

Método No Paramétrico de Segmentación de Imágenes Basado en Teoría de Información

Felipe Córdova y Miguel Videla, *Ingeniería Civil Eléctrica, Universidad de Chile*

Abstract—En el presente documento se presenta un enfoque de segmentación de imágenes basado en métricas de teoría de información y evolución de curvas basado en [1], caracterizando el problema como la maximización de la información mutua entre las distintas regiones de la imagen y la intensidad de los píxeles, sin asumir a priori, la densidad de probabilidad de la intensidad de los píxeles de la imagen, utilizando una estimación no paramétrica de la misma para abordar el problema propuesto.

Index Terms—Segmentación de imágenes, curvas de evolución, conjuntos de nivel, teoría de información, estimación de densidad no paramétrica.

I. INTRODUCCIÓN

LA segmentación de imágenes [2] consiste en particionar una imagen digital en N segmentos de mediante un criterio de homogeneidad, tal como, color, brillo o textura, obteniendo los bordes que delimitan cada partición, para extraer información relevante de dicha imagen. Este proceso es fundamental en diversas áreas, tales como visión computacional, procesamiento médico de imágenes, reconocimiento facial, imágenes satelitales, etc.

En la actualidad existen múltiples técnicas de segmentación de imágenes [3], las cuales se dividen en dos grandes categorías, los métodos basados en discontinuidades y los basados en similitudes, los cuales se enfocan en detectar cambios abruptos o similitudes en la intensidad de escala de grises, respectivamente, identificando puntos de interés, líneas y bordes. Los principales enfoques de segmentación se presentan a continuación:

1. *Segmentación basada en detección de bordes*: La segmentación basada en detección de bordes consiste en identificar puntos de una imagen donde se encuentre un cambio abrupto en la intensidad de la misma, lo cual puede ser obtenido mediante la intersección de las curvas gaussianas referentes a los histogramas de escalas de grises del fondo y la zona de interés respectivamente, obteniendo un umbral de discernimiento, o bien, mediante al cálculo del gradiente de cada punto.
2. *Segmentación basada en umbral*: La segmentación basada en umbral consiste en separar la zona de interés

del fondo de acuerdo a un valor de umbral de intensidad determinado, donde dicho umbral puede ser global, si la intensidad de los objetos de interés y el fondo son muy distinto, o en caso contrario, local, seleccionando umbrales para distintas subsecciones de la imagen.

3. *Segmentación basada en regiones*: La segmentación basada en regiones consiste en separar la imagen en regiones de acuerdo a un conjunto de criterios de similitud predefinidos, dividiéndose en dos grandes métodos, el método región creciente, en el cual se seleccionan puntos iniciales en la imagen, los cuales anexarán iterativamente píxeles de su vecindad que cumplan con los criterios de similitud preestablecidos hasta que ningún pixel de la vecindad cumpla los requisitos para ser anexado, y método de separación y unión, en el cual, se subdivide la imagen en subregiones arbitrarias inconexas, las cuales se separarán o unirán mediante un criterio de segmentación preestablecido.
4. *Segmentación basada en teoría*: La segmentación basada en teoría agrupa técnicas de diferentes campos tales como algoritmos genéticos, difusos, redes neuronales, clustering, etc. Los algoritmos de clustering consisten en agrupar los puntos de la imagen en categorías finitas de acuerdo a un criterio de similitud pudiendo ser clustering duro, en el cual existen bordes definidos de pertenencia a un conjunto, o clustering difuso, donde cada punto tendrá un cierto grado de pertenencia no binario a los distintos conjuntos. Los algoritmos basados en redes neuronales, consisten en mapear cada pixel de la imagen a neuronas de entrada de la red, donde la red es entrenada con un conjunto de imágenes de entrenamiento para ajustar sus parámetros, luego, las imágenes son segmentadas mediante el modelo entrenado.
5. *Segmentación basada en modelos*: La segmentación basada en modelos, consiste en identificar distintos objetos en una imagen en base al conocimiento previo de la forma geométrica de dichos objetos, requiriendo el conocimiento cabal de la forma de todos los objetos presentes en la imagen.

En el presente documento se propone un método de segmentación de imágenes mediante curvas de evolución [1], basado en el método de conjuntos de nivel [4], el cual modificará una curva inicial para separar la imagen en distintos segmentos basado en métricas de teoría de información, utilizando una estimación no paramétrica de las densidades de probabilidad de las variables aleatorias a utilizar, donde se presentará la motivación y la demostración de la validez del problema de optimización a resolver para un caso binario de regiones de segmentación, posteriormente se explicarán los conceptos básicos del método de evolución de curvas mediante conjuntos de nivel, luego se presentarán las estimaciones de densidades de probabilidad de las variables aleatorias propuestas. Finalmente se implementará el método descrito, analizando la efectividad y eficiencia de la solución obtenida, y aplicando distintos métodos de aproximación de los cálculos requeridos con el objetivo de optimizar el costo computacional y tiempo requerido para la ejecución del mismo, comparando su desempeño y eficiencia con el método de contornos activos geodésicos [9].

II. FUNCIÓN DE COSTO PARA SEGMENTACIÓN DE IMÁGENES USANDO MÉTRICAS DE TEORÍA DE INFORMACIÓN

A. Descripción del Problema

Consideramos el problema de segmentación de imagen en 2 regiones, estas zonas poseen distintas densidades de probabilidad para los valores de intensidad que toma cada pixel perteneciente a las mismas ambos parámetros desconocidos e inherentes de cada imagen, a priori no se considera ninguna forma particular de estas densidades, por lo que el problema se presenta de manera generalizada. Se denota la intensidad de un pixel x como $G(x)$, luego se escribe la función de densidad de probabilidad $p_1 \forall x \in R_1$ y $p_2 \forall x \in R_2$ como sigue:

$$\begin{aligned} \{G(x) \mid x \in R_1\} &\sim p_1 \\ \{G(x) \mid x \in R_2\} &\sim p_2 \end{aligned} \quad (1)$$

Se considera que la intensidad de pixeles $G(x)$ para cada región R_i es *i. i. d.* Luego el problema queda definido por las regiones R_1 y R_2 que son desconocidas a las cuales se les asocia las densidades de probabilidad para las intensidades de pixel p_1 y p_2 , también desconocidas.

El problema consiste en encontrar una curva \vec{C} que coincida con el borde entre las regiones R_1 y R_2 , por lo que se necesita encontrar una acción de evolución de la curva \vec{C} para que se aproxime de mejor manera al objetivo. Denotamos como R_- la región que se encuentra fuera de la curva cerrada \vec{C} y R_+ la zona interior, luego asociamos una etiqueta binaria $L_{\vec{C}}$ que mapea el dominio de la imagen Ω a un espacio binario $\{L_+, L_-\}$, la definición de la función corresponde a:

$$L_{\vec{C}}(x) = \begin{cases} L_+, & \text{si } x \in R_+ \\ L_-, & \text{si } x \in R_- \end{cases} \quad (2)$$

Por lo que el problema de segmentación de imágenes a partir

de una curva \vec{C} es equivalente a resolver un problema de etiquetación binario dependiente de la curva \vec{C} en donde se busca que se cumpla lo siguiente:

$$L_{\vec{C}}(x_1) = L_{\vec{C}}(x_2) \quad \forall x_1, x_2 \in R_i \quad i \in [1, 2] \quad (3)$$

B. Información Mutua entre Intensidad de Pixel $G(x)$ y Etiqueta del Pixel $L_{\vec{C}}(x)$

Para la solución del el problema es necesario introducir el concepto de información mutua entre la intensidad de pixel y la etiqueta asignada por la curva cerrada \vec{C} . Para propósitos de esta sección consideramos las densidades de probabilidad p_1 y p_2 conocidas, luego el candidato a solución \vec{C} es aquel que permite dividir la imagen en zonas las cuales concuerdan con R_1 y R_2 , ahora se toma un pixel $x \in \Omega$ de manera aleatoria, a este muestreo se le asocia la variable aleatoria X , la cual es uniformemente distribuida sobre Ω , por lo que ahora la variable aleatoria $L_{\vec{C}}(X)$ que toma valor L_+ con probabilidad $\frac{|R_+|}{|\Omega|}$ y L_- con probabilidad $\frac{|R_-|}{|\Omega|}$. Con lo mencionado anteriormente es posible denotar la densidad de probabilidad de la variable aleatoria $G(X)$ como sigue:

$$p_{G(X)}(\omega) = P(X \in R_1) p_1(\omega) + P(X \in R_2) p_2(\omega) \quad (4)$$

$$= \frac{|R_1|}{|\Omega|} p_1(\omega) + \frac{|R_2|}{|\Omega|} p_2(\omega) \quad (5)$$

Donde ω es el argumento de la función de densidad de probabilidad sobre el conjunto dominio Ω .

Es posible notar que la intensidad de pixel está directamente conectado con la pertenencia de x sobre los conjuntos R_1 y R_2 , a partir de este factor nace la intuición de que mientras mejor sea la división de C sobre el dominio de la imagen, en otras palabras que se cumpla (3), menor es la incerteza o *Entropía* que se tiene de $G(X)$, es por esto que se procede a caracterizar la información mutua entre $G(X)$ y $L_{\vec{C}}(X)$ como sigue:

$$\begin{aligned} I(G(X); L_{\vec{C}}(X)) &= H(G(X)) - H(G(X) \mid L_{\vec{C}}(X)) \end{aligned} \quad (6)$$

$$\begin{aligned} &= H(G(X)) - P(L_{\vec{C}}(X) = L_+) H(G(X) \mid L_{\vec{C}}(X) = L_+) \\ &\quad - P(L_{\vec{C}}(X) = L_-) H(G(X) \mid L_{\vec{C}}(X) = L_-) \end{aligned} \quad (7)$$

Donde $H(Y)$ es la entropía, o grado de incerteza, de la variable aleatoria Y , que queda definida como $-\sum_Y p_Y(y) \log(p_Y(y))$. Como las variables aleatorias en (7) se encuentran en su forma condicional, es necesario escribir de manera explícita su forma, de esta manera se obtiene como:

$$p_{G(X) \mid L_{\vec{C}}(X) = L_r}(\omega) = \frac{|R_1 \cap R_r|}{|R_r|} p_1(\omega) + \frac{|R_2 \cap R_r|}{|R_r|} p_2(\omega) \quad (8)$$

Con r etiquetas en $[-, +]$.

Con lo mostrado en (5) es posible intuir que el valor que

maximiza $I(G(x); L_{\vec{C}}(X))$ se obtiene cuando la curva cerrada \vec{C} aproxima de mejor manera las divisiones intrínsecas de la imagen en el sentido de (3), puesto que las entropías en (7) alcanzan su mínimo valor mientras menos incertidumbre posea su argumento, dado el hecho que $G(X)$ está relacionado con la etiqueta $L_{\vec{C}}$ en el sentido de (8) se observa que la incertidumbre de la densidad de probabilidad disminuye mientras mejor que la asignación de etiquetas. De hecho, la intuición mencionada anteriormente se verifica de manera formal en el *Apéndice A*, obteniéndose que el máximo valor de (7) se alcanza cuando se cumple (3), por lo que maximizar este criterio es objetivo del problema.

Es importante mencionar que para los efectos de esta sección se están considerando las densidades p_1 y p_2 como conocidas, en contra posición a la definición del problema en donde estas no lo son, es por ello que es necesario realizar una estimación de las densidades de probabilidad obteniéndose una estimación tanto para la información mutua $\hat{I}(\cdot)$ como para la entropía $\hat{H}(\cdot)$ denotadas de esa forma, luego la función a maximizar es equivalente a (7) reemplazando los valores de $I(\cdot)$, $H(\cdot)$, $p_i(\cdot)$ por sus valores estimados, el cálculo de estos queda explicado en la *Sección IV*.

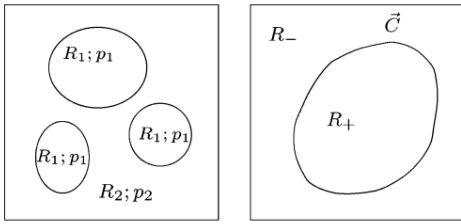


Fig. 1. A la izquierda: Ilustración de región de interés (R_1) y fondo (R_2), asociadas a las distribuciones de probabilidad p_1 y p_2 , respectivamente. A la derecha: Ilustración de la curva \vec{C} y de regiones interna (R_+) y externa (R_-) de la misma.

C. Función de Costo y Objetivo del Problema

A partir de lo concluido en la *subsección B* es necesario maximizar el valor de $\hat{I}(G(x); L_{\vec{C}}(X))$, que es equivalente a encontrar la curva cerrada \vec{C} que resuelve el problema de segmentación. Por otro lado, es necesario establecer algunas restricciones sobre la curva para que sea optima en un sentido de implementación, es por esto que se debe agregar a la función costo a definir una regularización o penalización sobre el largo de la curva que evita soluciones puntiagudas (jagged curves) que no representan una segmentación del todo óptima.

Considerando los puntos mencionados anteriormente se procede a construir la función de costo $E(\vec{C})$ como sigue:

$$E(C) = -|\Omega| \hat{I}(G(x); L_{\vec{C}}(X)) + \alpha \oint_{\vec{C}} ds \quad (9)$$

Donde $\oint_{\vec{C}} ds$ corresponde al largo de la curva \vec{C} y α corresponde a la penalización asociada al largo de la curva, por otro lado, es posible observar que la información mutua estimada $\hat{I}(G(x); L_{\vec{C}}(X))$ esta ponderada por el tamaño del

dominio de la imagen, esto se realiza para representar numéricamente la información mutua que posee toda la imagen.

Luego el problema de segmentación es análogo a la minimización del funcional $E(\vec{C})$.

III. CONJUNTOS DE NIVEL

Se propone representar la evolución de una determinada curva $\vec{C}(t)$, definiendo una superficie ϕ con una dimensión adicional llamada función de conjunto de nivel [4]. La función de conjunto de nivel recibe los puntos de entrada $\vec{x}(t) = (x, y)$, asignándoles una altura o distancia z a la interfaz en un determinado tiempo:

$$z = \phi(\vec{x}(t), t) \quad (10)$$

Donde se establece $\phi(\vec{x}(t), t) > 0$ para los puntos exteriores a la curva \vec{C} y $\phi(\vec{x}(t), t) < 0$ para los puntos interiores a la misma. De este modo, la curva $\vec{C}(t)$ queda determinada por:

$$\vec{C}(t) = \{x | \phi(\vec{x}(t), t) = 0\} \quad (11)$$

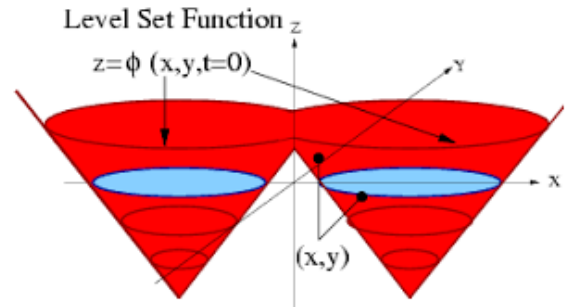


Fig. 2. En rojo: Función de conjunto de nivel ϕ . En azul: curva C obtenida en $t = 0$.

Para obtener la función de conjunto de nivel $\phi(\vec{x}(t), t)$, notamos que dado $\phi(\vec{x}(t), t = 0)$, es posible determinar ϕ para cualquier t mediante la ecuación de movimiento $\frac{\partial \phi}{\partial t}$, la cual por regla de la cadena resulta:

$$\frac{\partial \phi(\vec{x}(t), t)}{\partial t} = 0 \quad (12)$$

$$\phi_t + \nabla \phi x_t = 0 \quad (13)$$

Se observa que la velocidad o flujo de gradiente x_t puede ser obtenida mediante una fuerza F_n normal a la superficie de la curva, así $x_t = F_n(\vec{x}(t)) \cdot \vec{N}$, donde $\vec{N} = \frac{\nabla \phi}{|\nabla \phi|}$. De este modo, la ecuación de movimiento anterior puede ser reescrita de la siguiente forma:

$$\phi_t = -F_n |\nabla \phi| \quad (14)$$

Donde la última ecuación determina el movimiento de ϕ , es decir, dado ϕ en un tiempo $t = 0$, y su movimiento a través del tiempo, es posible determinar $\phi(\vec{x}(t), t)$ en cualquier tiempo t

mediante la evolución de $\phi(\vec{x}(t), t = 0)$, donde dicha curva inicial puede separarse o unirse en distintas otras curvas a lo largo del tiempo, siendo fácilmente determinada por la formulación propuesta.

Notamos que la velocidad o flujo de gradiente x_t se define sólo en el contorno de la curva, por lo cual la velocidad fuera de la curva queda a priori indeterminado, siendo necesario determinar la velocidad en todo el campo para establecer la correcta evolución de la función de conjunto de nivel. Para ello se mantiene constante la velocidad de cada punto de la curva en su dirección normal, estableciendo así, la velocidad de todo el campo.

Es posible obtener la curvatura κ de la superficie, mediante la siguiente ecuación:

$$\kappa = \nabla \frac{\nabla \phi}{|\nabla \phi|} \quad (15)$$

$$\kappa = \frac{\phi_{xx}\phi_y^2 - 2\phi_{xy}\phi_x\phi_y + \phi_{yy}\phi_x^2}{(\phi_x^2 + \phi_y^2)^{3/2}} \quad (16)$$

Para ejecutar el método de conjuntos de nivel de manera rápida, se propone utilizar la variante de banda estrecha [5], construyendo un tubo sobre el cual la función ϕ estará definida.

El tubo se compone por los puntos del espacio cuya distancia sea menor a una distancia máxima d a la curva construyendo una grilla de operación de k puntos, actualizando los valores de la función de conjunto de nivel sólo en los puntos de la vecindad $\{x; |\phi(x)| < d\}$, ahorrando el tiempo de computo del cálculo sobre todos los puntos del espacio de orden $O(n^3)$ a $O(nk^2)$. Se debe asegurar que el nivel cero correspondiente a la curva en evolución siga en el interior del tubo. Para ello se utiliza el tubo definido durante la mayor cantidad de iteraciones posibles, recalculándolo cuando la curva generada en alguna iteración este cerca del borde de este.

IV. ESTIMACIÓN DE NO PARAMÉTRICA DENSIDADES DE PROBABILIDAD

En esta sección se presentan los métodos asociado a una estimación no paramétrica de la densidad de probabilidades de las regiones denotadas por la curva \vec{C} y la evolución de \vec{C} a modo de alcanzar el óptimo que minimice el funcional $E(\vec{C})$.

A. Estimación del Diferencial de Entropía

Primeramente, se realiza una estimación no paramétrica de las entropías diferenciales que describen a la información mutua estimada en (16) utilizando el siguiente estimador:

$$H_n = -\frac{1}{n} \sum_{i=1}^n \log f_n(X_i) \quad (17)$$

Donde $f_n(X_i)$ es una estimación por el metodo de Kernel de la densidad de probabilidades asociada al vector aleatorio X . La estimación propuesta en (17) cumple con la propiedad de consistencia cuadrática media (18) como sus autores Ahmad y Lin muestran en [6]:

$$\lim_{n \rightarrow \infty} E\{(H_n - H(x))^2\} = 0 \quad (18)$$

Como se mencionó anteriormente la entropía $\hat{H}(G(X))$ no es dependiente de la curva \vec{C} , luego se pueden escribir las entropías condicionales como sigue.

$$\begin{aligned} \hat{H}(G(X)|L_{\vec{C}}(X) = L_+) &= -\frac{1}{|R_+|} \int_{R_+} \log \hat{p}_+(G(x)) dx \\ &= -\frac{1}{|R_+|} \int_{R_+} \log \left(\frac{1}{|R_+|} \int_{R_+} K(G(x) - G(\hat{x})) d\hat{x} \right) dx \end{aligned} \quad (19)$$

Donde \hat{p}_+ es el estimador de densidad de probabilidad, en (18) se aproxima el valor estimado como el promedio de \hat{p}_+ sobre todas las muestras, este a su vez se trata de un estimador del tipo KDE, que utiliza la función $K(z) = (1/\sqrt{2\pi\sigma})e^{-z^2/2\sigma^2}$ considerando σ un parámetro de $K(z)$.

Análogamente se escribe $\hat{H}(G(X)|L_{\vec{C}}(X) = L_-)$ como:

$$\begin{aligned} \hat{H}(G(X)|L_{\vec{C}}(X) = L_-) &= -\frac{1}{|R_-|} \int_{R_-} \log \left(\frac{1}{|R_-|} \int_{R_-} K(G(x) - G(\hat{x})) d\hat{x} \right) dx \end{aligned} \quad (20)$$

B. Flujo del Gradiente para Integrales Anidadas

Es posible observar que (19) y (20) son integrales anidadas, por lo que se pueden escribir de la manera:

$$\int_{R_+} f(\epsilon(x, t)) dx, \text{ considerando } \epsilon(x, t) = \int_{R_+} g(x, \hat{x}) d\hat{x} \quad (21)$$

Al escribir las integrales anidadas en la forma general (21) es posible derivar el flujo del gradiente para la curva \vec{C} como aparece en (22), de manera negativa con el objetivo de tener un mayor decrecimiento en la región integral.

$$\frac{d\vec{C}}{dt} = - \left[f(\epsilon(\vec{C})) + \int_{R_+} f'(\epsilon(x)) g(x, \vec{C}) dx \right] \vec{N} \quad (22)$$

En donde \vec{C} depende del tiempo t y luego R_+ también depende del tiempo t . La obtención de (22) se detalla formalmente en el Apéndice B.

Puesto que se tiene las expresiones de información mutua como integrales anidadas y utilizando la expresión obtenida en (22) es posible calcular el flujo de gradiente para el funcional energía $E(\vec{C})$ como se detalla en el Apéndice C el cual presenta como resultado:

$$\frac{d\vec{C}}{dt} = \left[\log \frac{\hat{p}_+(G(\vec{C}))}{\hat{p}_-(G(\vec{C}))} + \frac{1}{|R_+|} \int_{R_+} \frac{K(G(x) - G(\vec{C}))}{\hat{p}_+(G(x))} dx - \frac{1}{|R_-|} \int_{R_-} \frac{K(G(x) - G(\vec{C}))}{\hat{p}_-(G(x))} dx \right] \vec{N} - \alpha \kappa \vec{N} \quad (23)$$

En donde κ corresponde a la curvatura de \vec{C} y por lo tanto el término $-\alpha \kappa \vec{N}$ es la penalización para el largo de la curva, similar a lo formulado para (17). La evolución de la curva por flujo de gradiente se implementa con el método de conjuntos de nivel mencionado en la Sección III.

Un análisis por componentes para la ecuación (23) identifica que el primer factor corresponde a la razón de verosimilitud de la intensidad de pixel para un punto presente en la curva con respecto a las regiones R_+ y R_- , este elemento desplaza al punto a \vec{C} hacia la zona en donde la intensidad de pixeles presente ser más homogénea, esto se debe a que se estima la probabilidad de pertenencia como el promedio de comparaciones de kernel del pixel con respecto a todos los elementos de la región como es posible observar en (19) y (20).

El segundo y tercer factor de (23) resultan ser elementos que refuerzan el accionar del primer componente, es importante notar que el primer componente no considera que su acción de movimiento induce una actualización en los pixeles pertenecientes en ambas regiones, estos factores calculan por sí solos el promedio del aumento en la probabilidad de pertenencia por pixel de cada elemento perteneciente a una región con respecto a los puntos que forman la curva \vec{C} . De esta manera se identifica que el segundo y tercer elemento de la ecuación de movimiento para \vec{C} considera el impacto de desplazar la curva en una u otra región.

C. Costo Computacional del Flujo del Gradiente

Al momento de implementar el cálculo del gradiente obtenido en la sección anterior es necesario estimar el costo computacional asociado al mismo, donde $p_+(G(x))$ toma costo $\mathcal{O}(|R_+|)$ cuando se realiza la comparación del elemento x en relación a todos los puntos pertenecientes a la región R_+ por lo que al evaluar $p_+(\vec{C})$ tiene costo $\mathcal{O}(M|R_+|)$ donde M es el número de puntos pertenecientes a la curva \vec{C} , de la misma manera se estima $p_-(\vec{C})$ como $\mathcal{O}(M|R_-|)$ obteniéndose un cálculo total del primer componente como $\mathcal{O}(M|R_+ + R_-|)$. En relación al segundo componente es necesario calcular el valor de $p_+(G(x))$ para cada elemento perteneciente a R_+ al cual se le asocia un costo de $\mathcal{O}(|R_+|^2)$, los valores calculados se almacenan y son utilizados para el cálculo de la integral que recorre cada elemento de la región R_+ la cual se recalcula para cada punto perteneciente a la curva \vec{C} luego se asocia el costo de $\mathcal{O}(M|R_+|)$, se realiza el mismo análisis para la región R_- . El costo total del cálculo del gradiente para cada iteración del algoritmo resulta ser $\mathcal{O}(M|R_+ + R_-| + |R_+|^2 + M|R_+| + |R_-|^2 + M|R_-|) \approx \mathcal{O}(\# \text{numero pixeles}^2)$.

V. IMPLEMENTACIÓN

Se procede a realizar la implementación de la evolución de la curva de segmentación utilizando el método de nivel descrito en la Sección III añadiendo algunas modificaciones las cuales se incorporaron como sigue. Dado que se obtiene por medio de métricas de teoría de la información el gradiente de la curva como en una ecuación cerrada correspondiente a (23) de la forma:

$$\frac{d\vec{C}}{dt} = \beta \vec{N} - \alpha \kappa \vec{N} = (\beta - \alpha \kappa) \vec{N} \quad (24)$$

Donde β corresponde al componente de la velocidad que impulsa a la curva hacia la segmentación óptima y el componente $\alpha \kappa$ que simboliza la penalización del largo de curva, que pretende suavizar irregularidades en la curva, es importante mencionar que κ corresponde a la curvatura en 2 dimensiones de la curva y no la curvatura en 3 dimensiones sobre la superficie \emptyset .

Con la evolución de curva descrita en forma de (24) se procede a construir una superficie \emptyset que toma valores para cada posición de pixel de la imagen, específicamente se considera una superficie que toma nivel cero para los puntos pertenecientes a la curva, (11) expresa este concepto, para los puntos en la región exterior se fija el valor 1 y para la región interior -1. Luego el problema de evolución de la curva se transforma a la evolución de la superficie \emptyset por lo que es necesario representar una magnitud de evolución para cada punto en el espacio de pixeles de la imagen, es posible notar que (24) solo determina la magnitud de movimiento en dirección normal para puntos pertenecientes, si se tratase de evolucionar la superficie \emptyset solo actualizando los valores en la curva el comportamiento completo de la superficie sería irregular y crearía problemas en la evolución de la curva, como se propone en [5] es necesario fijar valores distintos de 0 para las regiones superiores e inferiores en nivel a la curva, es por esto y motivados de que la evolución de la curva debe ser normal, en el sentido 2 dimensional de la misma y no normal a la superficie \emptyset , se procede a realizar una extensión de las velocidades en dirección normal a la curva implementado como se observa en la figura 3.

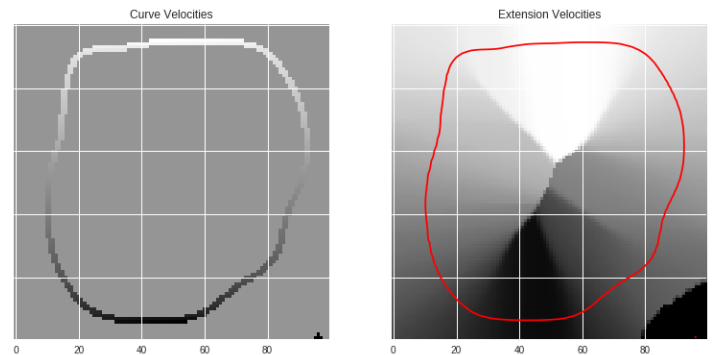


Fig. 3. Implementación y visualización de extensión de velocidades en dirección normal a la curva

Luego de definir como se actualizarán los valores de velocidad sobre el espacio de píxeles es necesario calcular la magnitud de las velocidades que serán propagadas según la ecuación (23) por lo que es necesario implementar la estimación de probabilidades p_+ y p_- como sigue:

$$p_*(x) = \frac{1}{|R_*|} \int_{R_*} K(x - \hat{x}) d\hat{x} \quad (25)$$

siendo $K(z) = (1/\sqrt{2\pi\sigma})e^{-z^2/2\sigma^2}$, siguiendo el análisis realizado en la Sección IV-C el costo computacional de realizar la estimación de la pertenencia de un punto sobre una región en particular resulta ser altísimo cuando las dimensiones de la imagen aumentan, es por esto que se procede a implementar el método de banda estrecha el cual puede ser resumido en considerar solo un vecindario d distante desde el conjunto de nivel 0, este método tiene la hipótesis que el vecindario cercano a la curva es lo suficientemente representativo como para determinar su siguiente paso en el flujo de gradiente, en la figura 4 se visualiza la banda estrecha para una determinada curva de nivel.

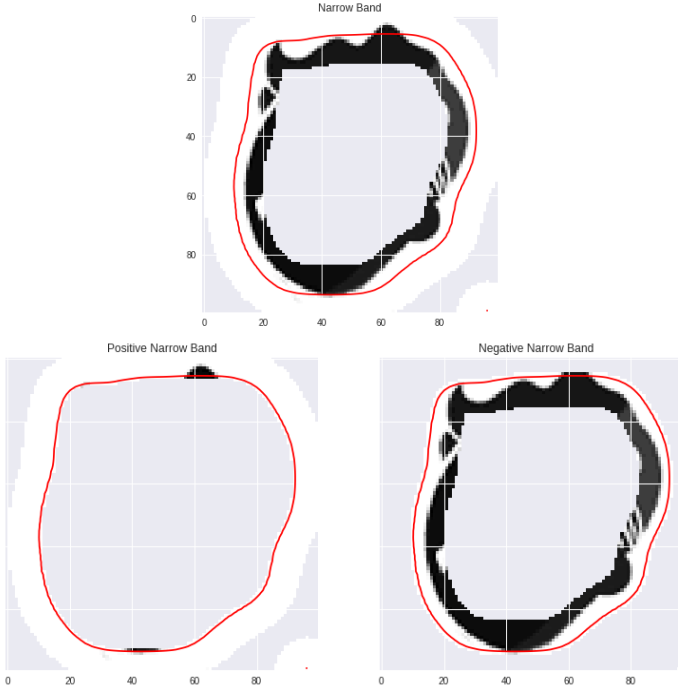


Fig. 4. Ejemplo de banda estrecha, sección interior y sección exterior de la misma.

Dado que el dominio R_+ y R_- se ve temporalmente disminuido dependiendo de cada iteración del algoritmo de optimización, el costo computacional disminuye considerablemente, siendo ahora mayor parte dependiente del largo de la curva de nivel. En cuanto a la longitud de la curva de nivel, es necesario considerar los caminos necesarios para representar la curva en un espacio discretizado como lo son los píxeles, es por esto que se realiza una aproxima una recta para cada “vértice” obtenido al determinar la pertenencia en cuadrícula de la curva de nivel 0 en el espacio de píxeles de la imagen. Finalmente, la longitud y complejidad de la curva

determinaran el costo computacional y por ello el tiempo de cálculo asociado a cada iteración.

Dado que ahora se considera una banda estrecha de actualización, es importante considerar que la evolución de la superficie \emptyset solo debe estar limitada en este ancho de banda puesto que ir más allá de los límites de este sería contraproducente con la hipótesis de que los puntos cercanos al nivel 0 son lo suficientemente representativos como para actualizar la curva en una evolución de magnitud reducida. Es por esto que se define \emptyset_t como la evolución de la superficie \emptyset en un tiempo Δt y esta solo toma valores distintos de 0 en la banda estrecha definida anteriormente.

Finalmente es necesario regularizar el movimiento de la curva dentro su banda estrecha, puesto que es posible que la curva evolucione violentamente en ciertos tramos alcanzando el límite de ancho de banda muy rápidamente, es por esto que se postula una normalización de las velocidades por punto en la curva como se menciona en [10], la cual se expresa de la siguiente manera:

$$\vec{V}_{norm} = \frac{\vec{V}}{2 \max(\vec{V})} \quad (26)$$

En donde \vec{V} representa el vector de velocidades obtenido en la ecuación (23). Lo que postula la (26) es fijar como máximo desplazamiento un movimiento, que en cierto sentido, mantendrá dentro de los márgenes del ancho de banda a la curva

Luego de calcular todos los parámetros necesarios para la actualización de la curva \vec{C} se procede a evolucionar la superficie \emptyset como:

$$\emptyset(t+1) = \emptyset(t) + \emptyset_t \quad (27)$$

El código desarrollado correspondiente a la implementación de los métodos de evolución de curva mediante conjuntos de nivel se encuentra disponible en [11], tanto para el método basado en teoría de información desarrollado en el presente documento, como el método de contornos activos geodésicos [9] para el desarrollo de comparativas de rendimiento y posterior análisis.

VI. RESULTADOS Y DISCUSIÓN

El algoritmo de segmentación de imágenes implementado se probó sobre distintas imágenes de diferentes dimensiones, donde en todas ellas era distinguible los objetos de interés del fondo, comparando los resultados obtenidos con un método similar de segmentación basado en contornos activos geodésicos [9]. Los resultados de segmentación de una imagen de 100x100 píxeles obtenidos por el método basado en teoría de información propuesto se presentan en la figura 5, mientras que los resultados obtenidos por el método de contornos activos geodésicos se presentan en la figura 6.

En la figura 5 se aprecia una curva rectangular inicial cercana los bordes de la imagen, la cual, se va deformando para adaptarse a los bordes superior e inferior de la figura de interés, estrechándose en las zonas laterales para lograr segmentar por

completo la figura de interés del fondo. Esta evolución de la curva genera dos largas y delgadas curvas residuales en los bordes laterales de la imagen, los cuales reducen su área hasta ser casi imperceptibles con el paso de las iteraciones. El método se iteró 50 veces tomando un tiempo total aproximado de 2 horas.

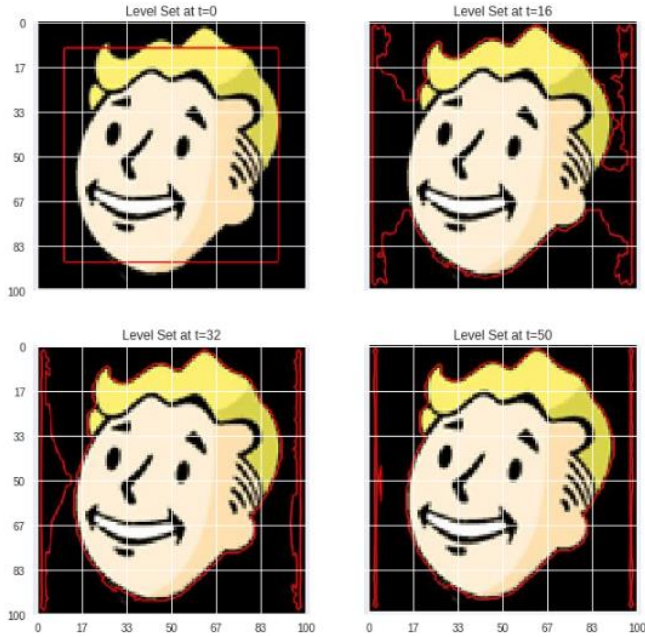


Fig. 5. Ejemplo de evolución de curva de segmentación utilizando método de contornos activos en base a teoría de información.

En la figura 6 se aprecia una curva inicial idéntica a la descrita en el caso anterior, la cual se contrae de manera cuasi-homogénea hasta acercarse lo más posible al contorno de la figura de interés. Como se observa en la figura 6, la curva obtenida no logra ajustarse de manera precisa a la figura de interés, debido a que los parámetros la función de detención $g(I)$ encargada de detener la evolución de la curva, no distingue con claridad la figura de interés, la cual se determina mediante los parámetros α para ponderar el valor del inverso del gradiente gaussiano de la imagen y σ correspondiente a la desviación estándar del filtro gaussiano utilizado para calcular el gradiente de la imagen descritos en [9], donde se utilizaron valores 1000 y 5, respectivamente. El método se iteró 200 veces tomando un tiempo total aproximado de 30 segundos.

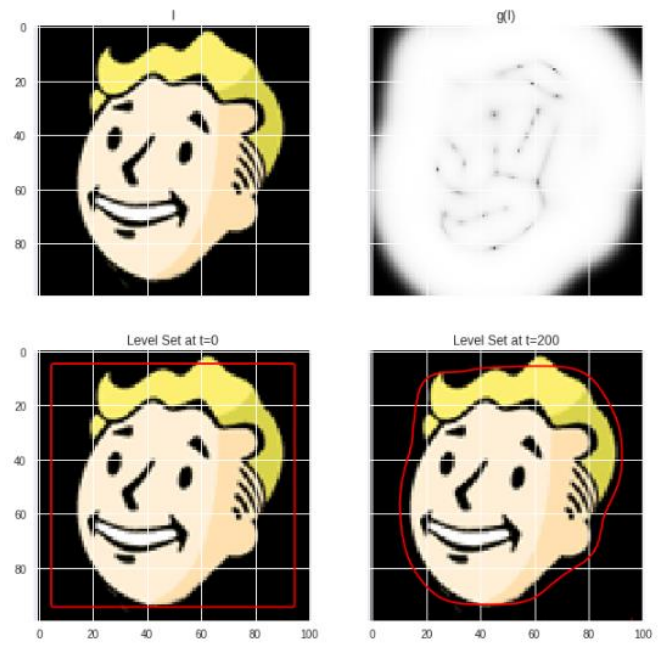


Fig. 6. Ejemplo de evolución de curva de segmentación utilizando método de contornos activos geodésicos junto a su respectiva función de detención $g(I)$.

Se observa que los tiempos de ejecución de los métodos presentados distan radicalmente, siendo el método de segmentación por medio de contornos activos mediante teoría de información el más costoso por lo expuesto en la sección IV-C. Para disminuir el costo computacional del algoritmo mencionado, se propone en [1] calcular una aproximación utilizando el método de FGT (Fast Gauss Transform) [7], el cual permite estimar $p_+(G(x))$ con sólo N puntos pertenecientes a la banda estrecha referida en [5]. Si se realiza sobre todos los puntos de M , con M el número de puntos pertenecientes a la curva \tilde{C} , el algoritmo entrega un costo de $\mathcal{O}(c(M + N))$ donde c es un factor que crece en relación a la exactitud la cual se requiera sobre el algoritmo. Formalmente corresponde al orden de la aproximación de Taylor utilizada sobre el FGT en donde valores inferiores a $c = 10$ son lo suficientemente cercanos para la estimación de la distribución de probabilidades. Luego es necesario solo tomar un submuestreo de N puntos de cada región para obtener un costo total de $\mathcal{O}(c(M + N + N) + c(N + N) + c(M + N) + c(N + N) + c(M + M))$, obteniéndose que una razonable elección es un múltiplo del número de puntos M considerados en la curva, por lo que el orden del costo computacional se considera $\mathcal{O}(M)$ mucho menor a $\mathcal{O}(\text{numero de pixeles}^2)$, pudiendo así, comparar efectivamente los tiempos de convergencia de ambos métodos presentados. Esta aproximación no fue implementada en el trabajo desarrollado debido a su alta complejidad, considerando que la ausencia de la misma no presentaba efecto alguno sobre el desempeño del algoritmo presentado.

Sin embargo, si es posible comparar el desempeño de los métodos presentados, donde se observa un ajuste más preciso de la curva de segmentación a la figura de interés usando el método de contornos activos en base a teoría de información con respecto al método de contornos activos geodésicos. Esto debido a que el flujo de gradiente utilizado para evolucionar el

conjunto de nivel de la curva de segmentación proviene de un funcional de energía que busca maximizar la información mutua entre la intensidad de los píxeles de la imagen y la zona relativa a la curva (9), por lo cual, es de esperar que dicho flujo gradiente logre un buen desempeño al segmentar la imagen de acuerdo a la densidad de probabilidad de la intensidad de sus píxeles. Además, el desempeño de segmentación del método de contornos activos geodésicos depende fuertemente de la correcta elección de los parámetros del filtro gaussiano utilizado para obtener la función de detención $g(I)$ presentada en la figura 6, donde se observa que con los parámetros escogidos, no se distingue con claridad los bordes de la figura de interés a segmentar. Es por esto, que el desempeño del método mencionado no logra segmentar con precisión la imagen. Esto no quiere decir que dicho método sea peor que el método basado en teoría de información, sino que el método geodésico depende de la adecuada elección de parámetros, los cuales varían dependiendo de la imagen de entrada a segmentar, donde en el ejemplo presentado no se escogieron adecuadamente los parámetros, mientras que el desempeño del método basado en teoría de información no se ve completamente determinado por esta elección paramétrica, por lo cual se dice que es un modelo no paramétrico, siendo esta una clara ventaja con respecto al método de segmentación geodésico.

VII. CONCLUSIONES

En el presente trabajo se ha estudiado e implementado un método de segmentación de imágenes no paramétrico utilizando un marco teórico a partir de la teoría de la información. Se ha mostrado la equivalencia entre la formulación del problema de segmentación y la maximización de un funcional dependiente de la información mutua entre la intensidad de píxeles y la correcta segmentación de áreas en la imagen. Se ha encontrado el flujo de gradiente para la evolución de la curva en forma de expresión cerrada y se ha implementado según el método de curvas de nivel y otras modificaciones un algoritmo capaz de evolucionar la curva y probar su desempeño.

Se realizó una comparación entre el desempeño del algoritmo de contornos activos geodésicos y el método propuesto y se obtiene como resultados que es posible implementar un método de segmentación de imágenes sin parámetros ajustables, como sería el caso de contornos activos geodésicos, pero este resultado se obtiene a costo de aumentar la cantidad de tiempo en cálculos.

Finalmente se concluye que el método propuesto cumple a cabalidad el objetivo de segmentar imágenes sin la utilización de parámetros ajustables de manera satisfactoria, además que su tiempo de cálculo puede ser considerablemente reducido si se implementan métodos de estimación probabilísticas aproximados, como el caso de FGT.

VIII. AGRADECIMIENTOS

Al equipo docente por la ayuda brindada, al laboratorio de procesamiento de imágenes por la ayuda en la implementación de la evolución por conjuntos de nivel, y especialmente, a nuestro queridísimo tercer integrante Hans Starke, por escoger

la temática del proyecto y abandonar el curso en las primeras semanas.

APÉNDICE A

En este apéndice se procede a demostrar la proposición mencionada en la *sección II*, la cual hace referencia al hecho que la curva cerrada \vec{C} que maximiza la información mutua $I(G(x); L_{\vec{C}}(X))$ cumple (3) o en otras palabras resuelve el problema de segmentación de imagen, para ello se consideran conocidas las distribuciones p_1 y p_2 y las regiones R_1 y R_2 .

Primero es posible notar que se cumple:

$$I(G(x); L_{\vec{C}}(X)) \leq I(G(x); X) \quad (28)$$

Obtenido directamente de la desigualdad de procesamiento de datos que postula que para $\forall f: A \rightarrow B$ y la variable aleatoria X que toma valores en A se cumple que:

$$I(Y; f(X)) \leq I(Y; X) \quad \forall v. a. Y \quad (29)$$

Luego $I(G(x); X)$ corresponde a un límite superior fundamental de $I(G(x); L_{\vec{C}}(X))$, luego lo que es necesario demostrar es que la igualdad se alcanza cuando se cumple (3) o \vec{C} es solución del problema de segmentación. Para ello descomponemos $I(G(x); X, L_{\vec{C}}(X))$ por regla de la cadena como sigue:

$$\begin{aligned} I(G(x); X, L_{\vec{C}}(X)) &= I(G(x); L_{\vec{C}}(X)) + I(G(x); X | L_{\vec{C}}(X)) \\ &= I(G(x); X) + I(G(x); L_{\vec{C}}(X) | X) \end{aligned} \quad (30)$$

Donde es posible notar que la igualdad para (29) se cumple cuando:

$$I(G(x); L_{\vec{C}}(X) | X) = I(G(x); X | L_{\vec{C}}(X)) \quad (31)$$

Para ello reescribimos (13) como:

$$\begin{aligned} H(G(x)) - H(G(x) | (L_{\vec{C}}(X) | X)) \\ = H(G(x)) - H(G(x) | (X | L_{\vec{C}}(X))) \end{aligned} \quad (32)$$

Luego todo el problema se reduce a demostrar que condicionar $G(X)$ ante $L_{\vec{C}}(X) | X$ o $X | L_{\vec{C}}(X)$ induce a la misma densidad de probabilidades. Por un lado es trivial que conociendo el valor de X es conocida la pertenencia de x a las regiones R_+ o R_- entonces $L_{\vec{C}}(X) | X$ define esta pertenencia, de la misma manera, $X | L_{\vec{C}}(X)$ también permite conocer la pertenencia de x a las regiones R_+ o R_- puesto que se impone el conocimiento de $L_{\vec{C}}(X)$.

Ahora suponiendo que la curva cerrada C es óptima en sentido (3) se tiene que $R_1 = R_+$ y $R_2 = R_-$ (o viceversa $R_1 = R_-$ y $R_2 = R_+$), sumado con la conclusión anterior, para ambos casos se conoce la pertenencia de x a los conjuntos R_1 y R_2 , luego

recordando (8):

$$\begin{aligned} \{G(x) \mid x \in R_1\} &\sim p_1 \\ \{G(x) \mid x \in R_2\} &\sim p_2 \end{aligned}$$

Con ello se tiene que:

$$H\left(G(x) \mid (L_{\vec{C}}(X) \mid X)\right) = H\left(G(x) \mid (X \mid L_{\vec{C}}(X))\right) \quad (33)$$

Luego siguiendo la cadena lógica antes mencionada, se logra comprobar que la curva cerrada \vec{C} que es solución del problema de segmentación de imágenes maximiza $I\left(G(x); L_{\vec{C}}(X)\right)$. ■

APÉNDICE B

En este apéndice se probará la derivación del flujo de gradiente de la curva $\vec{C}(t)$ mostrado en la *sección IV-B* para minimizar la integral de energía E , la cual es una región integral sobre el interior de la cura $R(\vec{C}(t))$.

Representando el funcional de energía de la curva como una simple región integral de la forma:

$$E(\vec{C}(t)) = \int_{R(\vec{C}(t))} f(x) dx \quad (34)$$

Donde el integrando no depende de la curva o del tiempo t . Derivando dicha ecuación con respecto al tiempo, se obtiene el siguiente resultado:

$$\frac{dE}{dt} = \oint_{\vec{C}(t)} \langle \vec{C}_t, f \vec{N} \rangle ds \quad (35)$$

Donde se denota $\vec{C}_t = \partial \vec{C}(t) / \partial t$. De la derivación anterior, la forma del flujo de gradiente de \vec{C} (el negativo de su gradiente para que la región integral decrezca más rápidamente), resulta:

$$\frac{d\vec{C}}{dt} = -f \vec{N} \quad (36)$$

Luego, se considera una clase general de regiones basadas en funcionales de energía E , donde el integrando f depende de otra familia de regiones integrales $\varepsilon(x, t)$ sobre $R(\vec{C}(t))$. Notar que las regiones integrales anidadas $\varepsilon(x, t)$ dependen de t , mientras que $R(\vec{C}(t))$ (Región interior de la curva \vec{C}), cambia de acuerdo a evolución de la curva \vec{C} en el tiempo. En la ecuación (21) se asume:

$$E(\vec{C}(t)) = \int_{R(\vec{C}(t))} f(\varepsilon(x, t)) dx \quad (37)$$

Donde:

$$\varepsilon(x, t) = \int_{R(\vec{C}(t))} g(x, \hat{x}) d\hat{x} \quad (38)$$

Denotando $\varphi(x, t) = f(\varepsilon(x, t))$, el funcional de energía queda reescrito como:

$$E(\vec{C}(t)) = \int_{R(\vec{C}(t))} \varphi(x, t) dx \quad (39)$$

Usando los resultados de Delfour y Zolesio [8], la derivada temporal del funcional de la forma (39) resulta:

$$\frac{\partial E(\vec{C}(t))}{\partial t} = \int_{R(\vec{C}(t))} \varphi_t(x, t) dx + \oint_{\vec{C}(t)} \langle \varphi(x, t) \vec{N}, \vec{C}_t \rangle ds \quad (40)$$

Donde $\varphi_t(x, t) = f'(\varepsilon(x, t))(\varepsilon_t(x, t))$. Notando que por construcción, $\varepsilon(x, t)$ tiene forma de una simple región integral (11) para cada x , cuyo integrando g no depende de \vec{C} . Por ende, ε_t puede escribirse, de acuerdo a (12), como:

$$\varepsilon_t(x, t) = \oint_{\vec{C}(t)} \langle \vec{C}_t, g(x, \vec{C}) \rangle ds \quad (41)$$

Adjuntando este resultado en la ecuación (40), se obtiene:

$$\begin{aligned} \frac{dE}{dt} &= \oint_{\vec{C}(t)} \langle \vec{C}_t, (f \circ \varepsilon) \vec{N} \rangle ds \\ &\quad + \int_{R(\vec{C}(t))} f'(\varepsilon(x, t)) \left[\oint_{\vec{C}(t)} \langle \vec{C}_t, g(x, \vec{C}) \rangle ds \right] \\ &= \oint_{\vec{C}(t)} \langle \vec{C}_t, (f \circ \varepsilon) \vec{N} \rangle ds \\ &\quad + \int_{R(\vec{C}(t))} \oint_{\vec{C}(t)} \langle \vec{C}_t, f'(\varepsilon(x, t)) g(x, \vec{C}) \vec{N} \rangle ds dx \\ &= \oint_{\vec{C}(t)} \langle \vec{C}_t, \left[f \circ \varepsilon + \int_{R(\vec{C}(t))} f'(\varepsilon(x, t)) g(x, \vec{C}) dx \right] \vec{N} \rangle ds \end{aligned} \quad (42)$$

Donde se obtiene el siguiente flujo de gradiente para la curva \vec{C} , donde se omitió la dependencia de t por simplicidad:

$$\frac{d\vec{C}}{dt} = - \left[f(\varepsilon(\vec{C})) + \int_{R(\vec{C}(t))} f'(\varepsilon(x, t)) g(x, \vec{C}) dx \right] \vec{N} \quad (43)$$

Obteniendo una expresión general para la derivada temporal de una curva \vec{C} en base a regiones integrales anidadas. ■

APÉNDICE C

En este apéndice se mostrará la derivación del flujo de gradiente de la curva \vec{C} (23) presentado en la *sección IV-C*. En primer lugar, el funcional de energía descrito en la ecuación (9), puede reescribirse de la siguiente forma:

$$E(\vec{C}) = -|\Omega| \hat{h}(G(X)) + E_+(\vec{C}) + E_-(\vec{C}) + \alpha \oint_{\vec{C}} ds \quad (44)$$

Donde las componentes $E_+(\vec{C})$ y $E_-(\vec{C})$ quedan descritas por:

$$E_+(\vec{C}) = Pr(L_{\vec{C}}(X) = L_+) \hat{h}(G(X)|L_{\vec{C}}(X) = L_+) \\ = - \int_{R_+} \log \left(\frac{1}{|R_+|} \int_{R_+} K(G(x) - G(\hat{x})) d\hat{x} \right) dx \quad (45)$$

$$E_-(\vec{C}) = Pr(L_{\vec{C}}(X) = L_-) \hat{h}(G(X)|L_{\vec{C}}(X) = L_-) \\ = - \int_{R_-} \log \left(\frac{1}{|R_-|} \int_{R_-} K(G(x) - G(\hat{x})) d\hat{x} \right) dx \quad (46)$$

Luego se procede a calcular el flujo de gradiente para E_+ , notando que el cálculo del flujo de gradiente para E_- es análogo, con un signo opuesto. Como el término $1/|R_+|$ depende de la curva \vec{C} , se separará E_+ en dos términos:

$$E_+ = -(E_+^1 + E_+^2) \quad (47)$$

$$E_+^1 = - \int_{R_+} \log |R_+| dx = -|R_+| \log |R_+| \quad (48)$$

$$E_+^2 = \int_{R_+} \log \left(\int_{R_+} K(G(x) - G(\hat{x})) d\hat{x} \right) dx \quad (49)$$

Notando que el término E_+^2 exhibe la forma de integral anidada mostrada en (21) y usando el resultado mostrado en (22), el flujo de gradiente de E_+^2 resulta:

$$\nabla_{\vec{C}} E_+^2 = - \left[\log \left(\int_{R_+} K(G(\vec{C}) - G(\hat{x})) d\hat{x} \right) \right. \\ \left. + \int_{R_+} \left(\frac{1}{\int_{R_+} K(G(x) - G(\hat{x})) d\hat{x}} \right) \cdot K(G(x) - G(\vec{C})) dx \right] \vec{N} \quad (50)$$

$$\nabla_{\vec{C}} E_+^2 = - [\log |R_+| + \log \hat{p}_+(G(\vec{C}))] \\ + \frac{1}{|R_+|} \int_{R_+} \frac{K(G(x) - G(\vec{C}))}{\hat{p}_+(G(x))} \vec{N} \quad (51)$$

Mientras que el flujo de gradiente de E_+^1 resulta:

$$\nabla_{\vec{C}} E_+^1 = -(\nabla_{\vec{C}} |R_+|) \log |R_+| - \nabla_{\vec{C}} |R_+| \\ = (1 + \log |R_+|) \vec{N} \quad (52)$$

Por lo tanto el flujo gradiente total del funcional de energía E_+ resulta:

$$\nabla_{\vec{C}} E_+ = -(\nabla_{\vec{C}} E_+^1 + \nabla_{\vec{C}} E_+^2) \\ = \left[-1 + \log \hat{p}_+(G(\vec{C})) \right. \\ \left. + \frac{1}{|R_+|} \int_{R_+} \frac{K(G(x) - G(\vec{C}))}{\hat{p}_+(G(x))} \right] \vec{N} \quad (53)$$

Mientras que la expresión del flujo de gradiente para E_- resulta:

$$\nabla_{\vec{C}} E_- = - \left[-1 + \log \hat{p}_-(G(\vec{C})) \right. \\ \left. + \frac{1}{|R_-|} \int_{R_-} \frac{K(G(x) - G(\vec{C}))}{\hat{p}_-(G(x))} \right] \vec{N} \quad (54)$$

Notando que el vector normal presenta signo opuesto a la expresión de flujo de gradiente obtenido para E_+ .

Finalmente, combinando las expresiones obtenidas, el flujo de gradiente de $E(\vec{C})$ resulta:

$$\frac{\partial \vec{C}}{\partial t} = \left[\log \frac{\hat{p}_+(G(\vec{C}))}{\hat{p}_-(G(\vec{C}))} + \frac{1}{|R_+|} \int_{R_+} \frac{K(G(x) - G(\vec{C}))}{\hat{p}_+(G(x))} \right. \\ \left. + \frac{1}{|R_-|} \int_{R_-} \frac{K(G(x) - G(\vec{C}))}{\hat{p}_-(G(x))} \right] \vec{N} \quad (55)$$

Obteniendo una expresión cerrada para el flujo de gradiente de la curva $\vec{C}(t)$. ■

REFERENCIAS

- [1] J.Kim, J.Fisher, A.Yezz, M.Çetin and A.Willsky, "A Nonparametric Statistical Method for Image Segmentation Using Information Theory and Curve Evolution" en *IEEE Transactions on Image Processing*, Vol. 14, No. 10, 1486-1502 Octubre 2005. [Online]. Available: <https://people.csail.mit.edu/fisher/publications/papers/kim04tip.pdf>
- [2] Rajeshwar Dass, Priyanka y Swapna Devi, Image Segmentation Techniques en *IJEET*, Vol. 3, Issue 1, Marzo 2012. [Online]. Available: <http://www.iject.org/vol3issue1/rajeshwar.pdf>
- [3] R. Kandwal, A. Kumar y S. Bhargava, "Review: Existing Image Segmentation Techniques" en *International Journal of Advanced Research in Computer Science and Software Engineering*: Volume 4, Issue 4, 153-156, Abril 2014. [Online]. Available: <https://pdfs.semanticscholar.org/ac6f/955c8487d2c5c03ee4b147fe0d75115c565b.pdf>
- [4] J. A. Sethian, Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision, and Material Science. Cambridge, U.K.: Cambridge Univ. Press, 1996.
- [5] D. Adalsteinsson and J. A. Sethian, "A fast level set method for propagating interfaces," *J. Comput. Phys.*, vol. 118, pp. 269-277, 1995. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.46.1716&rep1&type=pdf>

- [6] Ahmad, I. A. and Lin, P. E. A nonparametric estimation of the entropy for absolutely continuous distributions. *IEEE Trans. Information Theory*, 22, 372-375, 1976.
- [7] L. Greengard and J. Strain, "The fast Gauss transform," *SIAM J. Sci. Stat. Comput.*, vol. 12, no. 1, pp. 79-94, 1991.
- [8] M. C. Delfour and J. Zolesio, *Shapes and Geometries: Analysis, Differential Calculus, and Optimization*. Philadelphia, PA: SIAM, 2001 [Online]. Available: <http://www.sam.math.ethz.ch/~hiptmair/Seminars/SHAPECALC/Papers/DEZ11.pdf>
- [9] V. Calles, R. Kimmel and G. Sapiro, "Geodesic Active Contours", *International Journal of Computer Vision* 22 (1), pp.61-79, 1997. [Online]. Available: http://www.dtic.upf.edu/~vcaselles/papers_v/GAC_article.pdf
- [10] J. Larse and T. Greve, "An Overview of the implementation of Level Set methods, including the use of the Narrow Band method", pp. 5 [Online]. Available: <http://cs.au.dk/~tgk/courses/LevelSets/LevelSet.pdf>
- [11] F. Córdova and M. Videla, "Método No Paramétrico de Segmentación de Imágenes Basado en Teoría de Información", implementation on Python Notebook, Colaboratory [Online]. Available: https://colab.research.google.com/drive/1O1BgjO9OMnKpsSphnaXQed_aqy2t24GjV