V. Métricas de evaluación para algoritmos de segmentación

1. Introducción

Como se adelantó en la introducción de la memoria del presente proyecto, el desarrollo de algoritmos de segmentación es un área en el que se lleva empleado mucho esfuerzo. No obstante, es patente que hay menos interés puesto en la evaluación precisa de estos algoritmos que en su desarrollo. En el mundo de la investigación el desarrollo de nuevas técnicas de segmentación es un ámbito de gran interés, mientras que no siempre se dedica el esfuerzo suficiente en la evaluación precisa, sistemática y exhaustiva de dichas técnicas usando métricas de medición del desempeño estándar o uniformes. Aunque muchos desarrolladores de algoritmos de segmentación suelen proporcionar algo de información sobre la evaluación de sus algoritmos, aún están por desarrollarse marcos y enfoques globales y consistentes con los que poder evaluar el desempeño de los distintos algoritmos. Aunque han sido propuestas teorías unificadas, el problema no se ha resuelto aún de manera satisfactoria (10) (11) (12) (13) (14) (15). Como consecuencia, encontrar el algoritmo de segmentación adecuado para una tarea en particular, así como elegir los parámetros internos al algoritmo de segmentación que son óptimos para la tarea a resolver es aún un problema. Es por esta razón de vital importancia identificar las herramientas apropiadas para evaluar y comparar algoritmos de segmentación. A continuación se detallan las diversas herramientas que han venido siendo usadas por los desarrolladores a la hora de optimizar y evaluar sus algoritmos.

2. Métricas para la evaluación de algoritmos de segmentación

La tarea de diseñar o elegir una medida apropiada de la efectividad de la segmentación de objetos es una tarea en sí compleja. La métrica de evaluación debería proporcionar información relevante con la tarea, como por ejemplo si es diagnóstica o intervencionista. Por ejemplo, hay tareas como la segmentación de la aurícula izquierda que abordamos en este proyecto, que puede requerir ser llevada a cabo durante un procedimiento quirúrgico (como la ablación de la fibrilación auricular) o que, para otros propósitos diagnósticos, puede ser realizada fuera de procedimiento (como la valoración de características morfológicas o funcionales de la cámara cardiaca segmentada). La métrica elegida debe tener en cuenta este tipo de condicionantes, puesto que según valoremos la importancia de la eficiencia en el desempeño (*performance*), el resultado de la valoración o comparativa de distintos algoritmos puede variar de manera significativa. Sin embargo, en todo tipo de tareas en las cuales esté presente la segmentación de imágenes, hay consenso sobre la existencia de tres tipos de métricas que deben evaluarse, y estas son las que detallaremos en esta sección, a saber: exactitud, estabilidad y eficiencia (16).

1. Precisión

La precisión de una técnica de segmentación se refiere al grado en el que los resultados de la segmentación coinciden con la segmentación verdadera. Hay situaciones en las que la segmentación verdadera es conocida, y otras, como por ejemplo las que tratan con imágenes de pacientes o modelos para la investigación con animales, en las que la segmentación verdadera no se conoce. En dichas situaciones la segmentación verdadera se sustituye por una segmentación manual realizada por un experto, o realizada con otra técnica de imagen, o usando un algoritmo de segmentación del cual se conoce que produce resultados precisos. Idealmente la medida de la precisión debe reflejar el grado de discordancia entre la segmentación verdadera y aquella que se desea evaluar, y no debería depender de las dimensiones de la imagen (17) (18). Se exponen a continuación diversas métricas para evaluar la precisión.

A. Métricas basadas en la distancia

En algunos casos de segmentación la delimitación de la frontera es tan crítica que constituye el objetivo esencial de la segmentación. En estos casos, las métricas basadas en la distancia son importantes, y se usan para medir la distancia entre el borde de la segmentación realizada con el algoritmo a probar y el borde de la segmentación verdadera. Si se denomina *B* al contorno de la segmentación a evaluar y *T* al contorno de la segmentación verdadera, se pueden especificar los mismos como dos conjuntos de vértices tales que $B = \{b_i : i=1 \dots K\}$ y $T = \{t_n : i=1 \dots N\}$. Asimismo, la distancia entre b_i y T puede definirse como $d(b_i,T)=\min_n ||b_i - t_n||$.

Para cada imagen *j*, pueden calcularse 3 parámetros: (a) MAD_j, la diferencia media absoluta, que sería una medida del error medio de la segmentación, (b) $MAXD_j$, la diferencia máxima, una medida del error máximo de la segmentación, (c) PC_j, el porcentaje de los vértices (c_i) desde los cuales las distancias al contorno de la segmentación verdadera *T*, $d(b_i, T)$, es menor que *p* píxeles. Este parámetro se usa para calcular la fracción de puntos que pueden considerarse como muy próximos al contorno verdadero, es decir, que tienen una diferencia no significativa en términos médicos o biológicos.

$$MAD_j = \frac{1}{K}\sum_{i=}^{K} d(b_i, T)$$

$$MAXD_j = \underset{i \in [1,k]}{max}\{d(b_i, T)\}$$

$$PC_j = \frac{núm.elemnts\ en\ \{b_i \in B: d(b_i, T) < 5\ pixels\ \}}{K}$$

B. Métricas basadas en área o volumen

Otras segmentaciones se usan para medir el área o el volumen del objeto. En el caso de la segmentación de la aurícula izquierda, es útil para el diagnóstico conocer las dimensiones de la aurícula izquierda, tanto en volumen como en distancias características, como la distancia desde la pared anterior hasta la posterior, la distancia desde la cara septal de la aurícula hasta la pared libre, o el diámetro de cada una de las venas pulmonares. En estas situaciones se usan métricas basadas en área o en volumen para comparar las segmentaciones producidas por el algoritmo que se desea evaluar con la segmentación verdadera.

Pueden emplearse dos enfoques distintos para evaluar el algoritmo de segmentación. En el primer enfoque, la variable área o volumen se compara con el valor de la segmentación verdadera para calcular un porcentaje de error. Dado que la segmentación verdadera está a menudo sujeto a variabilidades introducidas por el operador o por el sistema que realiza la segmentación, y que la cantidad medida por el algoritmo es a menudo el resultado de interacciones del usuario con el algoritmo (19) (20), deben extraerse estadísticos media y varianza de los parámetros calculados para la evaluación usando métodos estadísticos estándar, como el análisis de la varianza (*ANOVA*) y el t-test (21) (22).

Otro conjunto de medidas para la evaluación del desempeño se inspira en la metodología definida en la diversa literatura acerca de la detección de objetos (11) (23) (24). Si S_s y S_t denotan las superficies o regiones interiores al borde en la imagen segmentada y en la segmentación verdadera, respectivamente, se define el área de verdaderos positivos (*VP*) como el área intersección de S_s y S_t, es decir, $S_{VP} = S_S \cap S_T$. Igualmente, el área de falsos positivos *FP* comprendería al área segmentada en la segmentación del algoritmo que no lo está en la segmentación verdadera, es decir, la resta de ambas áreas $S_{FP} = S_S - S_T$. Asimismo, el área de falsos negativos *FN* se correspondería con la resta opuesta, $S_{FN} = S_T - S_S$, y el área de verdaderos negativos *VN* se correspondería con $S_{VN} = ESCENA - S_s - S_T$. Con estas cantidades, las métricas para evaluar la precisión se definirían de la siguiente manera:

Fracción de verdaderos positivos: $FVP = \frac{S_{VP}}{S_T} = sensibilidad$

Fracción de verdaderos negativos: $FVN = \frac{S_{FN}}{S_T} = especificidad$

Fracción de falsos positivos: $FFP = \frac{S_{FP}}{ESCENA - S_T} = 1 - especificidad$

Fracción de verdaderos negativos: $FVN = \frac{S_{VN}}{ESCENA - S_T} = 1 - sensibilidad$

Con *ESCENA* se denota a la región que comprende todas las posibles regiones segmentadas, *FVP* sería la fracción del área de la segmentación verdadera que está también comprendida en la segmentación del algoritmo; la *FVN* sería la fracción del área de la segmentación verdadera que no se acertó a segmentar en la segmentación del algoritmo, la *FFP* sería la fracción del área de la segmentación del algoritmo que no aparece en la segmentación verdadera, y la *FVN* es la fracción del área del fondo de la escena que no pertenece al área segmentada ni en la segmentación verdadera ni en la segmentación del algoritmo.

Conviene resaltar que *FFP* y *FVN* pueden hacerse arbitrariamente pequeños escogiendo como *ESCENA* una región lo suficientemente grande. En ese caso, los valores calculados no pueden valorarse de una manera significativa. Sin embargo, con una elección adecuada de *ESCENA*, puede optimizarse un algoritmo de segmentación representando la *FVP* (sensibilidad) y la *FFP* (1-especificidad), tal y como se hace en el análisis *ROC*. Variando los parámetros de segmentación y calculando las métricas *ROC*, puede hallarse la solución óptima (11) (23) (24).

Dado que *FVN* y *FFP* pueden ser en la práctica difíciles de calcular por las razonas arriba expuestas, se ha propuesto otra métrica para valorar la precisión de la segmentación. Esta métrica mide la fracción de las regiones falsas segmentadas por el algoritmo, es decir:

Fracción falsa: $$FF = 1 - \frac{S_{FP} + S_{FN}}{S_T}$$

C. Métricas para evaluar el desempeño en un conjunto completo de imágenes

A la hora de evaluar el algoritmo de segmentación, es relevante que se use un número suficiente de imágenes para la evaluación. La demostración del desempeño y la comparación usando un número reducido de imágenes no es suficiente para alcanzar conclusiones estadísticamente significativas.

2. Estabilidad

La estabilidad de un algoritmo de segmentación da una idea de la repetibilidad de la técnica al usarse con un tipo particular de técnica de imagen. Existen fuentes diversas de variabilidad, que pueden deberse a interacciones subjetivas del algoritmo con el usuario, como por ejemplo la inicialización manual, la elección del objeto, y la interacción entre usuario y objeto. En la evaluación de la estabilidad de un algoritmo se ha de determinar la métrica apropiada para un conjunto de imágenes usando tanto el algoritmo como estimaciones repetidas de la segmentación verdadera, es decir, diversas segmentaciones manuales de la misma imagen. De este modo, repetir las segmentaciones proporciona una estimación de la varianza de la misma (25) (26) (27).

Usar estimaciones de la varianza tanto de la segmentación del algoritmo, como de la segmentación verdadera, permite comparar la variabilidad del algoritmo con la variabilidad en la determinación de la segmentación verdadera. Asimismo, puede compararse las variabilidades de diversos algoritmos con otro algoritmo o con el mismo algoritmo pero con un conjunto distinto de parámetros. Estas comparaciones deben incluir pruebas estadísticas como el análisis de varianza (*ANOVA*) y estadísticos F (21) (28). Se debe ser especialmente cuidadoso para asegurar que se emplea un número suficiente de pruebas de segmentación que permitan comparaciones significativas.

3. Eficiencia

La eficiencia de la segmentación da idea del uso práctico del algoritmo. Es frecuente que se mida la eficiencia a través del tiempo de segmentación. Sin embargo, la eficiencia debería incluir la cuantificación de todos los aspectos de la interacción con el usuario así como valorar que la técnica de segmentación sea apropiada o no para todas las imágenes. En consecuencia, además del tiempo de ejecución del algoritmo de segmentación, deben considerarse los tiempos de inicialización, edición e inspección para ser documentados también, junto con la tasa de error.

4. Comparación de algoritmos

Dado que un algoritmo de segmentación está sujeto a variabilidades diversas, como a la variación de los objetos que se segmentan y la interacción con el usuario, cualquier comparación de las métricas para la evaluar la precisión discutidas anteriormente debe realizarse usando los estadísticos apropiados. En consecuencia, las ventajas obtenidas con unos algoritmos respecto de otros, usando las métricas discutidas en esta sección, pueden no ser estadísticamente significativas por el hecho de que la ventaja es pequeña respecto de la variación de la segmentación y/o por el número insuficiente de casos usados para la evaluación.

3. Especificación de la métrica empleada en la evaluación

A continuación se detalla el conjunto de criterios que conforma la métrica de evaluación empleada en el presente proyecto para comparar el desempeño de los distintos algoritmos estudiados.

1. Precisión, Sensibilidad y Estabilidad

Los criterios para la evaluación de algoritmos de segmentación, como se ha visto en el apartado anterior, suelen clasificarse en dos grupos, métricas estadísticas, y métricas geométricas, según se evalúen propiedades relativas a las regiones segmentadas o a su contorno. Las métricas elegidas para la evaluación de los algoritmos estudiados son estadísticas, es decir, relativas a las regiones segmentadas. En particular, precisión y sensibilidad son las dos variables estadísticas fundamentales evaluadas en este estudio, con las que se va a caracterizar a los algoritmos para establecer una comparación que permita discernir lo adecuado o inadecuado de cada uno con respecto a su capacidad de segmentar “mal” o “bien”. La descripción detallada del significado de estos estadísticos ha sido abordada en la sección anterior -

Métricas para la evaluación de algoritmos de segmentación. Pueden definirse las variables estadísticas sensibilidad y especificidad, para la evaluación del desempeño de algoritmos de segmentación, de la siguiente manera:

$$sensibilidad = \frac{VP}{VP + FN}$$

$$especificidad = \frac{VN}{VN + FP}$$

Donde *VP*, *VN*, *FN* y *FP* representan los píxeles clasificados como verdaderos positivos, verdaderos negativos, falsos negativos y falsos positivos, respectivamente. Estas definiciones, ampliamente usadas en el dominio de las imágenes médicas, pueden resultar insuficientes, pues resulta a veces ambiguo cuál criterio de los dos es más relevante para el problema dado. La mayor desventaja de la sensibilidad/especificidad radica en su dependencia con la relación entre los tamaños de imagen y el objeto que se desea segmentar. Por esta razón se han definido métricas estadísticas independientes del tamaño de las imágenes (29). A continuación se detallan dos de ellas, *Dice Similarity Coefficient* (*DSC*) y χ^2:

$$DSC = \frac{2 \cdot VP}{2 \cdot VP + FP + FN}$$

$$\chi^2 = \frac{sensitivity - (VP + FP)}{1 - (VP + FP)} \cdot \frac{specificity - (1 - (VP + FP))}{VP + FP}$$

Para el objetivo de este proyecto, dado que se comparan las mismas imágenes para los tres algoritmos, es razonable asumir que la dependencia con la relación entre el tamaño de la imagen y el objeto segmentado afectará por igual a todas las variables precisión y sensibilidad obtenidas, con lo que podrán compararse sin que el hecho de usar estas variables afecte a la conclusión.

Asimismo, se ha optado por conceder la misma importancia a la sensibilidad que a la precisión. Como consecuencia de este criterio, y para facilitar más la comparación de resultados, se ha definido la similitud. Este punto es abordado en profundidad en la siguiente sub-sección, Similitud.

Así, en el contexto de este proyecto de se han calculado las variables estadísticas precisión y sensibilidad atendiendo a las siguientes expresiones:

$$precisión = \frac{VP}{VP + FP}$$

$$sensibilidad = \frac{VP}{VP + FN}$$

Así, la ***precisión*** nos dará una idea de la sobresegmentación de la segmentación obtenida con el algoritmo, respecto de la segmentación verdadera o *ground truth*, tomando su valor máximo, 1, cuando no existen zonas de falsos positivos (sobre-segmentadas), y tomando valores cercanos al mínimo, 0, cuando existan numerosas zonas de falsos positivos en comparación con los verdaderos positivos presentes.

De otra parte, la ***sensibilidad*** dará cuenta de la subsegmentación que existe en la imagen segmentada con respecto a la segmentación verdadera, tomando su valor máximo, 1, cuando no hay zonas de falsos negativos (sub-segmentadas), y tomando valores cercanos al mínimo, 0, cuando existan numerosas zonas con falsos negativos, es decir, zonas no detectadas como pertenecientes a la aurícula izquierda que en la segmentación verdadera sí pertenecen a la misma.

En lo que resta de esta sección se trata de analizar los distintos aspectos que nos interesan acerca de la precisión y la sensibilidad calculadas para cada algoritmo. ¿Nos interesa la precisión máxima que un algoritmo nos puede aportar? ¿O su sensibilidad máxima? ¿O una ponderación de ambas, independientemente del número de combinaciones sobre el espacio de parámetros que consiga ese valor máximo?

Para cada algoritmo se evalúan diversos estadísticos (dependiendo fundamentalmente, entre otros factores, del espacio de parámetros), que comprenden:

- Precisión máxima
- Sensibilidad máxima
- Curvas precisión y Sensibilidad frente al espacio de parámetros. Estabilidad
- Precisión y sensibilidad medias

Asimismo, en el caso de algunos algoritmos, por conveniencia y a la luz de los resultados obtenidos, se han calculado otros estadísticos adicionales, como varianza e Intervalo de confianza, para caracterizarlos de una manera más significativa y global, lo que se detalla en el apartado de evaluación de cada algoritmo correspondientemente.

Hablaremos de la ***estabilidad*** de los algoritmos refiriéndonos a la zona, en cuanto a rango de parámetros del algoritmo, en la que las segmentaciones obtenidas tienen la propiedad de tener tanto una precisión como una sensibilidad altas. En particular se ha escogido el valor del 90% como referencia. De esta manera, la zona estable de los algoritmos evaluados (cuando exista), es aquel rango de parámetros de entrada del algoritmo que hace obtener precisiones por encima del 90% y simultáneamente sensibilidades por encima del 90%.

2. Similitud

Además, para los distintos algoritmos también se calcula una variable que se ha definido para este proyecto: la ***similitud*** con la imagen *ground truth*. La similitud, en el contexto de este proyecto, se ha definido como una función de la precisión y la sensibilidad, con el fin de poder comparar los resultados obtenidos con un algoritmo dado para todas las imágenes probadas de una manera visualmente más directa. Se define la similitud de una imagen segmentada con la imagen *GroundTruth* en la Ecuación 2, siendo las variables p la precisión de la imagen segmentada y s la sensibilidad:

$$Similitud = 1 - \frac{\sqrt{(1-p)^2 + (1-s)^2}}{\sqrt{2}}$$

Ecuación 2: Definición de similitud

A continuación se analiza en detalle la deducción de esta ecuación, así como la simplificación que se introduce gracias a ella para la comparación de los distintos algoritmos.

A la hora de estimar la bondad de un algoritmo empleando distintas imágenes, (usando la métrica que hemos escogido para este proyecto, basada en la precisión y la sensibilidad), o bien al comparar distintos algoritmos empleando una imagen dada, se hace necesario establecer la precisión y la sensibilidad para cada una de ellas.

Supongamos, para la ilustración de la similitud, que deseamos evaluar la bondad de 10 segmentaciones realizadas con un algoritmo dado. Se tendrá, tras evaluar precisión y sensibilidad para cada una con respecto a la segmentación *GroundTruth*, diez pares de valores, un par por imagen, representando la precisión (en adelante, p) y la sensibilidad (en lo sucesivo, s) correspondientes. Representemos estos pares de puntos en el plano cartesiano (p, s), teniendo en cuenta que los valores de precisión están comprendidos entre 0 y 1, así como los de sensibilidad. Un valor $(p, s) = (1,1)$ se correspondería con una segmentación idéntica a la *GroundTruth*, es decir, con similitud máxima. Si hemos de comparar distintos pares de valores de p y s, hemos de apoyarnos en un criterio que nos permita discernir qué es mejor que qué. Tomando un ejemplo numérico, ¿Qué sería preferible, o considerado mejor, $(p, s) = (0.6, 0.39)$ o $(p, s) = (0.59, 0.4)$? Teniendo en cuenta que la precisión es inversamente proporcional a la sobresegmentación de la imagen segmentada con respecto a la *groundtruth*, y que la sensibilidad es inversamente proporcional a la subsegmentación, es necesario establecer un criterio para poder comparar ambas.

El criterio que se va a establecer para definir la similitud en el contexto de este proyecto es considerar igualmente importantes a precisión y sensibilidad. Esto podría ser de otra manera dependiendo del contexto para el que se analice. Es decir, si en el ámbito en que se está evaluando la segmentación es más sencillo solucionar un problema de sobresegmentación que de subsegmentación, o viceversa, bien por la facilidad de las herramientas usadas para llevar a cabo una u otra tarea en contraste con la otra, o por la dificultad inherente en la imagen para conseguir realizar una u otra operación, puede concederse mayor peso o penalización a una u otra medida. No obstante, para los objetivos de este proyecto y su campo de aplicación es aceptable considerar ambos eventos igualmente indeseables.

Una vez decididos por el criterio expuesto en el anterior párrafo, puede pasarse a representar los diez pares de puntos (p, s) del ejemplo que se estaba considerando para ilustrar la conveniencia de definir la variable similitud. La representación de estos 10 pares (p, s) se muestra en la Ilustración 17.

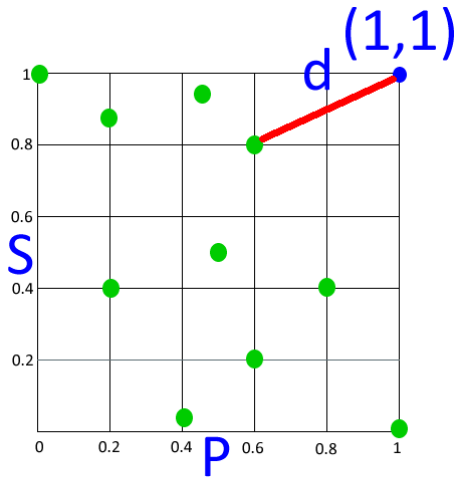


Ilustración 17: Distancia y similitud

Si en lugar de caracterizar los puntos por sus coordenadas (p,s), usamos la distancia al punto (1,1), de máxima similitud, tendríamos una medida de la ***distancia*** a dicho punto, o d (Ecuación 3).

$$d = \sqrt{(1-p)^2 + (1-s)^2}$$

Ecuación 3: Distancia a (1,1) o punto de máxima similitud

Dado que la distancia d puede variar entre 0 y $\sqrt{2}$ (la diagonal), se puede normalizar esta cantidad por su valor máximo, obteniendo así la distancia normalizada d_n al punto de similitud máxima, que varía entre $d = 0$ (punto de similitud máxima) y *d=1* (punto de similitud mínima).

$$d_n = {}^{d}/_{\sqrt{2}}$$

Ecuación 4: Distancia normalizada al punto de máxima similitud

Si finalmente invertimos el concepto distancia por cercanía, tenemos la definición de similitud:

$$Similitud = 1 - d_n = 1 - \frac{\sqrt{(1-p)^2 + (1-s)^2}}{\sqrt{2}}$$

3. Espacio de parámetros

En el contexto de este proyecto, denominaremos espacio de parámetros de cada algoritmo al número de parámetros que el algoritmo requiere por parte del operador, así como por el conjunto completo de valores que cada parámetro puede tomar. Este criterio es de utilidad para cuantificar el número de ejecuciones potencialmente diferentes (diferentes en cuanto a que puedan producir una salida potencialmente diferente a cualquier otra salida que resulte de una ejecución con alguna variación en los parámetros) que un algoritmo dado puede realizar sobre una misma imagen. La caracterización de los algoritmos por medio de su espacio de parámetros facilita el establecimiento de la estrategia de evaluación.

4. Eficiencia

Caracterizaremos asimismo cada algoritmo a través de su tiempo medio de ejecución. En el contexto clínico, y en particular, en la segmentación de la aurícula izquierda con las herramientas comerciales disponibles, es habitual que el ajuste de parámetros se realice siguiendo un método ensayo-error. Ello, unido a que a menudo la segmentación se realiza durante los procedimientos quirúrgicos, hace del coste computacional una variable esencial; Tiempos excesivamente elevados de ejecución combinados con métodos de ajuste de parámetros de ensayo-error pueden disparar los tiempos de segmentación total hasta hacer el uso de los algoritmos inviables para el empleo quirúrgico intra-procedimiento.

Las características del sistema en el cual se ha llevado a cabo la evaluación de los distintos algoritmos y la medición de tiempos de ejecución se muestran a continuación:

CPU	Intel Core™ 2 Quad Q9400@ 2.66GHz
RAM	4 GB
Sistema Operativo	Windows Vista Home Premium 64 bits

Tabla 2: Características Computador

Los tiempos calculados no son válidos para comparar la eficiencia de otros algoritmos fuera del contexto de este proyecto. No obstante, nos permitirán comparar los algoritmos evaluados en este trabajo de una manera fiable y reproducible, puesto que las condiciones con las que se midieron fueron homogéneas.

5. Interacción con el operador

Otro factor a tener en cuenta para caracterizar lo adecuado de cada algoritmo es la rapidez y la complejidad de la interacción necesaria con el operador. Todos los algoritmos evaluados en este proyecto son semiautomáticos, puesto que requieren parámetros que el operador ha de especificar.

En el contexto de este proyecto, la dimensión "interacción con el operador" se ha considerado analizando el número de parámetros que el operador ha de especificar, junto con la naturaleza de los mismos, en cuanto a lo que tengan de comprensibles/intuitivos para el operador. También debe tenerse en cuenta el coste (bien en términos de tiempo empleado, bien en *clicks*, si existe una interfaz visual elaborada para este fin, o en número medio de iteraciones ensayo-error necesarias para llegar a alcanzar la precisión y sensibilidad especificadas, o en una combinación de todo ello) que el operador ha de emplear para obtenerlos.

Los algoritmos evaluados en este proyecto tienen en común que su espacio paramétrico comprende un conjunto que abarca dos conceptos: localización de una parte de la estructura a segmentar (semilla, o *seed* en inglés), y nivel de intensidad de la zona a segmentar (*intensity threshold*). En el caso de algunos algoritmos es necesaria una sola semilla, en otros dos, así como en algún algoritmo se requiere un solo nivel de intensidad, y en otros más de uno. Obviamente ambos conceptos deberían ser comprensibles sin demasiada dificultad para un operador que lleve a cabo el proceso de segmentación. Sin embargo, la situación puede variar considerablemente, desde el punto de vista de coste (en tiempo, en *clicks*, o en iteraciones requeridas) entre tener que proporcionar niveles de intensidad de la zona a segmentar a no tener que hacerlo, ya que se disparan las combinaciones posibles. Sin la ayuda de una interfaz visual, proporcionar uno o varios niveles de gris (comprendido entre 0 y 65535, si se usan codificaciones de píxel de 16 bits, o entre 0 y 255, si se usan 8 bits) en/entre el/los que se encuentra la zona de interés puede ser una tarea algo tediosa, en función de la calidad de la imagen y el contraste de la cámara cardíaca a segmentar con respecto al resto (en el caso específico de la segmentación de la aurícula izquierda). Si a esta dificultad para proporcionar los parámetros que un algoritmo requiere se le añade un tiempo de cómputo elevado, se tiene una percepción de dificultad importante por parte del usuario. Este hecho se ha podido comprobar, especialmente para algunos algoritmos, durante las evaluaciones de tanteo que se hicieron previas a la evaluación total en bloque del espacio paramétrico.

Siempre que se mantengan las variables precisión y sensibilidad constantes (o dentro de unos rangos aceptables), disminuir la complejidad para el operador es sinónimo de automatizar el proceso de segmentación, y por tanto también lo es de disminuir la sensibilidad ante errores. Asimismo, la automatización contribuye a disminuir (o en el peor de los casos a acotar) los tiempos totales de segmentación y/o ajuste de parámetros, y ayuda también a estabilizar la repetibilidad de las segmentaciones.

Por todo ello, y dado que los conceptos que requieren los algoritmos evaluados son sólo dos (semilla y umbral), se ha optado por asignar una calificación subjetiva a cada algoritmo para valorar la experiencia de usuario. La calificación consiste en una escala de 1 a 3, representando la dificultad de la interacción requerida. Serán puntuados con 1 aquellos algoritmos con interacciones menos complejas, y con 3 aquellos que entrañen mayor dificultad en la interacción con el operador.

Esta calificación es relativa, en el sentido de que es útil solamente en el contexto de la comparación entre los cuatro algoritmos evaluados en este proyecto. Así, en el caso de que dos algoritmos exhibiesen en la misma cantidad todos y cada uno de los criterios anteriormente expuestos para todas las imágenes empleadas (mismas precisión y sensibilidad , medias, máximas, mínimas, mismos tiempos de ejecución, y mismo espacio paramétrico), la calificación asignada según el criterio de "interacción con operador"

serviría para orientar acerca de la superioridad de uno sobre otro para la tarea de la segmentación de la aurícula izquierda en imágenes por Tomografía Axial Computerizada.