

UN INDICADOR DE CALIDAD PARA EVALUAR SUPERFICIES SEGMENTADAS

D'Amato Juan^{a,b}, del Fresno Mariana^a y Vénere Marcelo^{a,c}

^a *PLADEMA, Universidad Nacional del Centro, 7000 Tandil, Argentina,*

^b *CONICET*

^c *CNEA*

{jpdamato, mdelfres, venerem}@exa.unicen.edu.ar, <http://www.pladema.net>

Palabras clave: medición de volumen, segmentación, indicador de error.

Resumen. Se presenta un indicador que permite evaluar la calidad de segmentaciones de imágenes digitales tridimensionales, comparando directamente la superficie de referencia con la obtenida mediante algún método de segmentación. El mismo se basa en el cálculo del volumen encerrado entre las dos superficies, clasificado en falso positivo y falso negativo. Se propone además un método robusto y computacionalmente eficiente para realizar este cálculo. Se incluye la validación del método implementado con casos sintéticos y segmentaciones de objetos con geometrías complejas.

1 INTRODUCCIÓN

Uno de los problemas que enfrenta la investigación en métodos de segmentación de imágenes digitales es contar con alguna métrica que indique la calidad de la segmentación obtenida. Generalmente, las propuestas se orientan a la evaluación de imágenes médicas, ya que dado el tipo de aplicaciones involucradas, la eficiencia de la segmentación es un aspecto crítico para establecer la validez de los algoritmos y su aplicabilidad dentro de ambientes clínicos. Se han propuesto numerosas definiciones y medidas de calidad con el fin de analizar la performance de los algoritmos y efectuar comparaciones (Chalana and Kim, 1997; Pham et al., 2000; Pommert and Höhne, 2002; Cardoso and Corte-Real, 2005; Udupa et al., 2006). En algunos trabajos se considera un análisis visual de los resultados por parte de expertos si bien hay un consenso casi generalizado en la necesidad de criterios objetivos de performance de los algoritmos de segmentación.

Típicamente, hay dos alternativas para efectuar una validación de métodos de segmentación. Una de ellas es comparar el resultado obtenido contra segmentaciones manuales efectuadas por usuarios expertos aunque, debido a la subjetividad inherente, no se puede garantizar un modelo de referencia único. En este sentido, se han propuesto algunos métodos estadísticos para cuantificar la performance de un algoritmo con respecto al resultado de segmentaciones por parte de expertos. Otra posibilidad es mediante el uso de fantasmas, ya sea físicos o computacionales (Collins et al., 1998). En el caso de imágenes médicas, los fantasmas físicos brindan un modelo del proceso de adquisición de la imagen pero por lo general no proveen una representación realista de la anatomía. En cambio, los fantasmas computacionales se derivan de una colección de imágenes, comúnmente de tomografías computadas (CT) o de resonancias magnéticas (MRI). Por ejemplo, en el sitio <http://www.bic.mni.mcgill.ca/brainweb/> del Montreal Neurological Institute (MNI) se encuentra disponible un fantoma digital 3D de MRI de cabeza, el cual ha sido utilizado en diferentes trabajos para la evaluación de resultados, ya que además permite efectuar simulaciones en base a diferentes configuraciones de ruido y uniformidad en la imagen. Sin embargo, debe tenerse en cuenta que algunos autores han señalado que este tipo de imágenes no siempre modelan las características encontradas en la práctica, ya sea en situaciones normales o patológicas (Warfield et al., 2004).

Se han propuesto diferentes métricas para la comparación respecto de fantasmas o imágenes de referencia (Zhang, 1996), algunas de las cuales presentan ciertas deficiencias. Por ejemplo, una de las métricas más simples, frecuentemente usada en estudios de confiabilidad, consiste en comparar el volumen encerrado en la imagen segmentada y en la imagen de referencia. Sin embargo, esta medida no constituye una métrica representativa ya que dos conjuntos de voxels pueden constituir volúmenes similares aunque las segmentaciones correspondientes sean muy diferentes o incluso disjuntas (Udupa et al., 2006). Una alternativa que permite considerar las propiedades espaciales de las estructuras segmentadas es mediante una comparación por superposición relativa. Se han propuesto métricas que analizan si existe coincidencia entre las segmentaciones o si se producen errores en la clasificación, ya sea por falsos positivos (elementos de la imagen que han sido segmentados por el algoritmo pero que no están incluidos en la imagen de referencia) o por falsos negativos (elementos que sí pertenecen a la segmentación de referencia pero no han sido detectados por el algoritmo). En todos los casos, estas medidas están orientadas al dominio discreto –es decir que consideran la proporción de voxels coincidentes o mal clasificados– basándose en una decisión binaria sobre su pertenencia o no a alguna de las segmentaciones. No obstante, al considerar el ajuste de superficies en el dominio continuo, las métricas basadas en conteo de voxels no resultan directamente aplicables. Si bien es posible identificar cuáles voxels resultan incluidos dentro

de la superficie detectada y calcular tales medidas, esta tarea no siempre resulta trivial dependiendo de la técnica de segmentación utilizada. Además, en este caso se desestima información valiosa sobre la posición exacta de la superficie y tampoco se pueden comparar finamente las diferencias entre distintas segmentaciones (especialmente considerando que la mayoría de los errores ocurren en la superficie, sobre los límites del objeto).

Cuando el resultado de la segmentación es una curva, una medida más adecuada para caracterizar el grado de desajuste entre la curva obtenida y la que se utiliza como referencia es medir la distancia entre ellas. En (Chalana and Kim, 1997) se plantea una métrica de distancia entre dos contornos sobre imágenes planas. El cálculo es directo cuando se cuenta con marcadores o puntos de control sobre los contornos; sin embargo, en su ausencia se requiere de un proceso de correspondencia artificial entre las curvas, que los autores resuelven mediante el cálculo de la distancia de Hausdorff. Un enfoque similar pero basado en la distancia entre superficies ha sido adoptado en (Cignoni et al., 1998) para el cálculo del error de aproximación en la simplificación de mallas, aunque posee algunas limitaciones en cuanto a la detección de cambios topológicos que pueden producirse en el proceso de simplificación. También se ha considerado la distancia de Hausdorff en (Gerig et al., 2001), aunque debido al alto costo computacional involucrado en su cálculo sobre un dominio tridimensional, los autores plantean utilizar una distancia euclideana como alternativa para la comparación entre las superficies evaluadas.

2 INDICADOR PROPUESTO

Dada una superficie segmentada de referencia R (la que se considera como el resultado correcto) para un determinado componente de interés dentro de una imagen digital, la idea es encontrar un indicador de calidad que mida cuán cerca está la superficie segmentada S (obtenida a partir de la aplicación de un cierto método) respecto de R . Para ello en primer lugar es necesario definir “distancia entre dos superficies” y posteriormente evaluarla con un algoritmo robusto y a un costo computacional razonable. Este es un problema geométrico para el que aún no se cuenta con soluciones obvias, y que tiene múltiples aplicaciones, incluso en temas tan dispares como física relativista (Anderson, 2004).

Una medida posible para evaluar esta “distancia entre dos superficies” y que puede utilizarse como indicador de calidad de la segmentación, es la siguiente:

$$Q = \frac{V_R \cap V_S}{V_R \cup V_S} \quad (1)$$

donde V_S es el volumen encerrado por la superficie S y V_R es el encerrado por R . Esta métrica alcanza el valor 1 cuando ambos volúmenes coinciden y es cercana a 0 cuando no comparten volumen clasificado.

En este trabajo proponemos utilizar como indicador de distancia al volumen encerrado entre las dos superficies contabilizado en forma separada como la suma de los falsos positivos (Q^+) y los falsos negativos (Q^-):

$$\left. \begin{aligned} Q^+ = \text{falsos positivos} &= \frac{V_{SR}}{V_R \cup V_S} \\ Q^- = \text{falsos negativos} &= \frac{V_{RS}}{V_R \cup V_S} \end{aligned} \right\} \quad (2)$$

donde V_{SR} es el volumen encerrado entre S y R cuando S es exterior y R interior y V_{RS} es

volumen entre R y S donde ahora R es exterior y S interior (ver Figura 1). Para que el indicador quede normalizado entre 0.0 y 1.0, siendo 1.0 la calidad máxima, se propone escribirlo como:

$$Q = 1 - \left(\frac{V_{SR}}{V_R \cup V_S} + \frac{V_{RS}}{V_R \cup V_S} \right) \quad (3)$$

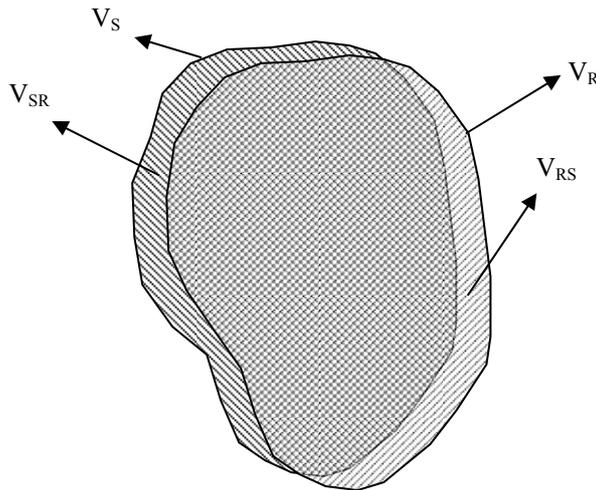


Figura 1: Volumen segmentado y volumen de referencia

Lamentablemente, tanto el cálculo numérico de las diferencias V_{SR} y V_{RS} , como el de la unión $V_R \cup V_S$ es un problema complejo en sí mismo. En este trabajo se propone resolverlo mediante un algoritmo basado en la discretización del espacio y el uso del test del rayo (Foley, 1996), el cual se detalla en el punto siguiente.

2.1 Cálculo del volumen encerrado entre dos superficies

Dada una superficie cerrada (que encierra un volumen) descrita mediante una triangulación orientada, la forma más sencilla para evaluar este volumen es generar una malla de tetraedros, tomando cada uno de los triángulos de la superficie y un mismo punto arbitrario en el espacio. La suma de los volúmenes de estos tetraedros así formados dará el volumen total encerrado independientemente del punto arbitrario elegido. Sin embargo, en el caso que estamos planteando –el volumen encerrado entre dos superficies–, esta idea no es aplicable ya que el resultado de este algoritmo será la diferencia de volumen entre las dos superficies independientemente de su posición relativa y no el volumen encerrado entre las dos.

En general es muy difícil obtener un método continuo y preciso para la evaluación de este volumen en el caso de superficies arbitrarias en el espacio, por lo cual es necesario utilizar técnicas basadas en aproximaciones discretas. La idea del algoritmo que proponemos para este cálculo es discretizar dos de las tres dimensiones, generando una matriz de $N \times M$ celdas, desde las cuales se originan rayos paralelos entre sí. Estos rayos pueden intersectar o no a las superficies, generando una serie de segmentos tal como se muestra en la Figura 2. Para cada rayo, los segmentos que atraviesan el volumen serán clasificados secuencialmente en “out” e “in”. La suma de las longitudes de los segmentos “in” por el área de la celda es una aproximación del volumen encerrado por las dos superficies. Obviamente, a medida que se

utilicen celdas más pequeñas (una discretización más detallada) el error de la aproximación será menor.

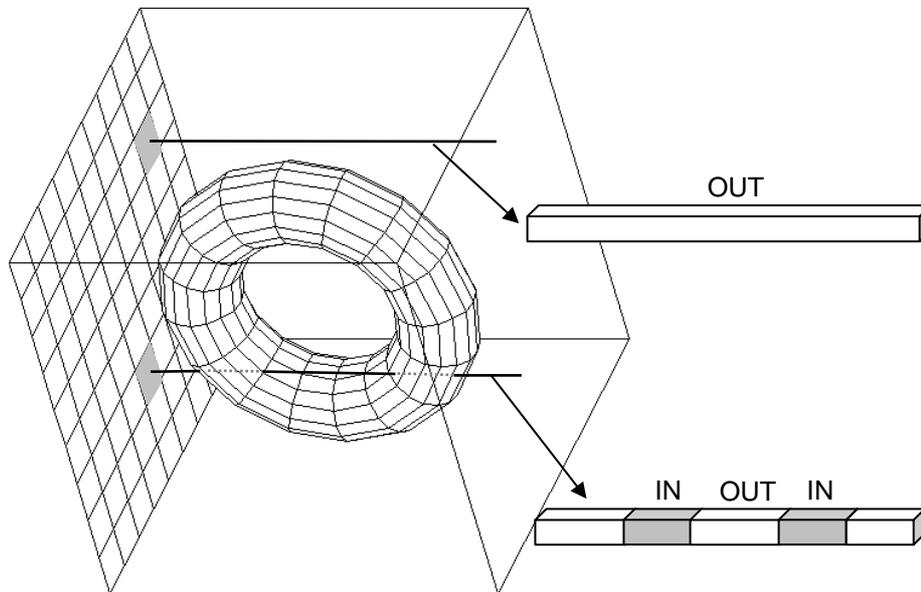


Figura 2: Discretización del espacio y generación de rayos

El volumen total entonces, se logra de la acumulación de los volúmenes parciales por celda. Para realizar la medición, se genera un rayo en dirección normal al plano de ésta y se guarda la distancia a los triángulos intersectados. Ordenando y numerando las intersecciones de acuerdo a la distancia al origen del rayo se observa que cada intersección con número impar indica que el rayo ingresó al volumen y las pares que el rayo ha salido del volumen. En la Figura 2 se muestra un caso con cero intersecciones y otro con cuatro.

El costo computacional del algoritmo así planteado es excesivamente alto ($O(\text{num_rayos} * \text{num_triángulos})$), pero esto puede mejorarse sustancialmente clasificando previamente los triángulos. Este paso de clasificación consiste en asociar los triángulos de la malla con las celdas, siempre y cuando el triángulo al ser proyectado al plano de la celda presente alguna superposición con ésta. En la Figura 3 se esquematiza esta clasificación, donde las celdas que tienen algunos triángulos asociados son las que reciben algo de “sombra” de la geometría.

La clasificación tiene un costo computacional $O(\text{num_triángulos} + \text{num_celdas})$ y posteriormente, si el número de celdas es suficiente, el cálculo de intersecciones tendrá un costo $O(\text{num_celdas})$ (asumiendo que cada celda tendrá muy pocos triángulos, lo cual es razonable para superficies medianamente regulares).

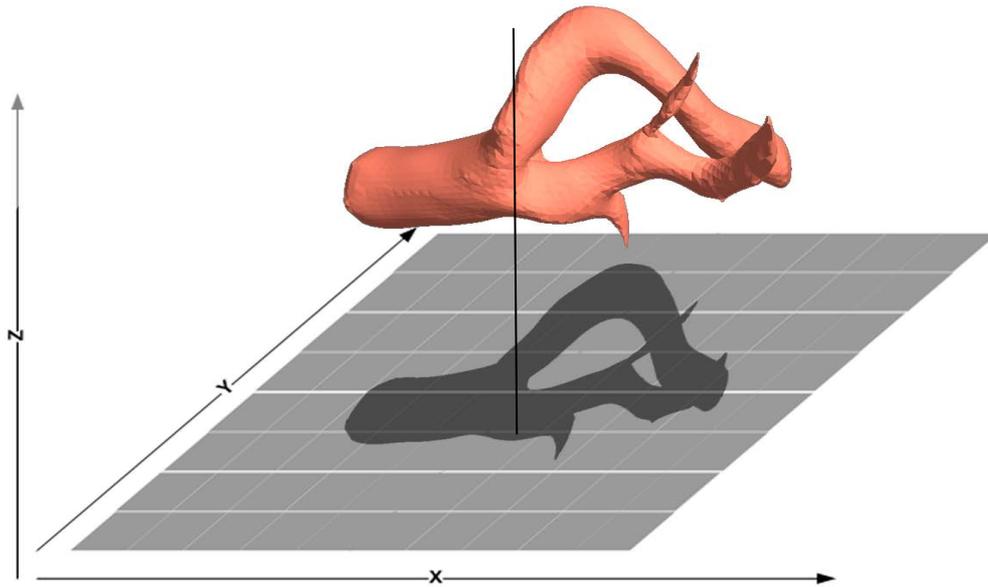


Figura 3: Proyección del objeto 3D sobre el plano de las celdas

El cálculo del volumen encerrado total es de la forma:

$$Volumen_{TOTAL} = \sum_{c=1}^{NC} sup_c \sum_{i=0}^{I/2} dist(Inter[c, 2 \cdot i], Inter[c, 2 \cdot i + 1]) \quad (4)$$

donde NC es el número de celdas de la discretización usada, I es el número de intersecciones correspondientes a cada celda, sup_c corresponde a la superficie de la celda y $dist(\bullet)$ evalúa la distancia euclidiana entre los puntos de intersección del rayo con cada segmento "in".

Para validar el algoritmo, medimos en primer lugar el volumen encerrado por una única malla arbitraria (en la forma convencional), y lo comparamos con el volumen obtenido mediante la técnica propuesta. En la Figura 4 mostramos el comportamiento del error al aumentar la cantidad de celdas, medido dentro de un volumen de $1 \times e9$ unidades cúbicas.

Como puede observarse, para resoluciones superiores a 150×150 celdas, el error relativo que se comete es inferior a 0.05%, valor más que aceptable para las aplicaciones reales.

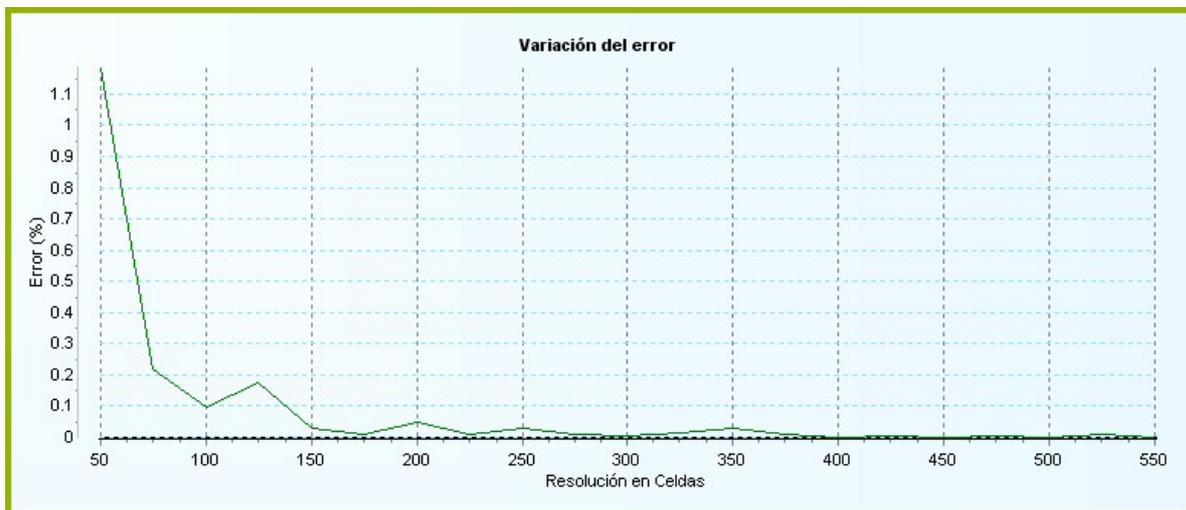


Figura 4: Variación del error relativo con respecto al número de celdas

2.2 Problemas de robustez

Aunque el método así planteado es correcto, existen casos puntuales en los que el test del rayo falla al evaluar alguna intersección por problemas de precisión. Estos casos de error se dan por ejemplo cuando el rayo atraviesa una arista compartida por dos triángulos. En esta situación el algoritmo puede detectar dos, una o ninguna intersección, fallando entonces en la clasificación “in” - “out”. Por ejemplo, en la Figura 5 se muestra un caso en que el algoritmo contabilizó intersecciones erróneas, por lo que se supuso que algunos tramos de rayo se encontraban dentro de un volumen y generó un sub-volumen interno incorrecto.

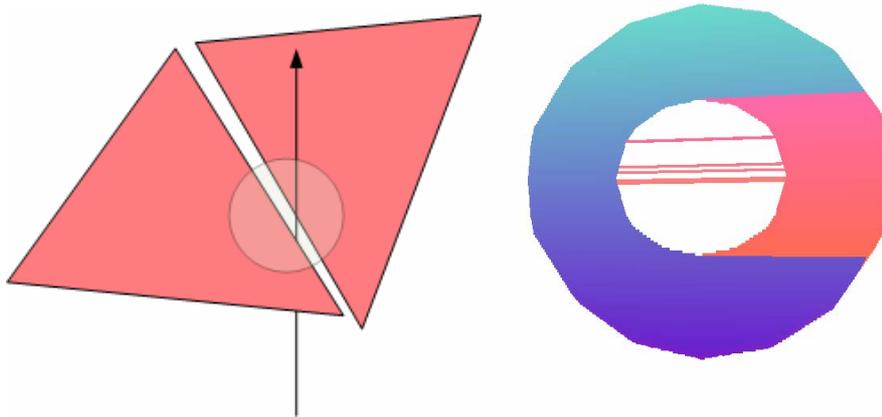


Figura 5: Caso de error cuando un rayo atraviesa una arista, contabiliza cero o dos cortes (izq.), el algoritmo supuso que se encontraba dentro del volumen y lo rellena parcialmente (der.)

Este problema puede generar errores importantes y si bien se supone que debería ser un caso excepcional, en la práctica puede presentarse frecuentemente ya que muchas veces las superficies segmentadas tienen patrones regulares orientados con los ejes cartesianos.

Para solucionar este problema, se propone una heurística razonable basada en pequeñas perturbaciones moviendo el rayo parcialmente para que no intersecte en la arista que une dos triángulos. La idea es lanzar más de un rayo por celda, paralelos entre sí, contar la cantidad de intersecciones generada por cada uno y quedarse con la cantidad que resulte mayoría. En nuestro caso una lógica dos de tres resultó suficiente para no tener este problema en ninguno de los casos estudiados.

3 EXTENSIÓN PARA LA CLASIFICACIÓN EN FALSOS POSITIVOS Y FALSOS NEGATIVOS

El caso que nos interesa es trabajar con dos objetos, pero ahora cada intersección implica tres casos que pueden llegar a ser usados posteriormente, ya sea que el rayo:

- Entró y salió de un único objeto
- Entró por el objeto 1 y salió por el objeto 2
- Entró por el objeto 2 y salió por el objeto 1

El indicador propuesto para la evaluación de diferencias de volúmenes es la suma de estos componentes. Solo si los dos objetos son coincidentes de forma exacta, el volumen encerrado será nulo. A continuación, se muestra el pseudo-código del algoritmo propuesto:

```

// Clasificación de los triángulos
Para todos los triángulos t de todos los objetos
t' = ProyecciónXY (t)
Para todas las celdas que se superponen a t'
c->lista_Triangulos-> añadir (t);

// computar el volumen
Para todas las celdas c
{
  Lista<Vertices> lt = intersección ( c->rayo, c->lista_Triangulos)
  lt . ordenarxDistancia()

  mientras i < lt.longitud / 2
    volumen_celda +=( lt[2*i+1]->z - lt[2*i]->z ) * c->superficie
}

```

3.1 Aplicación a casos sintéticos

Las pruebas realizadas inicialmente tenían el objetivo de validar el algoritmo para la medición de volumen. Para ello se utilizaron tres superficies con creciente complejidad: una esfera, una arteria carótida y una segmentación de cerebro a partir de una MRI del sitio del MNI. Las triangulaciones de estas superficies se pueden ver en las Figuras 6 y 7:

La Tabla 1 resume los resultados obtenidos para distintas discretizaciones (cantidad de celdas):

Superficie	Volumen (real)	Volumen aproximado con discretizaciones					
		30 × 30		120 × 120		240 × 240	
Esfera	258.021.241	257.874.168	0.057%	257.910.292	0.043%	258.013.500	0.003%
Carótida	58.296.467	57.760.140	0.920%	58.239.336	0.098%	58.295.301	0.002%
Cerebro	82.311.633	81.688.534	0.757%	82.300.109	0.014%	82.301.756	0.012%

Tabla 1: Error en la medición del volumen para superficies de creciente complejidad

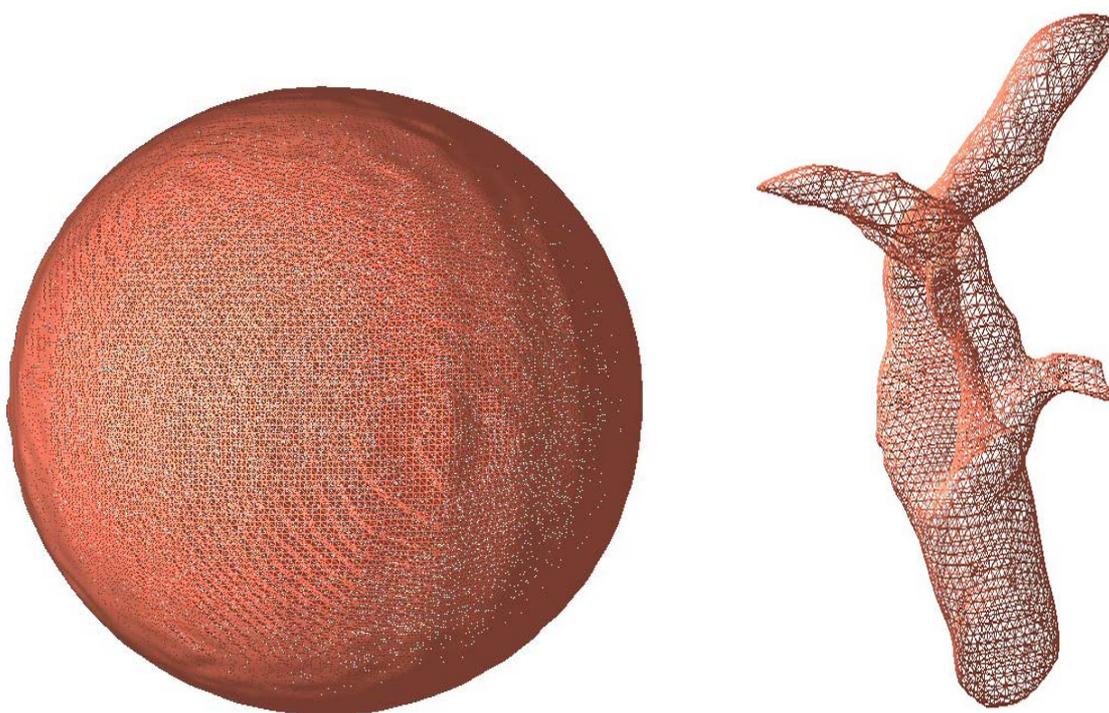


Figura 6: Superficies correspondientes a una esfera (izq.) y una carótida (der.)



Figura 7: Superficie correspondiente a una segmentación de cerebro

Las siguientes pruebas se realizan sobre dos superficies con volumen conocido, donde una contiene completamente a la otra (de forma de poder calcular el volumen encerrado como la resta de los dos volúmenes). En primer lugar se utilizaron dos esferas (ver Figura 8) y finalmente el cerebro incluido dentro de un cubo (Figura 9). La Tabla 2 muestra los errores para estos dos casos:

Superficies	30×30	60×60	120×120	240×240
Esfera/esfera	0.122%	0.026%	0.004%	0.001%
Cubo/cerebro	3.125%	0.902%	0.244%	0.014%

Tabla 2: Error en la medición del volumen encerrado entre dos superficies

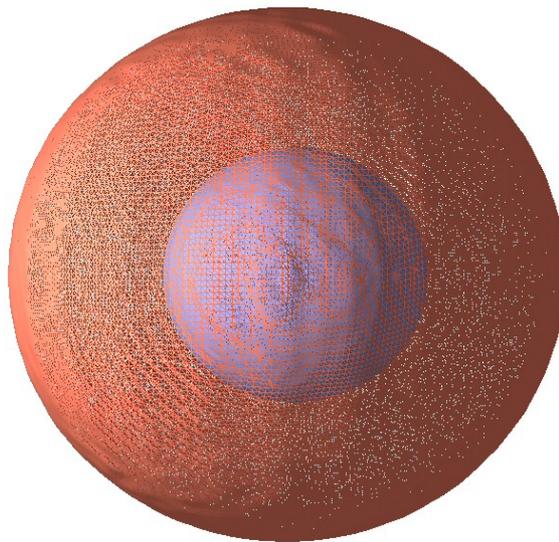


Figura 8: Triangulación para el caso esfera contenida dentro de otra

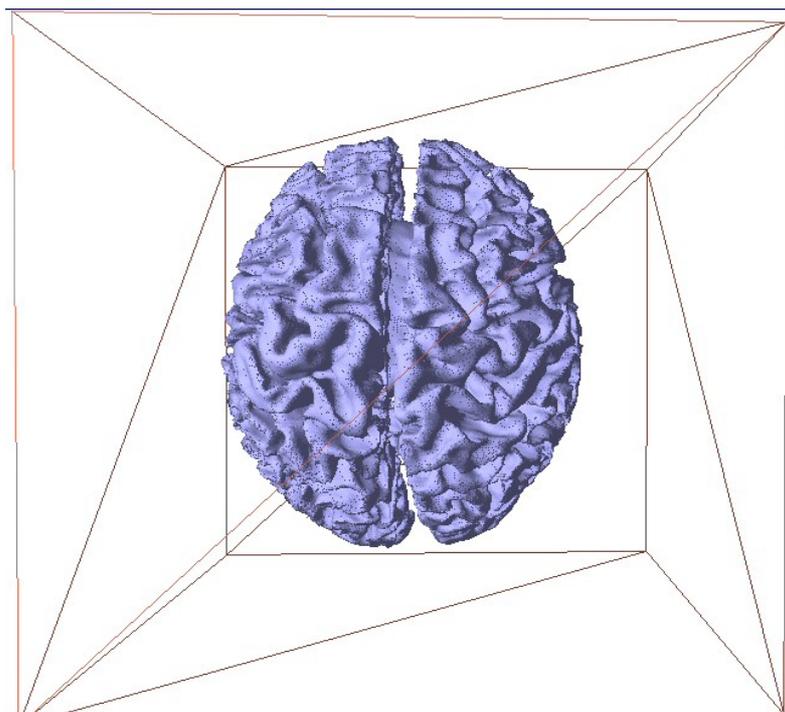


Figura 9: Triangulación para el caso cerebro contenido dentro de un cubo

Como puede observarse, los errores cometidos por la discretización son muy bajos, inclusive para geometrías tan complejas como el caso de la segmentación de cerebro, lo cual valida el método implementado, tanto desde el punto de vista de robustez como de precisión.

3.2 Aplicación a la evaluación de segmentaciones

El indicador propuesto se utilizó además para evaluar la calidad de segmentación en casos reales. El algoritmo de segmentación aplicado se basa en la integración de crecimiento de regiones y modelos deformables (Allevato et al., 2006). La Figura 10 (izq) muestra el caso para una segmentación de carótida, donde la superficie objetivo o de referencia está coloreada en rojo y la segmentación obtenida en azul. En este caso particular, el valor obtenido para el indicador de la ecuación (3) es de 0.892, con una resolución de malla de 50×50 celdas (resoluciones mayores no producen variación notable del indicador para esta segmentación). En la Figura 10 (der) se puede apreciar una segmentación de cerebro del sitio del MNI. En este caso, al comparar la superficie obtenida con la correspondiente a la segmentación de referencia se obtuvo un valor 0.879 para el indicador Q considerando una malla de 180×180 celdas. En la Figura 11 se muestran algunas facilidades que provee la herramienta desarrollada para la evaluación de calidad de las mallas de superficie. En ambos casos se ha aplicado transparencia al objeto de referencia para realzar la diferencia de volumen.

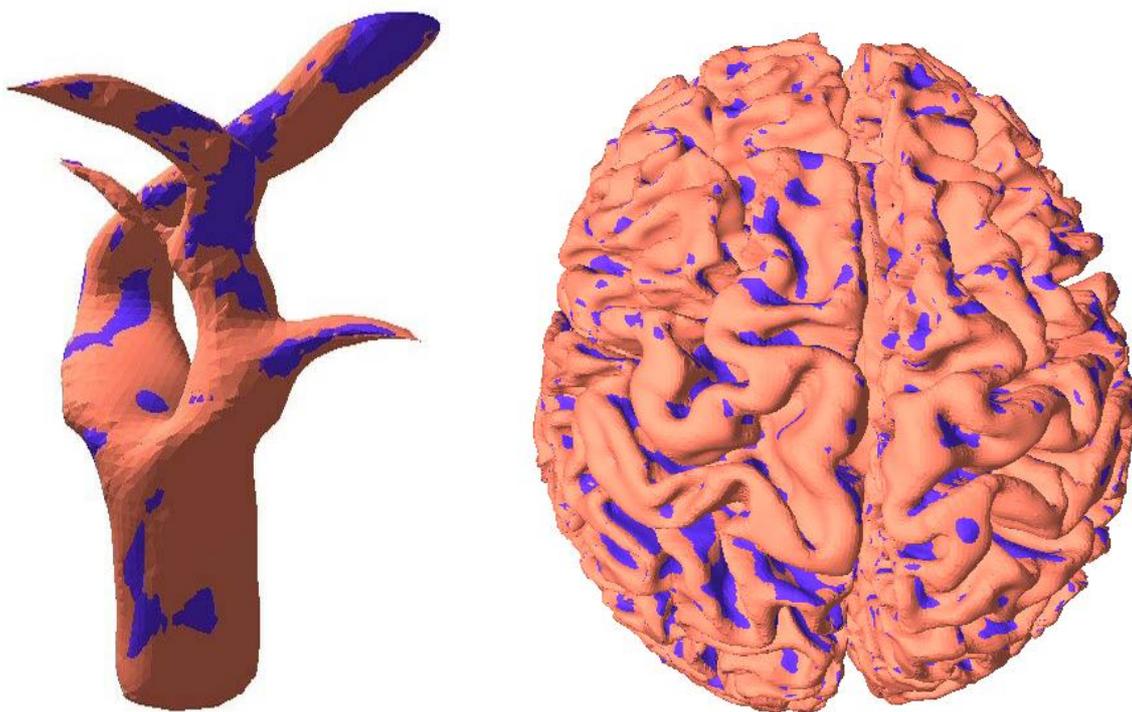


Figura 10. Comparación entre la segmentación obtenida y la superficie de referencia, en el caso de una carótida (izq) y de cerebro (der)

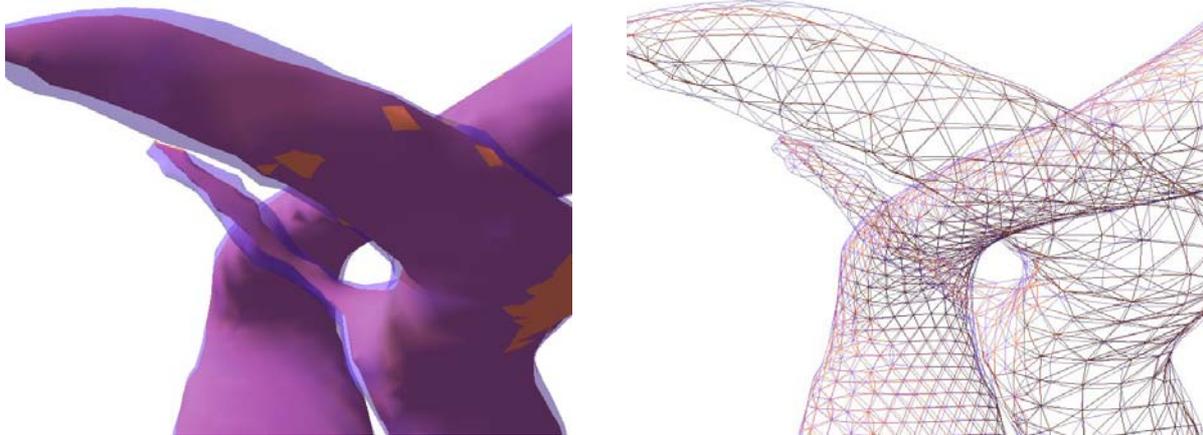


Figura 11. Segmentación de carótida: detalle de comparación entre las mallas de superficie referencia y obtenida por el algoritmo de segmentación, con efecto de transparencia

4 CONCLUSIONES

Se propuso un indicador de calidad que permite obtener una medida de la aproximación de una superficie segmentada a una de referencia. Para ello se definió la distancia entre dos superficies y se propuso un algoritmo robusto basado en la discretización del espacio y el uso del test del rayo, con un costo computacional muy aceptable. A modo indicativo, en la tabla 3 se incluyen los tiempos de cálculo para algunos casos estudiados, realizados a partir de una implementación en lenguaje Delphi 7.0, sobre un equipo PC, Athlon 64x2 Dual Core 2.41 GHZ con 2GB de memoria RAM y una placa Gráfica NVIDIA serie 8800.

Superficies	Núm. triángulos	100 × 100 celdas		200 × 200 celdas	
		clasificación	intersección	clasificación	intersección
Esfera/esfera	307,072	0.22 seg	2.81 seg	0.31 seg	3.48 seg
Cubo/cerebro	808,536	0.56 seg	7.69 seg	0.81 seg	9.22 seg

Tabla 3: Tiempos de cálculo para diferentes resoluciones

Las validaciones realizadas muestran que utilizando resoluciones superiores a 200×200 , los errores de discretización son suficientemente bajos, aún para geometrías tan complejas como la segmentación de cerebro mostrada en la Figura 9.

Los algoritmos propuestos fueron implementados en una herramienta computacional simple de usar y con facilidades gráficas, la cual solo requiere como dato de entrada la descripción de las dos superficies en forma de triangulaciones. La misma puede utilizarse libremente, bajándola de <http://www.exa.unicen.edu.ar/catedras/viscomp/volErrorAp.html>

REFERENCES

- Allevato E., Pedarré R., del Fresno M. y Vénere M., Segmentación de imágenes digitales 3D basado en regiones y contornos activos para la generación de mallas de superficie, *Mecánica Computacional*, V. XXV, 1149-1161, 2006.
- Anderson E., *Geometrodynamics: Spacetime o Space?*, PhD Thesis, University of London,

2004.

- Cardoso J. and Corte-Real L., Toward a Generic Evaluation of Image Segmentation, *IEEE Transactions on Image Processing*, 14:11, 1773-1782, 2005.
- Chalana V. and Kim Y., A methodology for evaluation of boundary detection algorithms on medical images, *IEEE Transactions on Medical Imaging*, 16: 642-652, 1997.
- Cignoni P., Rocchini C. and Scopigno R., Metro: measuring error on simplified surfaces, *Computer Graphics Forum*, 17:2, 167-174, 1998.
- Collins D., Zijdenbos A., Kollokian V., Sled J., Kabani N., Holmes C. and Evans A., Design and construction of a realistic digital brain phantom, *IEEE Transactions on Medical Imaging*, 17:3, 463-468, 1998.
- Foley J., Van Dam A., Feiner S. and Hughes J., *Introduction to Computer Graphics*, Addison-Wesley, 1996.
- Gerig G., Jomier M. and Chakos M., Valmet: a new validation tool for assessing and improving 3D object segmentation, *Proc. MICCAI 2001, LNCS 2208*, 516-523, Springer-Verlang, 2001.
- Pham D., Xu C. and Prince J., A Survey of current methods in medical image segmentation, *Annual Review of Biomedical Engineering*, 2: 315-337, 2000.
- Pommert A. and Höhne K., Evaluation of image quality in medical volume visualization: the state of the art, *Proc. MICCAI, Part II, Lecture Notes in Computer Science 2489*, Springer-Verlang, Berlin, 598-605, 2002.
- Udupa J., LeBlanc V., Zhuge Y., Imielinska C., Schmidt H., Currie L., Hirsch B. and Woodburn J., A framework for evaluating image segmentation algorithms, *Computerized Medical Imaging and Graphics*, 30: 75-87, 2006.
- Warfield S., Zou K. and Wells W., Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation, *IEEE Transactions on Medical Imaging*, 23:7, 903-921, 2004.
- Zhang Y., A survey on evaluation methods for image segmentation, *Pattern Recognition*, 29: 1335-1346, 1996.