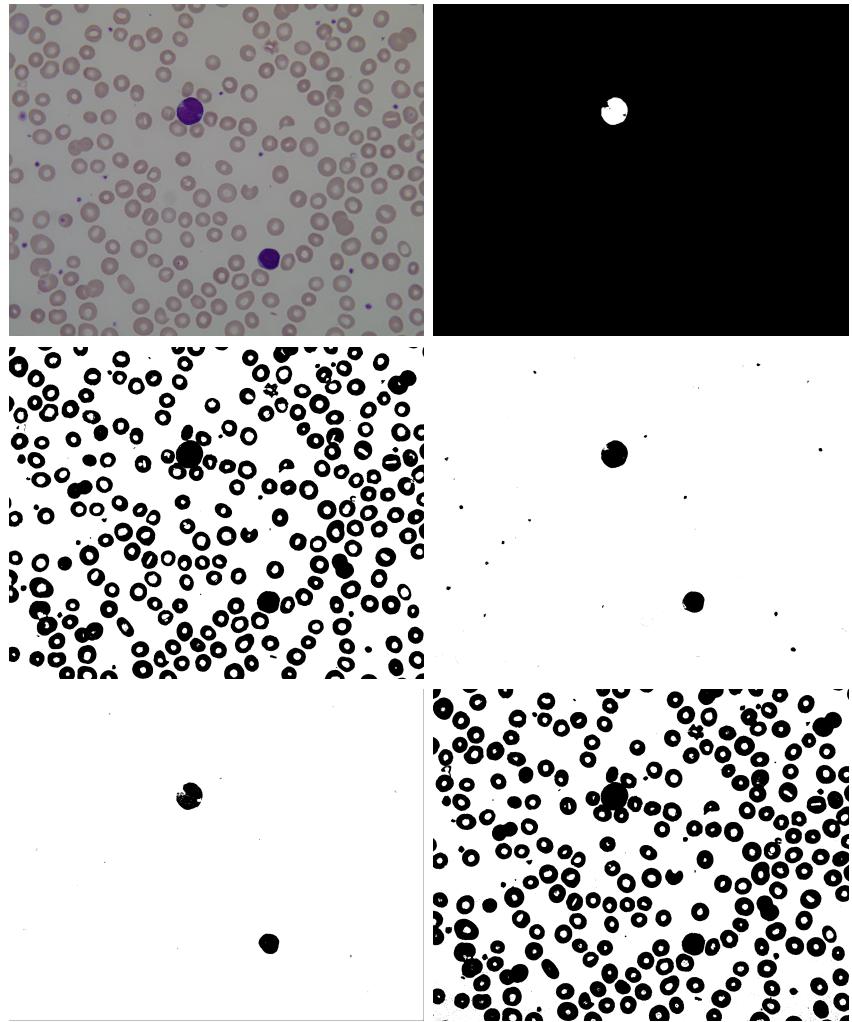


Figura 2: Análisis de la imagen *Im020_1.jpg*. En la primer fila parte izquierda está la imagen y en la parte derecha la región de interés. En la segunda fila en la parte izquierda está el método del Otsu estándar y en la parte derecha el método de Otsu con gradiente. En la tercera fila está el método difuso con entropía y en la parte deracha el método de Kapur. Se omite la versión DE ya que da los mismos resultados que la versión de Otsu estándar.

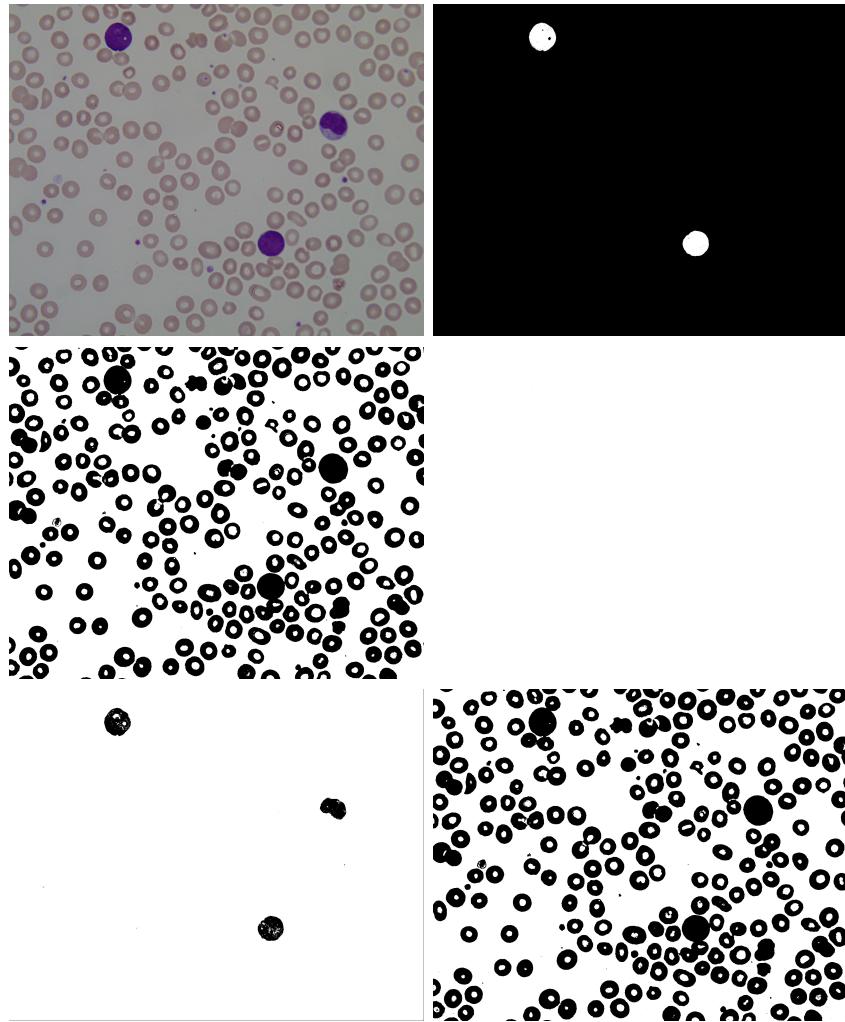


Figura 3: Análisis de la imagen *Im025_1.jpg*. En la primer fila parte izquierda está la imagen y en la parte derecha la región de interés. En la segunda fila en la parte izquierda está el método del Otsu estándar y en la parte derecha el método de Otsu con gradiente. En la tercera fila está el método difuso con entropía y en la parte deracha el método de Kapur. Se omite la versión DE ya que da los mismos resultados que la versión de Otsu estándar.

7.3. Umbralización Multi-nivel

En las imágenes 4 y 5 se presentan los resultados de segmentar las imágenes a tres niveles, principalmente se observa que el método del gradiente resalta tres objetos de una forma no conveniente, sin embargo el método

de Otsu estándar realiza una clasificación ideal, donde son encontrados los objetos de interés en una de las clases. Además se observa que Evolución Diferencial proporciona los mismos resultados de el método de Otsu estándar. Por otra parte el método de Kapur visualmente tiene problemas para clasificar los objetos en tres clases.

En figura 6 son graficadas la función objetivo del método de Kapur y del método de Otsu de las imágenes *Im020_1.jpg* y *Im025_1.jpg*. Principalmente, se puede observar que la función objetivo de Kapur es tiene forma casi concava ya que está basada en la entropía la cual posee únicamente un óptimo global. Por otra parte, se observa que la función objetivo de Otsu posee regiones excesivamente irregulares.

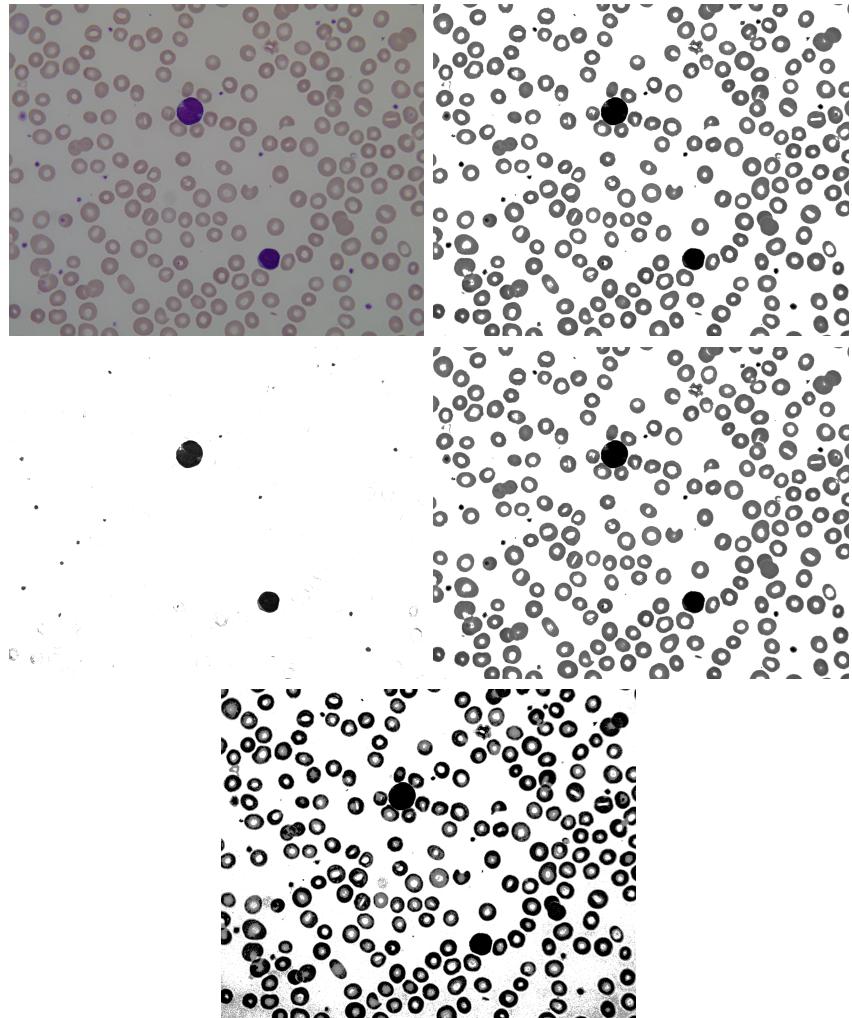


Figura 4: Análisis de la imagen *Img020_1.jpg* considerando tres niveles. En la primera fila en la parte izquierda está la imagen y en la parte derecha el resultado de segmentar con tres niveles. En la segunda fila en la parte izquierda está el método del gradiente y en la parte derecha el método de Otsu con Evolución Diferencial. En la última fila está el método de Kapur.

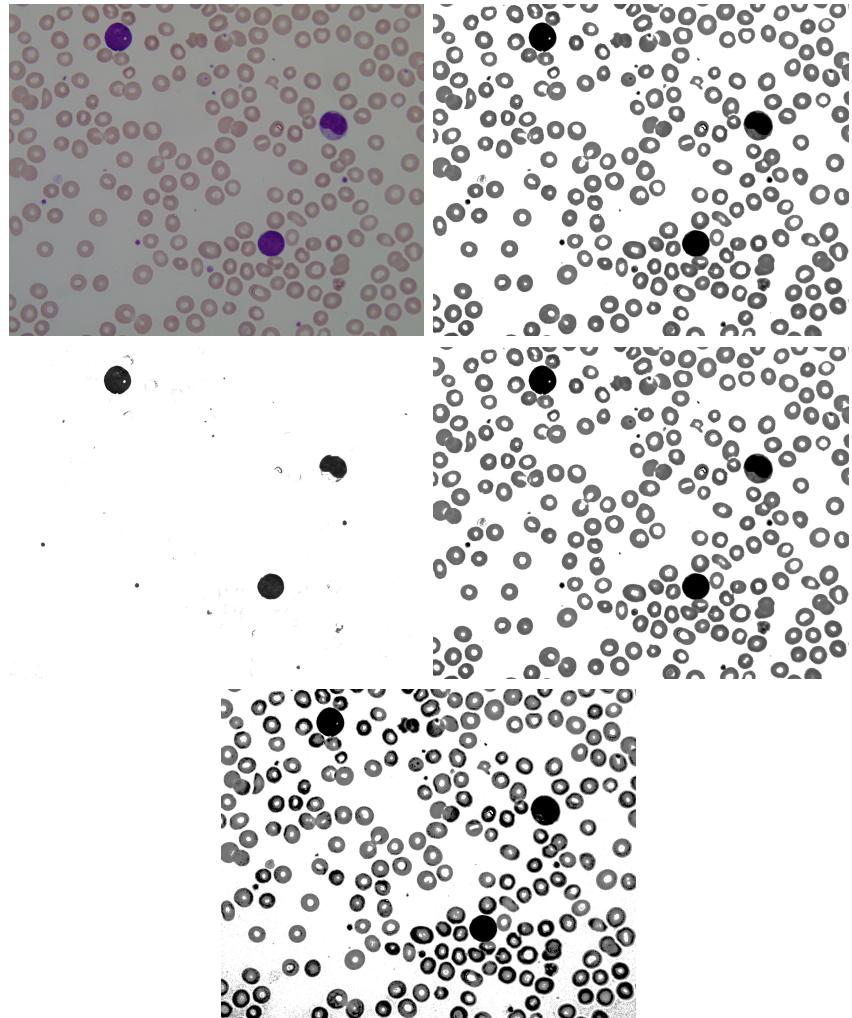


Figura 5: Análisis de la imagen *Img020_1.jpg* considerando tres niveles. En la primera fila en la parte izquierda está la imagen y en la parte derecha el resultado de segmentar con tres niveles. En la segunda fila en la parte izquierda está el método del gradiente y en la parte derecha el método de Otsu con Evolución Diferencial. En la última fila está el método de Kapur.

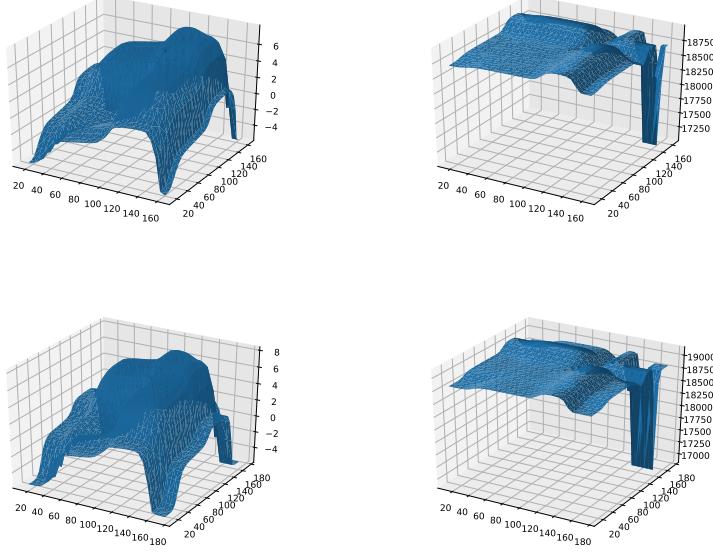


Figura 6: Función objetivo de la imagen *Im020_1.jpg* (primer fila) y *Im025_1.jpg* (segunda fila). En la parte izquierda de cada fila se ilustra la función objetivo de Kapur y en la parte derecha la función objetivo de Otsu.

8. conclusiones

En este trabajo se estudiaron cinco algoritmos para realizar umbralización de imágenes en escala de grises. Específicamente se puede concluir los siguiente.

- El modelo utilizado en el método de Otsu puede sobre-ajustar los parámetros y en resultado no proporcionar la mejor segmentación de la imagen.
- La propuesta de Evolución Diferencial basado en la función de Otsu obtiene los mismos resultados que el método de Otsu en umbralización de dos y tres niveles.
- El método de Otsu considerando más de cuatro niveles no es factible ya que implica un problema de combinatoria.
- El método del gradiente basado en la función de Otsu proporciona la mejor exactitud, esto se debe a que el optimizador no proporciona los

puntos óptimos y en resultado el model de Otsu no es sobre-ajustado, sin embargo en algunas imágenes el rendimiento de este método es muy malo.

- El modelo de Kapur posee propiedades deseables en el ámbito de optimización.

Basado en lo que se encontró, se hace énfasis que la parte de modelación del problema tiene un rol muy importante. Una propuesta sencilla consiste en un modelo el cual esté basado en la combinación lineal del método de Kapur y el método de Otsu, algo similar a lo aplicado en regresión con Lasso y Ridge (Elastic-Net).

- [1] M. Sezgin, B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, *Journal of Electronic imaging* 13 (2004) 146–166.
- [2] L.-K. Huang, M.-J. J. Wang, Image thresholding by minimizing the measures of fuzziness, *Pattern recognition* 28 (1995) 41–51.
- [3] N. Otsu, A threshold selection method from gray-level histograms, *IEEE transactions on systems, man, and cybernetics* 9 (1979) 62–66.
- [4] T. Pun, A new method for grey-level picture thresholding using the entropy of the histogram, *Signal processing* 2 (1980) 223–237.
- [5] J. N. Kapur, P. K. Sahoo, A. K. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, *Computer vision, graphics, and image processing* 29 (1985) 273–285.
- [6] A. Fenster, B. Chiu, Evaluation of segmentation algorithms for medical imaging, in: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, IEEE, pp. 7186–7189.
- [7] K. Price, R. M. Storn, J. A. Lampinen, Differential evolution: a practical approach to global optimization, Springer Science & Business Media, 2006.
- [8] D. M. Olsson, L. S. Nelson, The nelder-mead simplex procedure for function minimization, *Technometrics* 17 (1975) 45–51.

- 21 [9] R. Tanabe, A. Fukunaga, Success-history based parameter adaptation
22 for differential evolution, in: 2013 IEEE congress on evolutionary computa-
23 tion, IEEE, pp. 71–78.
- 24 [10] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, Introduction to
25 algorithms, MIT press, 2009.
- 26 [11] P. Viola, M. J. Jones, Robust real-time face detection, International
27 journal of computer vision 57 (2004) 137–154.