

REPORTE TÉCNICO
**Reconocimiento
de Patrones**

**Evaluación de calidad en la
segmentación de imágenes**

José Luis Gil Rodríguez

RT_042

junio 2011





CENATAV

Centro de Aplicaciones de
Tecnologías de Avanzada
MINISTERIO DE LA INDUSTRIA BÁSICA

RNPS No. 2142
ISSN 2072-6287
Versión Digital

SERIE AZUL

REPORTE TÉCNICO
**Reconocimiento
de Patrones**

**Evaluación de calidad en la
segmentación de imágenes**

José Luis Gil Rodríguez

RT_042

junio 2011



Tabla de contenido

1	Introducción	2
2	Segmentación de imágenes	4
2.1	Definición formal de segmentación de imágenes.....	5
2.2	Niveles de investigación en segmentación de imágenes	5
2.3	Cantidad de algoritmos de segmentación desarrollados.....	6
2.4	Una clasificación de algoritmos de segmentación de imágenes.....	7
2.5	Resumen de estados de arte sobre segmentación de imágenes.....	8
3	Evaluación en segmentación de imágenes	10
4	Métodos de evaluación utilizados en segmentación de imágenes.....	11
4.1	Métodos analíticos	14
4.2	Métodos empíricos no supervisados.....	15
4.2.1	Método de bondad ideal G1, basado en uniformidad intra-región	15
4.2.2	Método de bondad ideal G2, basado en contraste intra-región.....	16
4.2.3	Método de bondad ideal G3, basado en la forma de la región.....	17
4.3	Métodos empíricos supervisados.....	17
4.3.1	Método de diferencias D1, basado en el número de píxeles mal segmentados	18
4.3.2	Método de diferencias D2, basado en la posición de los píxeles mal segmentados	19
4.3.3	Método de diferencias D3, basado en el número de objetos en la imagen	20
4.3.4	Método de diferencias D4, basado en los valores de rasgos en objetos segmentados	21
4.3.5	Método de diferencias D5, basado en cantidades misceláneas.....	22
5	Dos nuevas métricas para evaluar segmentación de imágenes	23
5.1	Métrica RQM: medida de calidad real	23
5.2	Métrica TQM: medida de calidad teórica.....	24
5.3	Resultados experimentales	25
6	Comparación de los grupos de métodos.....	28
6.1	Generalidades para la evaluación	28
6.2	Terminología y taxonomía	29
6.3	Complejidad de la evaluación	30
6.4	Consideraciones sobre aplicaciones de segmentación.....	31
6.5	Evaluación y clasificación.....	32
7	Conclusiones y trabajos futuros	34
7.1	Conclusiones	34
7.2	Tendencias futuras	36
	Referencias bibliográficas	37
	Anexo	
	Glosario de términos	43

Evaluación de calidad en la segmentación de imágenes

José Luis Gil Rodríguez

Dpto. Reconocimiento de Patrones, Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV),
Ciudad de La Habana, Cuba.

jlil@cenatav.co.cu

RT _042, Serie Azul, CENATAV

Aceptado: 30 de diciembre de 2010

Resumen. Este reporte intenta reunir una selección de los últimos resultados del estado de arte en las investigaciones sobre evaluación en la segmentación de imágenes, una de las tareas cruciales del análisis de imágenes es que tiene el objetivo de extraer información (representada por datos) para ser usada en el proceso de segmentación de la imagen. En los últimos 40 años se han estudiado muchos métodos de segmentación, pero todavía no hay ninguna medida de evaluación eficaz y satisfactoria lo que hace difícil comparar métodos de segmentación diferentes, o incluso comparar diferentes parámetros de un sólo método. Un número de trabajos están basados en los principios que ya habían sido establecidos, otros trabajos han hecho modificaciones y mejoras a las técnicas previas, mientras que muy pocos han presentado ideas nuevas. Se ha propuesto un número pequeño de medidas efectivas de las llamadas “auto-suficientes” cuyo objetivo es evaluar diferentes criterios en los objetos, o evaluar el mismo criterio de diferente modo. Tales medidas normalmente trabajan bien en algunos casos, y pobremente en otros. En este sentido, se ofrece un resumen de la tendencia actual a usar la combinación de varias medidas de evaluación que obtienen el rendimiento de la segmentación de diferentes maneras, y sus medidas se combinan usando un “enfoque de aprendizaje automático” (machine learning approach) encargado de colegiar los resultados. Se incluyen en este reporte dos métricas supervisadas que permiten cuantificar la calidad de un resultado de segmentación procedente de cualquier tipo de imagen. La originalidad de estos métodos radica en la posibilidad de evaluar los resultados de la segmentación para todo tipo de imágenes naturales y artificiales con mosaicos de texturas (color, niveles de gris, texturadas o no), y ofrecen datos para comparar dos algoritmos que han segmentado la misma imagen. Los métodos se basan en el criterio de comparar las diferencias inter-regiones (imagen segmentada vs segmentada ideal) considerando las coincidencias pixel a pixel usando análisis de componentes conexas y un algoritmo de optimización combinatoria que soluciona el problema de asignación en un grafo ponderado, binario y completo. Las métricas RQM (medida de calidad real) y TQM (medida de calidad teórica) ofrecen el índice de coincidencia para cada región segmentada, el impacto de cada región segmentada en la segmentación global y la métrica (real/teórica) entre la segmentación obtenida y su segmentación ideal. Los resultados experimentales en imágenes con mosaicos de textura muestran la eficacia de estas técnicas. Finalmente, en las conclusiones se recopilan las tendencias para futuros trabajos de investigación.

Palabras clave: evaluación de calidad en la imagen segmentada, medidas de calidad de la imagen, comparación de diferentes métodos de segmentación, criterios supervisados y no supervisados, medida RQM, medida TQM.

Abstract. This report attempts to bring together a selection of the latest results of state-of-the art research in image segmentation evaluation, one of the most critical tasks of image analysis that has the objective of extracting information (represented by data) from image segmentation process. Many segmentation methods have been studied, but there is still no satisfactory effectiveness

measure which makes it hard to compare different segmentation methods, or even different parameterizations of a single method. A number of works are based on previously proposed principles, and several works have made modifications to and improvements on previous techniques, still other works have presented a few new ideas. A few stand-alone effectiveness measures have been proposed, but these measures examine different fundamental criteria of the objects, or examine the same criteria in a different way, so they usually work well in some cases, but poorly in the others. A co-evaluation framework, in which different effectiveness measures judge the performance of the segmentation in different ways, and their measures are combined by using a machine learning approach which merges the results is summarized. We included in this report two supervised criterion that enable to quantify the quality of a segmentation result from any kind of image. The originality of these methods lie in the possibility to evaluate the segmentation results for all kinds of images including textured ones (color image, gray-level image, textured / not textured image). These methods are based upon a criterion that takes into account the inter-regions disparities (segmented vs ground truth) by considering the coincidences pixel by pixel using connected components analysis and a combinatorial optimization algorithm which solves the assignment problem in a bi-partite weighted graph. The RQM metric (real quality measure) and TQM metric (theoretical quality metric) give us the coincidence index for every segmented region, the impact of the each region on the global segmentation and the coincidence metric (real/theoretical) between obtained segmentation and ground truth. Experimental results show the efficacy of this technique using textured images mosaics. Finally, in the conclusion session the future work research are summarized.

Keywords: segmented image quality evaluation, image quality measures, compare different segmentation methods, supervised and unsupervised criterias, RQM metric, TQM metric.

1 Introducción

El escaneado de imágenes y su procesamiento automatizado nació en 1956 en el “National Bureau of Standards” (NBS, ahora NIST, “National Institute of Standards and Technology”) [105]. En estos 53 años el ámbito de uso de los algoritmos desarrollados ha sido expansivo. El procesamiento digital de imágenes (PDI) ha dado lugar a una multitud de trabajos sobre temas muy variados: mejoramiento de imágenes, segmentación de imágenes, detección de movimiento, compresión, reconstrucción 3D, etc. [87]. Para estos propósitos se han diseñado algoritmos especializados en cada tema. En ellos, se mantiene un problema recurrente: ¿cómo evaluar la eficacia del resultado de un algoritmo de PDI?. Este problema es insoslayable e inevitable en toda publicación científica, o en la automatización de una cadena de procesamiento de imágenes.

Desde los primeros años del PDI, apareció un área de trabajo dedicada al diseño de software con una porción de ella dedicada a la *comprobación de algoritmos*. La comprobación de algoritmos es el proceso que determina si un algoritmo particular satisface ciertas especificaciones, por ejemplo relacionadas con criterios de exactitud y robustez [105]. La exactitud se vincula con la precisión del resultado, lo cual habla de la eficacia del algoritmo. La robustez es la capacidad del algoritmo de enfrentar diferentes complejidades en la imagen con soluciones aceptables. Otro criterio que interesa medirle al algoritmo es la eficiencia, lo que se vincula con el tiempo de cálculo empleado en la tarea. La comprobación de algoritmos se ha extendido a todos los temas mencionados del PDI y por tanto, cubre el mejoramiento y la segmentación de imágenes.

La principal limitación en el diseño de un algoritmo de PDI radica en la dificultad de demostrar que ese algoritmo trabaja con una medida aceptable de rendimiento. La comprobación de algoritmos tiene dos objetivos: 1) Proporcionar criterios cualitativos y cuantitativos que permitan evaluar al algoritmo, y 2) Proporcionar medidas que permitan comparar un algoritmo contra otro similar, asumiendo que los criterios usados son similares [105].

Pero, ¿cómo concebir un criterio para comparar resultados en PDI?. ¿Es posible diseñar criterios de rendimiento que midan sensibilidad, robustez, o exactitud?. Si, es posible y de hecho se han creado

cientos de medidas que evalúan el rendimiento de *algoritmos de mejoramiento de la imagen*, *algoritmos de segmentación de la imagen* y *algoritmos de segmentación de videos*. El centro de atención en este trabajo se ha puesto en la evaluación del rendimiento en imágenes segmentadas.

Los métodos más generales que se han usado en la creación de medidas de evaluación son los siguientes:

- Métodos analíticos para evaluar directamente el rendimiento del algoritmo de segmentación de imágenes.

La idea está centrada en medir y/o conocer las propiedades del algoritmo: considerando principios, requerimientos, utilidades y complejidad del algoritmo. Por ejemplo: es secuencial, paralelo, parte es secuencial+paralelo, tiempo de ejecución, orden de complejidad, el principio del algoritmo (reconocimiento estadístico de patrones, red neuronal, es un método variacional), y otras propiedades que describan al algoritmo sin tener en cuenta la imagen segmentada.

Los métodos analíticos son simples y proporcionan alguna información adicional, pero no pueden obtener todas las propiedades del algoritmo de segmentación. Se usan aisladamente.

- Métodos empíricos para evaluar directamente el resultado de la segmentación de imágenes.

La idea está centrada en evaluar indirectamente el rendimiento de los algoritmos de segmentación. Para ello se han planteado dos principios generales: uno “establece un ideal” de “buena” segmentación (goodness methods, por su designación en inglés); y el otro principio mide exhaustivamente las discrepancias o diferencias entre la imagen segmentada y una imagen idealmente segmentada (discrepancy methods, por su nombre en inglés) [122].

Los métodos empíricos son más realistas que los analíticos porque ofrecen esencialmente datos cuantitativos, a su vez los métodos empíricos “goodness”, que traducimos al español como métodos de “bondad ideal” son menos efectivos que los métodos empíricos de diferencias.

A los criterios de evaluación de algoritmos también se les llama métricas de rendimiento, medidas de rendimiento, o índices de rendimiento [128]. El criterio de evaluación puede ser simple, o compuesto si se combinan varias métricas.

La evaluación del rendimiento de un algoritmo, en el más amplio sentido de la palabra, se refiere a la medida de algún comportamiento del algoritmo como la exactitud-precisión, la robustez-adaptabilidad en imágenes con diferentes complejidades. Medir el rendimiento de un algoritmo permite conocer las características intrínsecas del algoritmo que deben ser enfatizadas, así como la evaluación de sus beneficios y sus limitaciones.

Cada grupo de método tiene sus propias particularidades que los distingue de otros grupos. Cada método también tiene sus propias características para ser identificado. Se considera que un buen método de evaluación de la segmentación no sólo permitiría comparar diferentes enfoques de ese proceso, sino también podría integrarse dentro del sistema de reconocimiento para seleccionar-correr adaptativamente los parámetros apropiados de la segmentación, lo que a su vez podría mejorar la exactitud del reconocimiento. Esto tan sólo es un ideal por alcanzar, ya que aún no se tiene el método ideal de evaluación.

El presente reporte está organizado en ocho secciones. La sección 2 hace un resumen sobre segmentación de imágenes, su definición formal, los niveles en que se ejecutan las investigaciones sobre segmentación de imágenes, se ofrecen datos acerca de la cantidad de algoritmos desarrollados y una clasificación de estos algoritmos para facilitar la realización de evaluaciones globales sistemáticas. Al final de la sección se hace una caracterización sobre varios “estados de arte” que resumen las características de los algoritmos de segmentación más destacados y la tendencia que han seguido. Se expone la estrecha relación que existe entre el desarrollo de los algoritmos dedicados a la segmentación de imágenes y el desarrollo de los algoritmos dedicados a su evaluación cualitativa y cuantitativa.

En la sección 3 trata sobre la producción de publicaciones y estados de arte dedicados a la evaluación de la segmentación de imágenes. Allí se describe la secuencia metodológica más completa

de pasos, para evaluar la segmentación, y las características que tiene una evaluación intra-algoritmo y una evaluación inter-algoritmos.

La sección 4 describe formalmente los métodos publicados para evaluar la segmentación de imágenes. En las sub-secciones 4.1, 4.2 y 4.3 se ofrece una breve descripción de los métodos que pertenecen a los tres grupos. Ellos están ordenados de acuerdo a la clasificación de los métodos dada con anterioridad. La justificación para hacer la clasificación de los métodos en analíticos y empíricos, así como la separación de los empíricos en métodos de diferencias y métodos de “bondad ideal” quedará esclarecida en la discusión comparativa entre grupos de métodos que se hace en la sección 6. En la sección 5 se presentan dos métodos de evaluación de la segmentación, y que fueron desarrollados en el proyecto teórico sobre segmentación de imágenes utilizando descriptores de textura ejecutado en CENATAV [24]. En la sección 7 se exponen las conclusiones de este reporte y las tendencias futuras para las investigaciones vinculadas con la temática de la segmentación de imágenes y su evaluación. La sección 8 incluye un glosario de términos con sus definiciones y el ámbito que cubren.

2 Segmentación de imágenes

La segmentación de la imagen consiste en subdividir la imagen en sus partes constituyentes y extraer aquellas partes de interés (objetos) [120], [128]. La segmentación es un proceso de agrupamiento (clustering, en inglés). Al “clustering” también se le llama aprendizaje no supervisado y es una técnica poderosa en clasificación de patrones. Es un proceso para agrupar dos o más términos basado en algún criterio que se define previamente, y así llegar a formar una colección grande. Con frecuencia, en el contexto de la segmentación de imágenes, el “clustering” se considera como la extensión multidimensional de la técnica de umbralamiento (thresholding, en inglés).

En el contexto de segmentación de imágenes, el clustering es reunir varios píxeles o grupos de píxeles con una propiedad similar y con ellos formar una región más grande [127]. La propiedad similar se refiere al color, o el nivel de gris de los píxeles, o la textura definida por un grupo de píxeles.

La segmentación de imágenes es una tarea crucial en el análisis automático de imágenes, la cual está en el nivel medio de la “Ingeniería de Imágenes” (Fig. 1). La ingeniería de la imagen es una nueva disciplina que abarca todas las técnicas aplicables a la imagen, y consta de tres capas que son: 1) procesamiento de imágenes, 2) análisis de imágenes, y 3) comprensión de imágenes [124]. El nivel de complejidad de los procesos ingenieriles crece de la capa 1 a la capa 3.



Fig. 1. La segmentación de imágenes forma parte de la joven disciplina llamada ingeniería de imágenes, cuya historia data de unos 10 años. La segmentación está ubicada en la capa de análisis de la ingeniería de imágenes. La comprobación de algoritmos se ha concentrado en las capas de procesamiento y análisis.

La percepción visual humana distingue los objetos de interés de manera natural, y ello se ejecuta instantáneamente ante cualquier escena por compleja que sea. De forma análoga, la segmentación automática (machine segmentation), es decir mediante máquinas, ha sido objeto de investigaciones en las últimas 4 décadas. En ese tiempo se ha propuesto un número grande de técnicas de segmentación de la imagen y video y se ha utilizado en varias aplicaciones de medicina, en la industria, la teledetección de la tierra, y otras [25], demostrando la importancia que tiene la segmentación y por tanto, el análisis de imágenes en la vida moderna del hombre.

La historia de la segmentación digital de imágenes usando las computadoras podría remontarse a unos 45 años atrás. En 1965, se introdujo el operador Roberts con el objetivo de detectar los bordes entre las diferentes partes de una imagen, de ahí que también sea llamado “detector de bordes Roberts” [85]. Esta fue la primera segmentación que se hizo. Desde entonces, este campo ha evolucionado muy rápidamente y ha sufrido cambios sustanciales. Se han diseñado varios miles de algoritmos de segmentación y se han incorporado en aplicaciones como las mencionadas. El número de algoritmos aumenta sostenidamente con una proporción de varios centenares por año desde que comenzó el siglo 21.

2.1 Definición formal de segmentación de imágenes

Una definición formal de segmentación de la imagen es la siguiente. Suponga que toda la imagen está representada por R y R_i , $i = 1, 2, \dots, n$ son regiones disjuntas no vacías de R , cumplen las siguientes condiciones (Fu, 1981):

$$\bigcup_{i=1}^n R_i = R; \quad (1)$$

$$\text{Para toda } i \text{ y } j, i \neq j, \text{ se verifica que } R_i \cap R_j = \emptyset; \quad (2)$$

$$\text{Para } i = 1, 2, \dots, n, \text{ se cumple que } P(R_i) = \text{TRUE}; \quad (3)$$

$$\text{Para toda } i \neq j, \text{ se cumple que } P(R_i \cup R_j) = \text{FALSE}; \quad (4)$$

Donde $P(R_i)$ es un predicado de uniformidad para todos los elementos del conjunto R_i y \emptyset representa un conjunto vacío. Además, la siguiente condición también es importante para la segmentación y a menudo se incluye entre las condiciones de la definición formal [15]:

$$\text{Para toda } i = 1, 2, \dots, n, R_i, \text{ es una componente conexa} \quad (5)$$

Cada una de las cinco condiciones arriba mencionadas tienen un significado particular. La condición (1) indica que la unión de las regiones segmentadas deben incluir a todos los píxeles de la imagen. La condición (2) indica que las diferentes regiones segmentadas no pueden interceptarse unas con otras. La condición (3) indica que los píxeles pertenecientes a la misma región i deben tener propiedades similares. La condición (4) indica que los píxeles pertenecientes a dos regiones i y j diferentes, tienen propiedades diferentes. La condición (5) indica que los píxeles pertenecientes a la misma región, como resultado de la segmentación, son componentes conexas [127].

2.2 Niveles de investigación en segmentación de imágenes

En nuestros días, no existe aún una teoría general para la segmentación de imágenes, aunque se ha dedicado mucho esfuerzo en desarrollar las técnicas de segmentación. Lo cierto es que, tradicionalmente el desarrollo de algoritmos de segmentación ha sido un proceso ad hoc, orientado a las aplicaciones donde el algoritmo trabaje bien bajo determinadas condiciones. Como resultado, se han explorado muchas direcciones de investigación y han sido adoptados principios muy diferentes unos de otros, por lo que han aparecido en la literatura una amplia variedad de algoritmos de segmentación con bases teóricas diferentes: estadísticos, redes neuronales, principios variacionales, etc. Ninguno de los

algoritmos de segmentación desarrollados hasta hoy, “son aplicables a todos los tipos de imágenes, y tampoco los diferentes algoritmos son igualmente adecuados para una aplicación particular” [127].

Según Zhang, ([125], [129]), las investigaciones en segmentación de imágenes se están llevando a cabo en tres niveles (Fig. 2). El nivel 1, es básico y cubre el desarrollo de algoritmos de segmentación de imágenes; el nivel 2, centra su atención en la evaluación de los algoritmos desarrollados; y el nivel 3, está dedicado al estudio sistemático de los métodos de evaluación de la segmentación. Este reporte se concentra principalmente en el nivel 2.

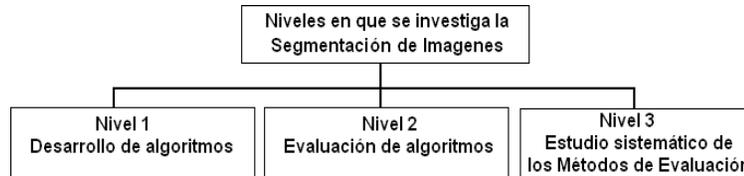


Fig. 2. Las investigaciones en segmentación de imágenes que se han realizado y las actuales en curso, pueden ser agrupadas en tres niveles. Un nivel básico que genera los algoritmos, un nivel intermedio que produce métodos para evaluar su rendimiento y un tercer nivel joven aún, encargado de estudiar con profundidad métodos para seleccionar los mejores algoritmos de segmentación dada la fuente de la imagen y la aplicación que se desea con la segmentación.

2.3 Cantidad de algoritmos de segmentación desarrollados

En los últimos 45 años, las investigaciones y el desarrollo de algoritmos de segmentación de imágenes y videos han seguido un crecimiento muy rápido. Se ha desarrollado un elevado número de algoritmos de segmentación y este número crece continuamente cada año. La Tabla 1 contiene el número de registros encontrados cada 5 años, en la base de datos bibliográfica “EI Compendex” (<http://www.ei.org>) la más abarcadora que existe hoy sobre investigaciones ingenieriles. La búsqueda se hizo sobre el término “image segmentation” para investigar solamente los campos “Subject/Title/Abstract”. Se observa que el número de artículos publicados experimenta un crecimiento muy rápido con un aumento exponencial. Es interesante notar que no se aprecia ninguna señal de disminución del ritmo de aumento.

Tabla 1. Cantidad de registros encontrados en la base de datos bibliográfica “EI Compendex” con el término “image segmentation” [124].

1965	1970	1975	1980	1985	1990	1995	2000	2005	Total
1969	1974	1979	1984	1989	1994	1999	2004	2009	
10	20	233	680	1499	3423	7665	12727	-	26257

La tabla muestra de manera enfática el número de algoritmos de segmentación de imágenes y videos ya desarrollados, dedicados a la descripción de estas técnicas de clasificación. Con ese resumen numérico es fácil percibir una idea general acerca de los progresos alcanzados en las investigaciones y las aplicaciones con segmentación de imágenes, a nivel mundial, en estas 4 décadas.

Debido al incremento del número de algoritmos diferentes para segmentar imágenes, surgió el problema de como evaluar el rendimiento de esos algoritmos, lo que se ha convertido en algo indispensable en el estudio de la segmentación. Esta conclusión queda reforzada conociendo que existen varias modalidades tecnológicas para adquirir diferentes imágenes (por ejemplo, sólo en medicina: TAC, RMN, US, ...) y un número grande de usuarios solicitando aplicaciones de segmentación con esas imágenes. Entonces, el problema de cómo seleccionar el, o los algoritmos apropiados de segmentación, se ha tornado en una tarea de importancia práctica actual. Para ese fin, han sido propuestas un número de técnicas de evaluación de segmentación de imágenes que se describen en el epígrafe 4.

2.4 Una clasificación de algoritmos de segmentación de imágenes

Con tantas publicaciones que han aparecido en la literatura y tantos algoritmos de segmentación de imágenes que se han desarrollado y se siguen desarrollando, se hizo necesario establecer una clasificación que contuviera a todos esos algoritmos y facilitar la realización de estudios sistemáticos sobre ellos.

Hacer una clasificación de los algoritmos en grupos, en principio, es un problema para definir los conjuntos y los sub-conjuntos de dichos algoritmos conociendo la diversidad de bases teóricas. Sin embargo, teniendo en cuenta las condiciones que cumple la definición del concepto segmentación [20], se ha creído que los grupos de algoritmos, después de hacer una clasificación que respete las bases del proceso de segmentación y sus objetivos, deben satisfacer las siguientes cuatro condiciones ([121], [126]):

1. Cada algoritmo debe estar en un grupo,
2. Todos los grupos unidos, deben incluir a todos los algoritmos,
3. Los algoritmos de un mismo grupo deben tener propiedades comunes, y
4. Los algoritmos de grupos diferentes deben tener ciertas propiedades que los diferencie.

Para llevar a vias de hecho las cuatro condiciones anteriores, hay que encontrar “ciertos criterios” que faciliten la clasificación de los algoritmos de segmentación de imágenes. Los dos criterios de clasificación podrían ser: 1) *una propiedad vinculada al pixel segmentado* (discontinuidad, similaridad), y 2) *una propiedad vinculada a la estrategia del segmentador* (paralelo, secuencial). Entonces, los algoritmos quedarían clasificados a partir de las cuatro propiedades siguientes:

- Algoritmos de segmentación basados en bordes (discontinuidad)
- Algoritmos de segmentación basados en regiones (similaridad)
- Algoritmos con programación paralela
- Algoritmos con programación secuencial

Las dos primeros grupos hacen uso de la propiedad del pixel segmentado, según proceda de una segmentación de bordes, o proceda de una segmentación de regiones; que son los dos paradigmas bien conocidos en segmentación de imágenes. Las dos últimos grupos se forman a partir de la estrategia de funcionamiento interno implementada en el algoritmo, según sea programación paralela o secuencial.

Teniendo en cuenta los dos criterios de clasificación escogidos, y las cuatro propiedades a que ellos dan lugar, *todos los algoritmos de segmentación de imágenes* pueden ser clasificados en cuatro grupos, a saber: G1, G2, G3, y G4. Los resultados se muestran en Tabla 2.

Tabla 2. Clasificación general de los algoritmos de segmentación de imágenes [127].

Clasificación	Segmentación Basada en Bordes (discontinuidad)	Segmentación Basada en Regiones (similaridad)
Algoritmo paralelo	G1: Algoritmo paralelo basado en bordes	G3: Algoritmo paralelo basado en regiones
Algoritmo secuencial	G2: Algoritmo secuencial basado en bordes	G4: Algoritmo secuencial basado en regiones

Se observa en la Tabla 2 que los algoritmos satisfacen las cuatro condiciones propuestas por Zhang [121]: Cada algoritmo queda ubicado en un grupo, todos los grupos unidos construyen el universo de algoritmos, los algoritmos de un mismo grupo tienen ciertas propiedades comunes y los algoritmos de grupos diferentes tienen ciertas propiedades que los diferencian unos de otros.

Algunos ejemplos simples y típicos de cada grupo son los siguientes, los cuales pueden ser encontrados fácilmente en la literatura:

- G1: Detector de bordes, operador gradiente, operador de bordes fuzzy, operador SUSAN.
- G2: Modelo de contornos activos ACWE para bordes, Modelo de contornos activos clásico (snake), Modelo de búsqueda en grafos, programación dinámica, ...

- G3: K-means paralelo, Modelo basado en umbrales, clasificación del pixel, agrupamiento en el espacio de rasgos.
- G4: K-means secuencial, Modelo de contornos activos ACWE para regiones, Modelo basado en grafos, red SOM, red ART2, red fuzzy ART, crecimiento de regiones, corte, y mezcla, watersheds morfológico.

2.5 Resumen de estados de arte sobre segmentación de imágenes

Junto con el desarrollo de los algoritmos de segmentación de imágenes, en estos últimos 45 años (1965-2009) se han presentado en la literatura, varios “estados de arte” que resumen los algoritmos de segmentación más destacados ([14], [134], [135], [83], [104], [20], [89], [79], [29], [70], [78], [5], [91], [6], [75], [92]). Todos ellos son estados de arte generales y abarcadores del universo de algoritmos de segmentación de imágenes hasta el año de su publicación. El último estado de arte fue realizado por [127]. Este en particular, ofrece una visión general del desarrollo de la segmentación en imágenes y videos con énfasis en mostrar el número de algoritmos de segmentación que han sido desarrollados, describe la técnica de clasificación de algoritmos presentada en el epígrafe anterior, y analiza las publicaciones “estados de arte” vinculadas con la segmentación de imágenes.

Observando la Tabla 1 acerca de la producción de algoritmos de segmentación, y los años en que se realizaron los estados de arte, se pueden extraer algunas conclusiones. En los primeros 10 años (1965-1974) no fueron publicados ninguno y la razón es que los resultados de las investigaciones se acumularon en esta década. Los trabajos importantes sobre segmentación de este período fueron recogidos en los estados de arte de las siguientes décadas. Todos los estados de arte, están comprendidos en las dos décadas siguientes (1975- 1994), excepto el último. La razón por la que no se hicieron estados de arte en la última década (1995-2004) puede estar relacionada con el hecho de la explosión de algoritmos y técnicas de segmentación que fueron presentadas en ese intervalo, lo que complica la realización de un estudio exhaustivo, detallado y comprensible. Se conoce que, en este mismo período está concentrada la atención en la evaluación y selección de algoritmos de segmentación para hacer aplicaciones.

En los últimos 10 años no se han realizado estados de arte generales que tengan un cubrimiento global e integral en segmentación de imágenes, aunque si han sido presentados algunos estados de arte especializados enfocados:

Estados de arte enfocados al estudio de un grupo particular de algoritmos de segmentación

Muchos algoritmos de segmentación han sido desarrollados usando herramientas teóricas (matemáticas), como son lógica fuzzy, algoritmos genéticos, redes neuronales (NN), reconocimiento estadístico de patrones, y otras teorías; o basados en algún modelo, tal como el modelo ACM (contornos activos), modelo basado en umbrales (thresholding, en inglés), modelo watershed (parteaguas, es español), y otros modelos. Los algoritmos tratados en los estados de arte tienen sus principios basados en esas teorías matemáticas y esos modelos. Los siguientes párrafos presentan algunos ejemplos.

1. *Estudio de un proceso de segmentación interactiva* [72]: Los métodos completamente automáticos en ocasiones funcional mal y producen resultados incorrectos. Teniendo en cuenta esta realidad, se hace necesaria la participación y los conocimientos del operador humano; por ejemplo para identificar los patrones que se están usando en el proceso de segmentación de imágenes médicas. Con la participación del hombre se busca desarrollar criterios cualitativos para evaluar un método de segmentación interactiva. En este estado de arte se estudian técnicas computacionales destinadas a la interacción hombre-máquina en segmentación de imágenes. Este estudio toma en cuenta, el tipo de información proporcionada por el usuario, y cómo esta información afecta a la parte computacional, el propósito de la interacción en el proceso de segmentación y se comparan varios métodos de diálogo hombre-máquina.

2. *Buscando el mejoramiento de algoritmos de segmentación*. [19]: Desarrollan un estado de arte acerca de algoritmos de segmentación que combinan técnicas basadas en bordes y regiones. Este estudio presenta siete estrategias diferentes para fundir tales informaciones.
3. *Estudio de la segmentación con ASM para una aplicación particular*. Behiels et al. [2]: Revisan los algoritmos ASM (Active Shape Model). Este modelo ofrece una manera particular para encontrar la frontera de los objetos en la imagen. Para ello se utilizan varias funciones de energía y rasgos de la imagen. Los autores están interesados en la segmentación de estructuras anatómicas en radiografías.
4. *Estudio de la segmentación con umbrales para una aplicación particular*. Marcello et al., [61]: Desarrollan un estado de arte sobre métodos de segmentación basados en umbrales (thresholding, en inglés). Estas técnicas son muy populares, rápidas y relativamente simples. Los autores tienen el propósito de evaluar su rendimiento en imágenes de teledetección para aplicaciones oceanográficas.

Estados de arte enfocados a una aplicación particular de segmentación de imágenes

La segmentación de imágenes tiene muchas aplicaciones. En cada aplicación particular, podrían ser desarrollados y adaptados varios algoritmos de segmentación. En este sentido se han hecho algunos estados de arte en áreas de aplicación particulares.

1. *Aplicación en medicina* [77]: Desarrollan un estudio sobre los métodos de segmentación automáticos y semi-automáticos de imágenes médicas anatómicas. Son discutidas y comparadas las ventajas y las desventajas de estos métodos para desarrollar aplicaciones con imágenes médicas. En esta área, la segmentación de imágenes se usa para automatizar o facilitar la delineación de estructuras anatómicas y otras regiones de interés.
2. *Aplicación en medicina* [45]: Desarrollan un estudio sobre los algoritmos relacionados con la extracción de vasos en bio-ingeniería, ya que este es esencialmente un proceso de segmentación. Se exponen seis grupos de técnicas para esta aplicación particular.
 - a) Técnicas de reconocimiento estadístico de patrones.
 - b) Métodos basados en modelos.
 - c) Métodos basados en seguimiento (tracking, en inglés).
 - d) Métodos basados en inteligencia artificial.
 - e) Métodos basados en redes neuronales.
 - f) Otros métodos de detección de objetos, por ejemplo tubos.
3. *Aplicación en video* [47]: Desarrollan un estudio sobre técnicas que operan tanto sobre videos comprimidos, como no comprimidos. El video es considerado como un tipo particular de imagen y su segmentación es una extensión de la segmentación de imágenes. La segmentación temporal de los datos de videos se usa para determinar el límite de las tomas. En el estudio se tienen en cuenta las transiciones abruptas y graduales de las tomas. Se discuten exhaustivamente el rendimiento, los méritos relativos y las limitaciones de los métodos existentes.
4. *Aplicación en video* [53]: Revisan la segmentación en tiempo real de secuencias de video no comprimidos para la búsqueda y recuperación basadas en contenido. Dependiendo de la información usada para detectar los cambios de las tomas, se seleccionan los algoritmos: basados en el pixel, el histograma, en bloques, en rasgos y movimiento. Los autores calculan la complejidad, como un criterio que debe ser considerado, especialmente en aplicaciones de segmentación de video en tiempo-real.
5. *Aplicación en video* [80]: Desarrollan un estudio sobre cuatro clases de técnicas diseñadas especialmente para la detección de sombras en movimientos. En muchas aplicaciones de visión, se necesita hacer la detección de sombras en movimiento. El movimiento de la sombra se considera como un objeto en un video, y la detección de sombras es un problema de la segmentación video.

3 Evaluación en segmentación de imágenes

Haciendo una búsqueda en la base de datos bibliográfica “EI Compindex” del término “image segmentation”, y considerando solamente el campo título, produjo una salida de 5000 registros [124]. Esto demuestra la existencia de una elevada producción de algoritmos para segmentar imágenes. Su explicación pudiera estar en el hecho que “no hay una teoría general para desarrollar algoritmos de segmentación de imágenes” y en consecuencia, aparecen algoritmos para dar respuesta a las aplicaciones en los diferentes dominios particulares. Dado que, el número de algoritmos de segmentación de imágenes es creciente (vea además el epígrafe 2.3) la evaluación de la segmentación ha alcanzado una importancia crucial en la comunidad de segmentación de imágenes.

Desde hace 40 años aproximadamente, se han estado haciendo muchos esfuerzos destinados a la evaluación de los algoritmos de segmentación y como resultado de esos esfuerzos acumulados se pueden encontrar en la literatura más de 100 artículos sobre evaluación entre las últimas décadas del siglo pasado y lo que va del presente. A su vez, se han hecho varios estados de arte y estudios sistemáticos intentando caracterizar los métodos de evaluación existentes para imágenes ([109], [121], [116], [120], [122], [123], [125], [127], [128], [95], [105]) y video [129].

La evaluación de la segmentación es una tarea compleja que involucra a varias fases en el proceso. Liu & Dori [58] identificaron un protocolo de evaluación del rendimiento con tres elementos esenciales: adquisición de la verdad terreno (ground-truth), un procedimiento de cotejo, y la definición de métricas cuantitativas. Más tarde Mao & Kanungo [59] definieron un procedimiento de cinco pasos para evaluar cuantitativamente a un algoritmo de segmentación de imágenes, siguiendo una metodología de evaluación empírica. Esos autores propusieron la siguiente metodología:

1. Crear el conjunto de datos de prueba con su verdad terreno (ground-truth);
2. Formulación de las métricas de evaluación;
3. Selección óptima de los parámetros libres del algoritmo de segmentación (si los hubiera);
4. Evaluación del algoritmo de acuerdo a los valores de las métricas; y
5. Cálculo de la significación estadística de la evaluación anterior.

Esta metodología de evaluación, propuesta por Mao & Kanungo [59], parece ser la más completa pero no la única ([35], [43], [41], [42], [38]) y tampoco está universalmente aceptada. En realidad, la evaluación de las técnicas de segmentación ha continuado ganando más y más atención, y se están utilizando numerosas metodologías, además de la anterior. Por ejemplo, Zhang-Hui et al., [113] diseñaron y construyeron un software (framework, en inglés) para la evaluación usando algunas de las medidas existentes. En materia de evaluación tampoco hay una teoría general aceptada y es un tema donde queda trabajo por hacer. Un área de interés creciente es, ¿cómo caracterizar los diferentes métodos que existen para evaluar las técnicas de segmentación? [122]. De hecho, se necesita estudiar cuidadosa y sistemáticamente: los diferentes criterios de evaluación propuestos, su aplicabilidad, procedimientos de uso, ventajas, y limitaciones.

La evaluación de la segmentación de imágenes es un proceso genérico que incluye imágenes a color, en tonalidades de gris, de teledetección, “imágenes de rango” (en inglés, range images) y video. Aunque hay estudios específicamente orientados a estos tipos de imágenes, por ejemplo:

- Estudios dedicados a la evaluación de segmentación en “imágenes de rango” pueden encontrarse en ([33], [63], [36], [37], [39], [40]).
- Estudios dedicados a la evaluación de segmentación en imágenes de teledetección consulte los trabajos de [61], [66].
- Estudios dedicados a la evaluación de segmentación en imágenes a color [4], [56], [57], [16].

Los trabajos de investigación existentes sobre evaluación de segmentación ha sido clasificado en dos niveles:

- a) un nivel vinculado con el desarrollo de métodos para evaluar algoritmos, y

- b) un segundo nivel que se enfoca hacia el estudio sistemático de los métodos de la evaluación.

Este último sólo tiene una historia de aproximadamente 10 años y no son muchos los investigadores involucrados en este perfil [123].

A su vez, la evaluación de algoritmos de segmentación de imágenes (Fig. 3) requiere de técnicas para caracterizar un algoritmo (proceso intra-algoritmo), y técnicas para comparar algoritmos (proceso inter-algoritmos).

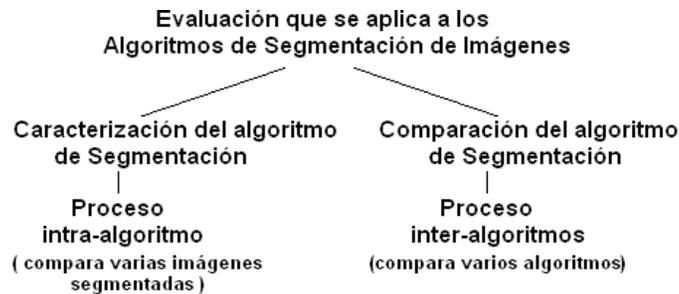


Fig. 3. La evaluación que se aplica a los algoritmos de segmentación de imágenes requiere de técnicas intra-algoritmo e inter-algoritmos.

Proceso de evaluación intra-algoritmo: El propósito de evaluar a un algoritmo de segmentación específico es conocer cuantitativamente su comportamiento ante varias imágenes y ayudar en consecuencia, a definir los parámetros que necesita ese algoritmo ante diferentes aplicaciones buscando obtener el mejor rendimiento.

Proceso de evaluación inter-algoritmo: El propósito de comparar diferentes algoritmos es ranquear sus rendimientos y proporcionar pautas para seleccionar los algoritmos adecuados según las aplicaciones, así como, para promover la creación de nuevos algoritmos tomando en cuenta eficazmente, los puntos fuertes de varios algoritmos.

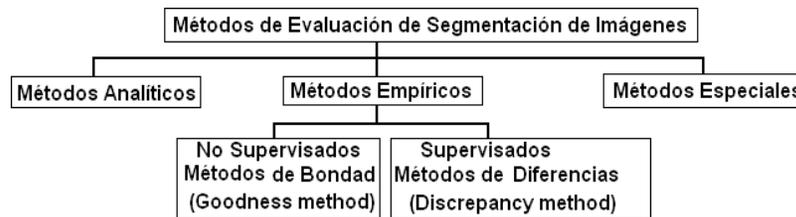


Fig. 4. Los métodos de evaluación de los resultados de la segmentación de imágenes han sido clasificados en métodos analíticos y métodos empíricos. Hay un resto que clasifican como métodos especiales. Los principales métodos de bondad “Goodness = G” y de diferencias (D = Discrepancy) se presentan en secciones siguientes.

4 Métodos de evaluación utilizados en segmentación de imágenes

Los métodos de evaluación de segmentación de imágenes pueden ser clasificados en métodos analíticos y métodos empíricos ([120], [122], [128]) (Fig. 4). Los métodos analíticos tratan los algoritmos de segmentación directamente y están dirigidos a examinar su principio de funcionamiento; mientras que, los métodos empíricos juzgan la imagen segmentada de acuerdo a un criterio predefinido (métodos no supervisados), o comparando con una imagen de referencia (métodos supervisados), para evaluar indirectamente el rendimiento de los algoritmos. Hay un tercer grupo denominado métodos especiales que utilizan varios criterios como, la evaluación visual de los resultados de la segmentación (vea [88]),

cantidad de operaciones de edición, y correlación entre la imagen original y la imagen bi-nivel segmentada [128]. Estos métodos han atraído cierta atención.

La evaluación empírica es prácticamente más eficaz y la más utilizada que la evaluación analítica [120]. Los avances recientes, en materia de evaluación de la segmentación, están hechos principalmente en el desarrollo de las técnicas de evaluación empíricas [128]. Los métodos empíricos se clasifican en: grupo de métodos de bondad (Goodness methods, en inglés), y grupo de métodos de diferencias (Discrepancy methods, en inglés). Estos grupos de métodos usan diferentes criterios empíricos para juzgar el rendimiento de los algoritmos de segmentación.

Los Métodos de Bondad pueden realizar la evaluación sin la ayuda de imágenes de referencia, mientras que los Métodos de Diferencias necesitan alguna imagen de referencia para arbitrar la calidad de la segmentación.

Los métodos para evaluar la segmentación de imágenes producidos en el siglo pasado y en el actual, están recogidos en los estados de arte publicados por [120],[121], [122], [123], [124], [125] [127], [128], [129]. En el estado de arte de Zhang [120], considerado como un estudio exhaustivo, están recogidos los ocho criterios empíricos principales que han sido publicados (tres de bondad y cinco de diferencias). Con los ocho métodos se evalúa empíricamente la imagen segmentada e indirectamente al algoritmo de segmentación. En las Tablas 3 y 4, tomadas de Zhang [122], [128], aparecen los métodos-criterios-medidas clasificados en clases y grupos. En las secciones 4.2 y 4.3 se presentan las descripciones formales de ellos.

Tabla 3. Listado de criterios empíricos y el grupo al cual pertenece el método. Es la Tabla 1 de [122].

Clase	Nombre del criterio-método-medida de evaluación	Grupo del método
G-1	Uniformidad intra-región	Bondad (Goodness)
G-2	Contraste inter-región	Bondad (Goodness)
G-3	Forma de la región	Bondad (Goodness)
D-1	Número de píxeles segmentados mezclados	Diferencias (Discrepancy)
D-2	Posición de los píxeles segmentados mezclados	Diferencias (Discrepancy)
D-3	Número de objetos en la imagen	Diferencias (Discrepancy)
D-4	Valores del rasgo en los objetos segmentados	Diferencias (Discrepancy)
D-5	Cantidades misceláneas	Diferencias (Discrepancy)

Tabla 4. Listado complementario de criterios empíricos y el grupo al cual pertenece el método. Tabla 2 de [128].

Clase	Nombre del criterio	Grupo del método
G-4	Moderado número de regiones	Bondad (Goodness)
D-5a	Consistencia de la región	Diferencias (Discrepancy)
D-5b	Diferencia del nivel de gris	Diferencias (Discrepancy)
D-5c	Divergencia simétrica (entropía cruzada)	Diferencias (Discrepancy)
S-1	Cantidad de operaciones de edición	Especial
S-2	Inspección visual	Como Diferencias
S-3	Correlación entre la imagen original y la imagen segmentada	Como Bondad

En la Fig. 5 de Zhang [120], [122] puede verse claramente la inter-relación que existe entre; algoritmo de segmentación, método de evaluación, y proceso de segmentación de la imagen. En esa figura se muestra un esquema general que vincula orgánicamente a las tres partes del proceso de segmentación: los algoritmos de segmentación (izquierda), los tipos de métodos de evaluación (analíticos, de bondad, de diferencias), y la segmentación de imágenes propiamente (rectángulo central punteado).

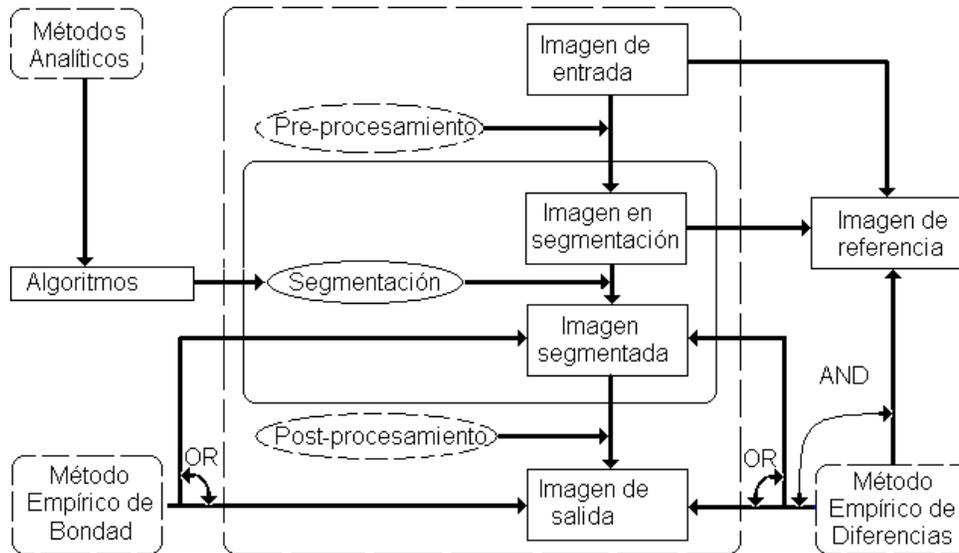


Fig. 5. Esquema general de segmentación y evaluación de imágenes.

El proceso de segmentación incluye: la lectura de la imagen de entrada, alguna operación de pre-procesamiento para mejorar la imagen (opcional), empleo de un algoritmo para ejecutar la segmentación, obtención de la imagen segmentada, alguna operación de post-procesamiento sobre la imagen segmentada (opcional), hasta obtener la imagen de salida. Esta secuencia de operaciones está enmarcada por el rectángulo central punteado y de arriba-abajo. Las flechas de izquierda a derecha indican los momentos en que se aplican las operaciones de pre-procesamiento, la segmentación y la operación de post-segmentación. Las flechas verticales muestran las etapas del proceso de segmentación desde la imagen de entrada, pasando por la imagen en segmentación, la imagen segmentada hasta la imagen de salida. La imagen de salida pudiera ser la misma que la segmentada si sobre ella no se ejecutó alguna operación de post-segmentación. Los procesos posteriores, como son la extracción de rasgos se hace sobre la imagen de salida.

Los puntos de acceso de los tres grupos de métodos de evaluación se muestran en las esquinas de la Fig. 5 con rectángulos punteados. Observe que hay una condición OR entre las dos flechas que apuntan a la “imagen segmentada” y la “imagen de salida”, y que proceden tanto del “método de diferencias empírico” como del “método empírico de bondad”. Por otra parte, hay una condición AND entre la flecha que va desde “método empírico de diferencias” hasta “imagen de referencia” y las dos flechas OR que enlazan a la “imagen segmentada” y la “imagen de salida”. Se aprecia claramente que los métodos analíticos actúan directamente sobre los algoritmos de segmentación. Los métodos empíricos de bondad juzgan la imagen segmentada o la imagen de salida, pero evalúan indirectamente el rendimiento de los algoritmos. Para aplicar los métodos empíricos de diferencias es necesario constar con la imagen de referencia, la cual se puede obtener manual o automáticamente a partir de la imagen de entrada, o a partir de la imagen en segmentación. Los métodos empíricos de diferencias comparan la imagen segmentada o la imagen de salida contra la imagen de referencia y usan sus diferencias para evaluar el rendimiento de los algoritmos de segmentación.

Cada grupo de métodos tiene sus propias particularidades que los distingue de otros grupos. Cada método también tiene sus propias características para ser identificado. En las siguientes tres secciones 4.1, 4.2 y 4.3 se ofrece una breve descripción formal de los métodos que pertenecen a los tres grupos.

Para la evaluación de la segmentación, usando diferentes medidas de las existentes, varios autores han construido sus propias versiones de softwares. Las siguientes publicaciones hacen referencias a algunos de esos software: [102], [113] y [99].

4.1 Métodos analíticos

En inglés, *Analytical methods*. Los métodos analíticos evalúan a los algoritmos de segmentación independientemente de su salida, no trabajan sobre los resultados de la segmentación. Ellos evalúan la efectividad basado completamente en las propiedades y los principios de funcionamiento del algoritmo. En general, los métodos analíticos sólo evalúan ciertas propiedades de los algoritmos de segmentación, tales como su estrategia de procesamiento (paralelo, secuencial, iterativo, o mixto), complejidad del procesamiento, eficiencia del algoritmo, la resolución de la segmentación y alguna otra. Estas propiedades normalmente no son decisivas para diferenciar el rendimiento entre los algoritmos de segmentación. Además, no todas las propiedades del algoritmo pueden ser obtenidas por estudios analíticos. Por consiguiente, los métodos analíticos generalmente no son métodos buenos para evaluar la tarea de segmentación de un sistema de reconocimiento [120].

Cuando se usan los métodos analíticos para evaluar a los algoritmos de segmentación se elimina la influencia causada por los experimentos de evaluación como hacen los métodos empíricos [113].

La dificultad continúa siendo “la carencia de una teoría general para la segmentación de imágenes” [30]. Se han hecho algunos intentos iniciales en la dirección de obtener una teoría unificada sobre segmentación de imágenes, por ejemplo, en relación a los modelos de la imagen y segmentación [90], pero no se ha encontrado ninguna solución formal todavía. Aún después de la desaparición física de Rosenfeld, los métodos analíticos sólo trabajan con algunos modelos particulares o con algunas propiedades de los algoritmos.

Liedtke et al., [50] propusieron un método analítico. Hicieron un estudio evaluativo de varios algoritmos tomando en cuenta el tipo y la cantidad de conocimientos que a priori habían incorporado en los diferentes algoritmos de segmentación. El conocimiento presente en los algoritmos está determinado principalmente, por la naturaleza del algoritmo. Ese conocimiento incorporado a priori, normalmente es información heurística y los diferentes tipos de conocimientos son difícilmente comparables. Entonces, la información que brinda su método analítico es general, desigual y cualitativa.

En ciertos casos el método analítico puede proporcionar información cuantitativa sobre el algoritmo de segmentación. Abdou & Pratt [1] crearon un procedimiento estadístico que implementa una “razón de probabilidad de detección” y con él analizaron el rendimiento de varios detectores de bordes. Sea U el umbral de decisión del borde, P_c es la probabilidad de detección correcta, y P_f es la probabilidad de detección falsa.

$$P_c = \int_U^{\infty} P(t/edge)dt \quad (6)$$

$$P_f = \int_U^{\infty} P(t/no\ edge)dt \quad (7)$$

El gráfico de P_c versus P_f en términos del umbral U puede proporcionar un índice de rendimiento de los detectores de bordes. A criterio de Zhang [120] este índice debe ser útil para evaluar los algoritmos de segmentación basados en la detección de bordes y sugiere ver [75]. Si se compara este índice con el índice basado en conocimientos a priori discutido más arriba, este índice está concebido más preciso y puede ser útil ante detectores del borde simples [1].

Es posible implementar y obtener ciertas propiedades de los algoritmos de segmentación, por ejemplo; la estrategia de procesamiento, complejidad del procesamiento, gasto en tiempo (eficiencia) del algoritmo, etc. ([22],[116]). Tales propiedades pueden ayudar a seleccionar algoritmos adecuados

para aplicaciones particulares. La estrategia de procesamiento de un algoritmo de segmentación puede ser paralela, secuencial, iterativa o mixta. Los algoritmos paralelos son convenientes para generar implementaciones rápidas. Sin embargo, ante imágenes que están severamente contaminadas por la presencia de ruidos, los algoritmos paralelos con frecuencia muestran un rendimiento más pobre que los métodos secuenciales [23].

4.2 Métodos empíricos no supervisados

En inglés, *Empirical Goodness Methods*. La traducción al español no es totalmente precisa. Podría ser “Métodos Empíricos de Bondad”. El vocablo *goodness* desea expresar que son métodos que cumplen con un ideal de calidad. Se conocen como métodos de evaluación no supervisados o métodos autosuficientes. Los métodos que pertenecen a este grupo evalúan el rendimiento de los algoritmos juzgando la calidad de la imagen segmentada. Para llevar a cabo esa evaluación se deben definir medidas de calidad que miden cuantitativamente los resultados de los algoritmos de segmentación de acuerdo a alguna caracterización que el humano ha hecho sobre las propiedades de una segmentación “ideal”. Las medidas “deben” satisfacer las condiciones de esa segmentación ideal (por ejemplo, una imagen linda) [120]. Dicho con otras palabras, la calidad de la imagen segmentada se evalúa mediante alguna medida *goodness* que tiene la cualidad de ser buena. Y es buena, porque responde al ideal predefinido por el hombre, pero ese ideal no es un ground truth. “Estos métodos caracterizan a diferentes algoritmos de segmentación simplemente calculando la medida *goodness* basada en la imagen segmentada, sin el conocimiento a priori de la segmentación correcta [120]. El beneficio de estos métodos es que ... no necesitan ser evaluados contra una imagen de referencia segmentada manualmente [114].

Los métodos de bondad evalúan el rendimiento de los algoritmos indirectamente, juzgando las imágenes segmentadas con ciertas medidas de calidad, establecidas de acuerdo a la ya mencionada intuición humana, y como ese ideal puede ser concebido de muchas maneras, han sido propuestas varias ideas para cubrir diferentes aspectos de un “ideal”, o de una “buena” segmentación. Los métodos o medidas de bondad están clasificados en tres grupos como se muestra en la Tabla 3.

Las ideas fundamentales que se están utilizando en el diseño de los métodos empíricos no supervisados son:

- Uniformidad intra-región ([104], [55], [91], [76], [57], [4].
- Contraste inter-región ([73], [55]).
- Forma de la región [91], y
- Evaluación del borde [110].

Por tratarse de medidas *no supervisadas*, no significa que sean mejores que las *supervisadas*. De hecho, estas últimas son las más precisas y más usadas para evaluar a los algoritmos de segmentación. El beneficio mayor de los métodos de bondad parece ser la funcionalidad no supervisada que ellos tienen, y al estar libres de referencias “ground truth” pueden ser usados en sistemas automatizados de evaluación en tiempo real (on-line). Algunas de estas medidas se han utilizado en el diseño de algoritmos de segmentación.

4.2.1 Método de bondad ideal $G1$, basado en uniformidad intra-región

En inglés, *Goodness Method based on intra-region uniformity*. Para Nazif & Levine [69] una segmentación ideal debe producir una imagen que posea alta uniformidad intra-región, la cual está relacionada con la propiedad de similaridad entre los elementos de la región.

La uniformidad de un rasgo, que pertenece a una región se puede calcular en base a la varianza de ese rasgo evaluada en cada pixel perteneciente a esa región [55]. La medida de uniformidad para una imagen en niveles de gris (GLUM; en inglés, Uniformity Gray Level Measure) se define como:

$$GLUM = \sum_i \sum_{(x,y) \in R_i} \left[I(x,y) - 1/A_i \sum_{(x,y) \in R_i} I(x,y) \right]^2 \quad 8$$

Donde,

$I(x,y)$: Imagen en niveles de gris,
 A_i : Área de R_i ,
 R_i : $i^{\text{ésima}}$ región segmentada,

La medida de uniformidad normalizada (NUM; en inglés, Normalized Uniformity Measure) propuesta por Sahoo et al., [91] tiene la expresión:

$$NUM = 1 - GLU/C \quad (9)$$

Donde,

C : Es un factor de normalización

La uniformidad intra-región de una imagen segmentada es una propiedad deseada porque aporta una cualidad que puede ser cuantificada. Esta propiedad también se puede medir a través de la entropía local de orden superior basado en teoría de la información [74]. Estos dos autores propusieron un método de umbral que maximiza la entropía local de segundo orden de las regiones objeto y su fondo (background). Asumiendo un umbral U, ellos calculan la entropía H^2 , a partir de la siguiente expresión:

$$H^2(U) = - \sum_{i=0}^U \sum_{j=0}^U p_{ij} \ln(p_{ij}) \quad (10)$$

Donde,

p_{ij} : Probabilidad de ocurrencia del par (i, j) dentro del objeto/background.

Pal & Bhandari [76] utilizaron la entropía H^2 como una medida de la homogeneidad de la región en imágenes segmentadas para evaluar el rendimiento de los resultados de segmentación.

4.2.2 Método de bondad ideal G2, basado en contraste intra-región

En inglés, *Goodness Method based on inter-region contrast*. Levine & Nazif [55] consideraron que una segmentación adecuada debe producir imágenes que tienen alto contraste entre regiones adyacentes. Ellos definieron la medida de contraste para una imagen $I(x, y)$ en niveles de gris (GLCM; en inglés, Gray Level Contrast Measure) de la siguiente manera:

$$GLCM = \frac{|\bar{I}_{obj} - \bar{I}_{backg}|}{\bar{I}_{obj} + \bar{I}_{backg}} \quad (11)$$

Donde,

$I(x,y)$: Imagen expresada en niveles de gris, en el punto (x,y),
 \bar{I}_{obj} : Nivel de gris promedio en el objeto,
 \bar{I}_{backg} : Nivel de gris promedio en el background,

Observe que una idea similar ya había sido usada por Otsu [73] en la evaluación de valores de umbral con “buena calidad” (*goodness*) cuando desarrolló su algoritmo de selección del umbral basado en el histograma.

4.2.3 Método de bondad ideal G3, basado en la forma de la región

En inglés, *Goodness Method based on region shape*. Para satisfacer la intuición humana de una segmentación “ideal” no solamente se han utilizado los “niveles de gris de la imagen” para el diseño de medidas de calidad, sino que también se ha usado la “forma de la región segmentada”. Sahoo et al., [91] propusieron una medida de la forma en la región (RSM; en inglés, Region Shape Measure) para evaluar varios algoritmos de selección de umbral, que definieron con la expresión siguiente:

$$RSM = \frac{1}{C} \left\{ \sum_{(x,y)} Sgn[I(x,y) - \bar{I}_{N(x,y)}] g(x,y) Sgn[I(x,y) - U] \right\} \quad (12)$$

Donde,

$Sgn[.]$: Función paso unitario,

$I(x,y)$: Imagen expresada en niveles de gris, en el punto (x,y) ,

$\bar{I}_{N(x,y)}$: Valor promedio de los niveles de gris, para la vecindad $N(x,y)$, centrado en el pixel (x,y) ,

$g(x,y)$: Valor gradiente en el pixel $I(x,y)$,

U : Valor de umbral seleccionado para la segmentación,

C : Factor de normalización.

4.3 Métodos empíricos supervisados

En inglés, *Empirical Discrepancy Methods*. Los métodos de diferencia empírica son conocidos como métodos de evaluación supervisada. Estos métodos evalúan los algoritmos de segmentación comparando la imagen segmentada resultante contra una imagen de referencia que ha sido segmentada manualmente y que a menudo se le llama “la norma de oro” [27]. Los métodos de diferencia empírica tienen la ventaja, por encima de los métodos de bondad empírica, que ellos hacen una comparación directa entre la imagen segmentada y la imagen de referencia, por lo cual se cree que ellos proporcionan una evaluación de resolución más fina y por tales motivos, son los métodos de evaluación objetiva que normalmente se usan. Sin embargo, la generación manual de una imagen de referencia es un trabajo difícil, subjetivo y altamente consumidor de tiempo. Y para la mayoría de las imágenes, especialmente las imágenes naturales, generalmente no se puede garantizar que una imagen segmentada manualmente sea mejor que otra. Además, tal método de evaluación no se puede usar dentro de un algoritmo de segmentación con el objetivo de encontrar los mejores parámetros de un algoritmo de segmentación parametrizado [113].

En aplicaciones prácticas de segmentación, se pueden tolerar algunos errores en la imagen segmentada. Por otro lado, si la imagen que se está segmentando es compleja y el algoritmo es totalmente automático, el error es inevitable [27].

La diferencia entre la imagen segmentada y la imagen idealmente segmentada, que es la referencia del mejor resultado esperado, se puede utilizar para evaluar el rendimiento de los algoritmos. Tanto la imagen segmentada como la imagen referencia, se obtienen a partir de la imagen original de entrada. En los casos en que las imágenes de prueba sean imágenes sintetizadas, las imágenes de referencia se pueden obtener a partir de un procedimiento de generación [130]; mientras que, cuando son imágenes reales se utilizan imágenes de referencia construidas manualmente con la ayuda de inspección visual.

Los métodos de evaluación de este grupo toman en cuenta las diferencias, que se les miden a varios parámetros, entre la imagen segmentada y la imagen de referencia. Estos métodos tratan de determinar cuán lejos está la imagen segmentada de la imagen de referencia. Un valor más alto de la medida de diferencias implicaría un error más grande en la imagen segmentada relativa a la imagen de referencia, y esto indica que el algoritmo de segmentación aplicado tiene un rendimiento más bajo.

El parámetro o medida, que habitualmente se utiliza en la evaluación del proceso de codificación de imágenes, es la diferencia entre la imagen original y la imagen decodificada. Allí la medida de diferencia que se aplica es el valor medio cuadrático de la Relación Señal Ruido (SNR) [26]. Sin embargo, la codificación de la imagen es un “procesamiento de la imagen”, y la segmentación de la imagen es un “proceso de análisis”, donde las imágenes de entrada y salida pertenecen a dominios diferentes y contienen materias diferentes. No obstante, para evaluar la segmentación de imágenes han sido propuestas y usadas otras medidas de diferencias.

Los experimentos comparativos demuestran que los métodos de diferencias son más efectivos que los métodos de bondad. Los cinco grupos de métodos de diferencias están relacionados en la Tabla 3.

4.3.1 Método de diferencias D1, basado en el número de píxeles mal segmentados

En inglés, *Discrepancy based on the number of mis-segmented pixels*. Considerando la segmentación de imágenes como un proceso de clasificación de píxeles, el porcentaje de píxeles mal clasificados es la medida de las diferencias que rápidamente nos viene a la mente [108]. Suponga que una imagen posee N clases de píxeles, y que es posible construir la matriz de confusión C de dimensión N, donde cada entrada C_{ij} representa el número píxeles de la “clase j”, clasificados en la “clase i” por el algoritmo de segmentación. Entonces, se pueden calcular dos tipos de errores para cada clase k, los cuales se utilizan para describir el rendimiento clase-por-clase del algoritmo [108]. El Error Tipo I para la clase k se define como $E_{\text{tipo I}}^{(k)}$:

$$E_{\text{tipo I}}^{(k)} = 100 \times \left[\left(\sum_{i=1}^N C_{ik} \right) - C_{kk} \right] / \left[\sum_{i=1}^N C_{ik} \right] \quad (13)$$

Donde, el numerador representa el número de píxeles de la clase k, no clasificados como k; y el denominador es el número total de píxeles de la clase k.

El Error Tipo II para la clase k se define como $E_{\text{tipo II}}^{(k)}$:

$$E_{\text{tipo II}}^{(k)} = 100 \times \left[\left(\sum_{i=1}^N C_{ki} \right) - C_{kk} \right] / \left[\left(\sum_{i=1}^N \sum_{j=1}^N C_{ij} \right) - \sum_{i=1}^N C_{ik} \right] \quad (14)$$

Donde, el numerador representa el número de píxeles de otras clases llamada clase k. El denominador es el número total de píxeles de otras clases. En las ecuaciones (13) y (14), cada clase está igualmente pesada.

Weszka and Rosenfeld [104] usaron un enfoque similar para medir las diferencias entre una “imagen ideal” (correcta) y una imagen umbralizada. Bajo la suposición que la imagen está formada de objetos y background, cada uno con una distribución específica de niveles de gris, los autores calcularon para cualquier valor de umbral, la probabilidad de clasificar mal un objeto como background, o viceversa. Esta probabilidad proporciona un índice de los resultados de la segmentación que pueden usarse por evaluar algoritmos de selección de umbral. En su trabajo, semejante probabilidad se minimiza en el proceso de seleccionar un umbral apropiado.

Otra medida de diferencias es la Probabilidad del Error (PE) y se basa en el mismo principio ya definido. Para un problema de dos clases, la medida PE se calcula como sigue [52]:

$$PE = P(O) \times P(B \parallel O) + P(B) \times P(O \parallel B) \quad (15)$$

Donde,

- $P(B \parallel O)$: es la probabilidad del error en la clasificación de objetos como background,
- $P(O \parallel B)$: es la probabilidad del error en la clasificación de background como objetos,
- $P(O)$: es la probabilidad a priori de los objetos en la imagen,
- $P(B)$: es la probabilidad a priori del background en la imagen.

Para un problema multi-clases los autores Lim & Lee [56], ofrecen una definición general de PE.

La evaluación de detección de bordes ha utilizado la idea de calcular diferencias basadas en el número de píxeles error. En este sentido Fram & Deutsch [18] utilizaron el método estadístico MLE (en inglés, Maximum Likelihood Estimation) del mejor ajuste de un modelo matemático a ciertos datos. Lo aplicaron para obtener la fracción de bordes detectados correctamente. Este esquema pudiera ser extendido para obtener que fracción de píxeles segmentados es incorrecta y aplicarlo en la evaluación de segmentación de imágenes.

4.3.2 Método de diferencias D2, basado en la posición de los píxeles mal segmentados

En inglés, *Discrepancy based on the position of mis-segmented pixels*. Las medidas de diferencias basadas sólo en el número de píxeles mal segmentados no tienen en cuenta la información espacial de estos píxeles. Entonces, es posible que imágenes segmentadas de manera diferente pudieran tener el mismo valor de evaluación, si las medidas sólo cuentan el número de píxeles mal segmentados. Para superar este problema, se han propuesto algunas medidas basadas en diferencias que tienen en cuenta la posición espacial en que ocurre el error.

Una manera es usar la distancia entre el píxel mal segmentado y el píxel más cercano que en realidad pertenece a la clase mal segmentada. Sea N , el número de píxeles mal segmentados en toda la imagen, y $d(i)$ es una distancia medida desde el $i^{\text{ésimo}}$ píxel mal segmentado y el píxel más cercano que en realidad está en la clase mal segmentada. La medida de diferencias (DM) basada en estas distancias [108] la definieron como sigue:

$$DM = \sum_{i=1}^N d^2(i) \quad (16)$$

Donde,

- N : Número de píxeles mal segmentados en toda la imagen,
- DM : Medida de distancias.

En la ecuación 16, cada distancia está elevada al cuadrado. Los autores normalizan la medida de las diferencias (NDM), para eliminar la influencia del tamaño de la imagen y darle un rango de valores adecuado.

$$NDM = 100 \times \sqrt{DM}/A \quad (17)$$

Donde,

- A : Número de total píxeles en la imagen (área),

NDM: Medida de distancias normalizada utilizada en segmentación de regiones.

Para evaluar detectores de bordes (segmentación de bordes) una medida de diferencias que se utiliza con frecuencia es la distancia media cuadrática propuesta por Pratt [81]. Se le llama también “Figure of Merit”:

$$FOM = \frac{1}{N} \sum_{i=1}^N \frac{1}{1 + p x d^2(i)} \quad (18)$$

Donde,

$$N = \max(N_{pbi}, N_{pbs}),$$

N_{pbi} : Número de píxeles del borde ideal (pbi),

N_{pbs} : Número de píxeles del borde segmentado (pbs),

$d(i)$: Distancia entre la posición del $i^{\text{ésimo}}$ píxel del borde segmentado y su posición ideal,

p : Parámetro de escala.

Strasters & Gerbrands [93] usaron FOM para evaluar resultados de la segmentación, pero considerando que N es el número de píxeles en la imagen, y $d(i)$ denota la distancia entre el píxel $i^{\text{ésimo}}$ y su clase correcta. Además, ellos definieron una versión modificada de *FOM* nombrada FOM_e para extender el rango de valores de *FOM* próximos a la segmentación perfecta:

$$FOM_e = \begin{cases} \frac{1}{N_e} \sum_{i=1}^{N_e} \frac{1}{1 + p x d^2(i)} & \text{if } N_e > 0 \\ 1 & \text{if } N_e = 0, \end{cases} \quad (19)$$

Donde,

N_e : denota el número de píxeles mal segmentados.

4.3.3 Método de diferencias $D3$, basado en el número de objetos en la imagen

En inglés, *Discrepancy based on the number of objects in the image*. Para obtener una segmentación perfecta, es necesaria la siguiente condición: que coincidan el número de objetos de cada clase cuando se comparan las imágenes de referencia y segmentada. Una sustancial no-coincidencia del número de objetos, indica una diferencia grande entre las imágenes de referencia y segmentada. Yasnoff & Bacus [109] propusieron calcular la “cantidad de objetos coincidentes” basados en la teoría de las probabilidades. Para medir la cantidad de objetos coincidentes, los autores usan la probabilidad F_{OCA} que tienen los dos números S_i y R_i de representar muestras de la misma distribución.

$$F_{OCA} = \int_L^{\infty} \frac{1}{2^{M/2} \Gamma(M/2)} Z^{(M-2)/2} e^{-Z/2} dz \quad (20)$$

Donde,

F_{OCA} : Medida basada en la cantidad de objetos coincidentes entre la imagen referencia y la imagen segmentada. OCA: en inglés, Object-Count-Agreement.

$M = N - 1$: denota el número de grados de libertad,

$\Gamma(\cdot)$: denota la función Gamma, y

L : se puede calcular por la ecuación que sigue,

$$L = \sum_{i=1}^N \frac{S_i - R_i}{p \times R_i} \quad (21)$$

Donde,

- S_i : Número de objetos de la clase i , en la imagen segmentada,
- R_i : Número de objetos de la clase i , en la imagen de referencia,
- N : Número de clases,
- p : es un parámetro de correlación.

Sobre la base de una idea similar [93] propusieron otro esquema pesado de fragmentación que llamaron *FRAG* y lo definieron de la siguiente manera:

$$FRAG = \frac{1}{1 + p \times |T_N - A_N|^q} \quad (22)$$

Donde,

- FRAG*: Medida de fragmentación basada en la cantidad de objetos en las imágenes de referencia y segmentada,
- T_N : Número de objetos en la imagen de referencia,
- A_N : Número de objetos en la imagen segmentada,
- p, q : parámetros de escala.

4.3.4 Método de diferencias *D4*, basado en los valores de rasgos en objetos segmentados

En inglés, *Discrepancy based on the feature values of segmented objects*. El análisis de la imagen está relacionado con la extracción de información de la imagen; una imagen contiene datos que pueden ser salidas del proceso de análisis [111]. Aquí los datos de salida se obtienen de valores medidos sobre los objetos que se encuentran en las imágenes segmentadas. En este sentido, una pregunta fundamental en análisis de la imagen es, si una medición hecha sobre los objetos de las imágenes segmentadas es tan precisa como la medición hecha en las imágenes originales. De acuerdo a esta medición, la imagen segmentada tendrá la mayor calidad si los rasgos extraídos a los objetos, machean de forma precisa con los rasgos de la imagen original. En práctica, la imagen segmentada tiene alta calidad si la medición hecha en ella permanece inalterable, respecto a la tomada en la imagen original [13].

El objetivo final del proceso de segmentación de imágenes, visto en el contexto del análisis de imágenes, es obtener mediciones de rasgos sobre los objetos segmentados [131]. La exactitud de las mediciones obtenidas en la imagen segmentada, con respecto a la imagen de referencia, ofrece una medida de las diferencias que es útil para evaluar calidad. A esa exactitud se le llama “Última Exactitud Medida” (en inglés, *UMA*: ultimate measurement accuracy) para resaltar el objetivo final de la segmentación. El valor “Absoluto” de *UMA* se denota como *AUMA* y su valor “Relativo” se denota por *RUMA*. Ambos fueron definidos por Zhang & Gerbrands [132] como sigue:

$$AUMA_f = |R_f - S_f| \quad (23)$$

$$RUMA = \frac{|R_f - S_f|}{R_f} \times 100\% \quad (24)$$

Donde,

$AUMA_f$: Valor “Absoluto” de la métrica UMA,

$RUMA_f$: Valor “Relativo” de la métrica UMA,

R_f : Valor del rasgo obtenido a partir de la imagen “Referencia”,

S_f : Valor del rasgo obtenido a partir de la imagen “Segmentada”.

Las expresiones de $AUMA_f$ y $RUMA_f$ son genéricas y pueden representar diferentes rasgos medidos sobre los objetos segmentados. El valor de esas métricas representa medidas de diferencias entre las imágenes segmentada y de referencia. Los rasgos que pueden ser medidos en los objetos son densitometría (densidad óptica del material) y rasgos geométricos o estáticos. Algunos ejemplos de rasgos geométricos de los objetos son: área, perímetro, esfericidad, excentricidad, factor de forma, curvatura absoluta media normalizada, etc. [119]. De todos ellos, el área de los objetos es el más adecuado para estimar la calidad de imágenes diferentemente segmentadas [132], [119].

4.3.5 Método de diferencias D5, basado en cantidades misceláneas

En inglés, *Discrepancy based on miscellaneous quantities*. Hay otras medidas de diferencias que pueden describir las discrepancias entre la imagen de referencia y la imagen segmentada. La medida propuesta por [54] es una medida de distancia 2-D (bi-dimensional) basada en dos componentes. Una componente da la “medida del error de sub-mezcla” y la otra componente da la “medida del error de sobre-mezcla”. La primera de las componentes es proporcional a la cantidad para la cual, las regiones en la imagen segmentada, solapan a las regiones de la imagen de referencia. La segunda componente significa la cantidad para la cual, las regiones segmentadas dividen a las regiones de referencia. Es obvio que encontrar esas cantidades es una tarea difícil y dependiente del algoritmo de segmentación.

Para describir las diferencias entre la imagen segmentada y la imagen de referencia, se puede usar tanto la información sobre los niveles de gris, como la información espacial. Strasters & Gerbrands [93] definieron una “figura de certeza” (FOC; en inglés, *figure of certainty*) para tener en cuenta esas informaciones. La medida FOC se define como sigue:

$$FOC = \frac{1}{N} \sum_{i=1}^N \frac{1}{1 + p + q |f_i - g_i|^q} \quad (25)$$

Donde,

FOC : Medida denominada figura de certeza (certidumbre),

f_i : nivel de gris del $i^{\text{ésimo}}$ pixel en la imagen de referencia,

g_i : nivel de gris representativo de la región que comprende al $i^{\text{ésimo}}$ pixel en la imagen segmentada,

N : Número total de píxeles en la imagen

p, q : parámetros de escala.

(Nota: ambas imágenes son tomadas como máscaras para extraer f_i y g_i de la imagen a segmentar)

Si se consideran ambas imágenes, la segmentada y la de referencia, como distribuciones de probabilidad, las diferencias entre ellas podrían ser reflejadas por su divergencia.

Suponga que la imagen segmentada tiene N regiones y p'_i representa la probabilidad a posteriori de un pixel que está en la $i^{\text{ésima}}$ región; mientras que p''_i está en la imagen de referencia. Pal & Bhandari [76] propusieron la métrica “divergencia simétrica” (SD ; en inglés, symmetric divergence) con la expresión la siguiente:

$$SD = \sum_{i=1}^N (p'_i - p''_i) \ln \frac{p'_i}{p''_i} \quad (26)$$

Donde,

SD : Métrica de Divergencia Simétrica,

N : Número de regiones segmentadas,

p'_i : representa la probabilidad a posteriori que un pixel pertenezca a la $i^{\text{ésima}}$ región segmentada,

p''_i : representa la probabilidad a posteriori que un pixel pertenezca a la $i^{\text{ésima}}$ región de referencia,

Conociendo el número N de regiones segmentadas, las probabilidades p'_i y p''_i en las imágenes segmentada y de referencia respectivamente, se calcula la métrica de “Divergencia Simétrica SD ” como la medida del rendimiento para algoritmos de segmentación.

5 Dos nuevas métricas para evaluar segmentación de imágenes

A continuación se proponen dos nuevas métricas para evaluar la calidad en la segmentación de imágenes. Para medir cuantitativamente las diferencias que existen, pixel a pixel, entre dos imágenes segmentadas se concibieron e implementaron dos nuevas métricas que nombramos “Medida de Calidad Real (RQM)” y “Medida de Calidad Teórica (TQM)”. Ambas métricas pertenecen a la categoría de métricas supervisadas, porque toman en consideración el conocimiento que brinda la segmentación ideal ó ground truth para evaluar por comparación las imágenes segmentadas. Entonces, se trata de métodos objetivos de evaluación de algoritmos de segmentación.

La originalidad de estos métodos radica en su capacidad para evaluar cuantitativamente los resultados de la segmentación ante imágenes naturales, imágenes sintetizadas artificialmente, imágenes homogéneas o texturadas, con mosaicos de texturas (color, niveles de gris, texturadas o no), y ofrecen datos para comparar, clase por clase, algoritmos que han segmentado la misma imagen. Los métodos se basan en el criterio de comparar las diferencias intra-regiones (imagen segmentada vs segmentada ideal) considerando las coincidencias pixel a pixel usando análisis de componentes conexas y el método Húngaro. Este es un algoritmo de optimización combinatoria que soluciona el problema de asignación en un grafo binario, ponderado y completo. El tiempo de ejecución es del orden $\log(n)$.

5.1 Métrica RQM: medida de calidad real

La Métrica RQM es una medida supervisada cuantitativa destinada a conocer la coincidencia real que existe entre la imagen segmentada y su versión idealmente segmentada, lo que se mide en una escala comprendida entre [0..1]. La métrica permite calcular la calidad real obtenida por el método de

segmentación sobre esa imagen. Esta medida fue concebida para comparar dos imágenes segmentadas, de la siguiente manera [100]:

Sean, $S_i = \{S_1, S_2, \dots, S_m\}$: la colección de regiones en la imagen segmentada por un algoritmo,
 $I_j = \{I_1, I_2, \dots, I_n\}$: la colección de regiones idealmente segmentadas o imagen ground truth,

El problema de encontrar cual S_i corresponde a cual I_j constituye un problema de apareamiento de máximo costo entre las regiones de S y las regiones de I . Ese es el problema que debe ser resuelto, y para ello se construye un grafo bipartito, ponderado y completo con las colecciones de regiones S e I .

1. Un grafo bipartito ponderado, significa: a) Asignar a cada región, tanto de S_i como de I_j , un nodo del grafo (grafo bipartito). b) Trazar una arista entre los nodos de S_i e I_j , con peso C_{ij} (grafo ponderado).
2. Si $m \neq n$, se le agregan “falsas regiones de tamaño cero” a la colección de menor cantidad de nodos.
3. Se aplica el algoritmo de Munkres también llamado algoritmo Húngaro o [68] para encontrar el apareamiento de costo máximo entre ambas colecciones.

Una vez pareadas las regiones de S_i e I_j , entonces se calculan C_j y d_j con las expresiones siguientes.

$C_j = \frac{|S_i \cap I_j|}{|S_i \cup I_j|}$, es el índice de coincidencia para cada región teniendo en cuenta que S_i e I_j representan a la misma región.

$d_j = \frac{|S_j \cap I_j|}{N}$, es el impacto real de la región j en la segmentación global; y N representa el número total de píxeles en la imagen.

Se define la medida de calidad real RQM (en inglés, Real Quality Measure) como sigue:

$$RQM = \sum_{j=1}^n d_j C_j$$

El valor numérico de RQM representa la medida de coincidencia total entre las regiones S e I . Vea que este valor está construido acumulando el aporte de cada región S_j que coincide exactamente, esto es coincidencia real, con la región idealmente segmentada. No es difícil apreciar que $0 \leq RQM \leq 1$, y para $RQM = 1$, se obtiene la coincidencia perfecta entre S e I .

La medida RQM es posible aplicarla entre dos imágenes segmentadas cualquiera, cuya condición es que las regiones estén representadas por etiquetas de valor entero, lo cual permite construir y comparar sus componentes conectadas. Por supuesto, que los resultados de la medida RQM son independientes del método que lleva a cabo la segmentación de la imagen, siendo esta una vía para evaluar indirectamente el rendimiento de un algoritmo de segmentación de imágenes siempre que se conozca la segmentación ideal (ground truth) de la imagen fuente.

5.2 Métrica TQM: medida de calidad teórica

La Métrica TQM es una medida supervisada cuantitativa destinada a conocer la coincidencia ideal o teórica que existe entre la imagen segmentada y su versión idealmente segmentada, lo que se mide en una escala comprendida entre [0..1]. Esta medida se define a partir de una modificación, simple pero conceptual, de la Métrica RQM.

Se mantienen con igual significado:

$S_i = \{S_1, S_2, \dots, S_m\}$: la colección de regiones en la imagen segmentada,

$I_j = \{I_1, I_2, \dots, I_n\}$: la colección de regiones en su imagen ground truth,

$C_j = \frac{|S_j \cap I_j|}{|S_j \cup I_j|}$, es el índice de coincidencia para cada región teniendo en cuenta que S_i e I_j representan a la misma región.

Se modifica la forma de calcular el impacto que aporta cada región segmentada en la segmentación total. Se considera sólo el “área ground truth de la región j ”, y no el área común de j entre las imágenes segmentada y de referencia). Ahora el impacto de la región j se calcula mediante la expresión:

$d_j = \frac{|I_j|}{N}$, impacto ideal de la región I_j en la segmentación global; y N sigue siendo el número total de píxeles en la imagen.

Análogamente, se define la medida de calidad ideal TQM (en inglés, Theoretical Quality Measure) como sigue:

$$TQM = \sum_{j=1}^n d_j C_j$$

Observe que, ahora el término d_j pondera al índice de coincidencia de la región j por la fracción teórica máxima, que aporta esa región a la segmentación total, y no como en la medida RQM, donde la ponderación se hace por la fracción $\frac{|S_j \cap I_j|}{N}$ que depende del área de coincidencia común entre la región S_j y la región I_j . En el momento en que se construye la segmentación ideal o teórica de la región I_j , automáticamente queda predeterminada la fracción $\frac{|I_j|}{N}$ que le da el peso específico a la región j en la segmentación total.

El valor numérico de TQM es una medida de la coincidencia total entre las regiones S e I , es decir, entre las regiones S_j segmentadas por el algoritmo y las regiones I_j segmentadas por el interpretador humano. Vea que la medida está construida acumulando el área común de cada región C_j , ponderada por la fracción teórica máxima d_j que le corresponde a cada una de las j regiones. Cuando las áreas de las regiones S_j e I_j coinciden perfectamente, entonces $TQM = 1$, mientras que en ausencia de coincidencia $TQM = 0$. Por ello, siempre se cumple que $0 \leq TQM \leq 1$.

Análogamente a la medida RQM, la medida TQM, es posible aplicarla entre dos imágenes segmentadas cualquiera, cuya condición es que las regiones estén representadas por etiquetas de valor entero, lo cual permite construir y comparar sus componentes conectadas. Los resultados de la medida TQM son independientes del método que lleva a cabo la segmentación, y por tanto es posible utilizarla para evaluar indirectamente el rendimiento de un algoritmo de segmentación de imágenes siempre que se conozca la segmentación ideal de la imagen fuente.

5.3 Resultados experimentales

La Fig. 6 muestra un ejemplo de aplicación de las métricas RQM y TQM usando una imagen en tonos de gris sintetizada por la Universidad de Bon que contiene una mezcla de cinco texturas. El ground truth fue construido manualmente por este autor. El método de segmentación que fue evaluado se denomina Método acM basado en el modelo variacional de los contornos activos sin bordes combinado con conjuntos de nivel [100]. El método acM recibió a la entrada un vector de seis descriptores de textura procedentes de tres modelos teóricos de textura. Cada modelo se aplicó sobre la imagen original Fig. 6 a) y se generaron las descripciones de los cinco tipos de textura que contiene el mosaico.

- Modelo Estadística de Primer Orden: produjo los descriptores b) “valor mínimo”, y “c) “valor de la varianza” (deviation). Se utilizó una ventana de operación de tamaño 7x7 píxeles.
- Modelo Estadística de Segundo Orden del Espectro de Textura: produjo los descriptores d) “micro-estructura horizontal”, y e) “micro-estructura vertical”. Se utilizó una ventana de operación de tamaño 15x15 píxeles y un delta de similitud de 5 unidades.

- Modelo Espectral Banco de filtros 2D Gabor de Energía: fueron seleccionados los rasgos de energía obtenidos en la orientación 22.5° azimutal con los valores de frecuencia central 0 y 32 ciclos/ ancho-imagen.

Los parámetros de funcionalidad interna del método acM fueron los siguientes:

- $Miu=1$; peso para ajustar la variabilidad del funcional. El valor 1 ofrece la máxima capacidad de ajuste del contorno a los objetos presentes,
- $classNumber=5$; número de clases a segmentar,
- $ccFilterSize=1500$ pixeles; tamaño de la menor componente conectada que se permite. Aquellas de menor área son filtradas aplicando un filtro de componentes conexas,
- $iterNum=30$; número de iteraciones del funcional.

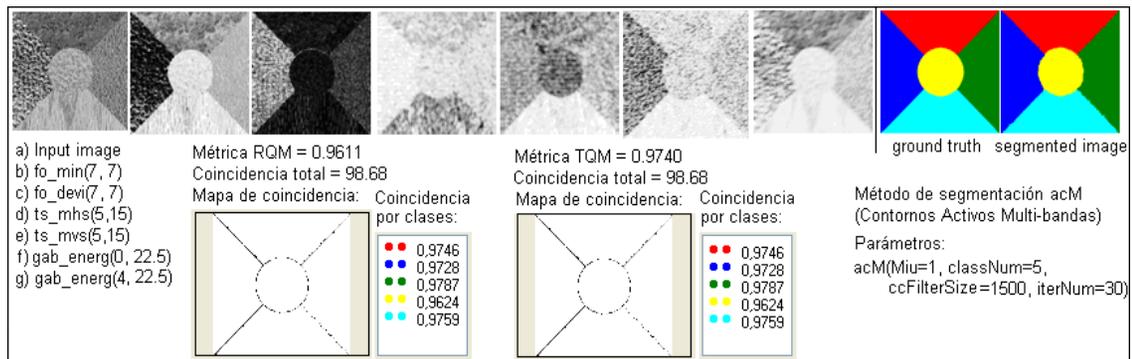


Fig. 6. Segmentación del mosaico de cinco texturas de la Universidad de Bon y su evaluación cuantitativa mediante las métricas RQM y TQM. Observe que se entrega además, el valor numérico de coincidencia por clases de textura. Los errores de segmentación se producen en las fronteras entre diferentes tipos de textura.

La Fig. 6 también muestra el valor de coincidencia de la métrica RQM (0.9611), y TQM (0.9740), los mapas de coincidencias (errores en la segmentación con puntos negros), el % de coincidencia total y las coincidencias entre las clases segmentadas vs clases ideales. La métrica RQM es ligeramente menor que la métrica TQM, esta última es más exigente que el valor realmente alcanzado. La proximidad del valor numérico de RQM a TQM es una medida de alta coincidencia entre las imágenes segmentada y su ground truth. Cuando la coincidencia es perfecta ambas medidas son iguales ($RQM=TQM=1$).

La Fig. 7 muestra la comparación gráfica y cuantitativa de varios métodos de segmentación con la misma imagen de entrada. La imagen que se analiza consiste de un mosaico de texturas desafiante y que se utiliza con frecuencia en la literatura para probar la eficacia de algoritmos de segmentación de textura. Los métodos que se comparan son:

1. Método de segmentación de textura basado en la integración de regiones y bordes [98]. La entrada del método son rasgos de textura extraídos con un banco de filtros de Gabor. La imagen segmentada que obtiene, comparada con su imagen de referencia, da una medida $RQM=0.902$.
2. Método de segmentación de textura basado en agrupamiento Mean Shift [28]. La entrada del método es un conjunto reducido de rasgos de textura extraídos con un banco de filtros de Gabor. La imagen segmentada que obtiene, comparada con su imagen de referencia, da una medida $RQM=0.861$.
3. Método acM para segmentación de textura basado en el modelo variacional de los contornos activos sin bordes combinado con conjuntos de nivel y un filtro de componentes conexas incluido en cada etapa de minimización del funcional [100]. El método recibió a la entrada un vector de descriptores de textura procedentes de varios modelos de textura: incluyendo rasgos obtenidos con un banco de filtros de Gabor 2D de energía y otros rasgos extraídos con modelos estadísticos de segundo orden. La imagen segmentada que obtiene nuestro método, comparada con su imagen de referencia, da una medida $RQM=0.897$.

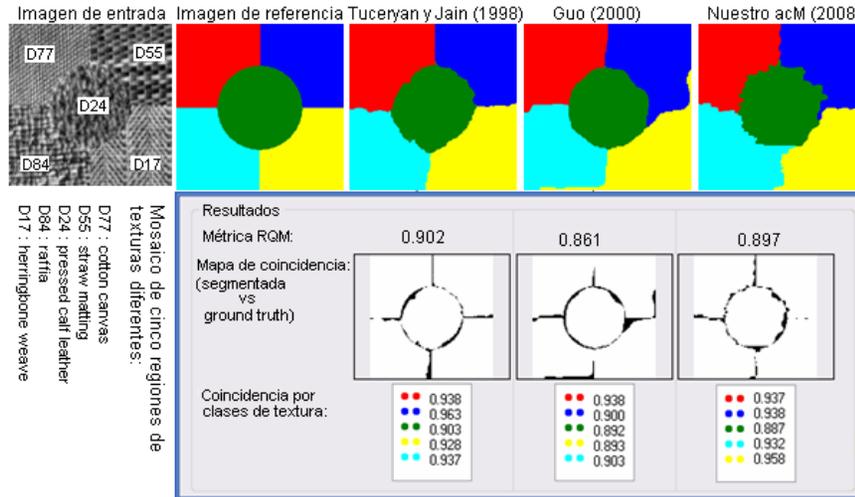


Fig. 7. Muestra varios resultados de segmentación de textura con distintos algoritmos sobre la misma imagen. Las comparaciones cuantitativas entre los algoritmos se hacen usando la métrica RQM. Tuceryan y Jain [96] usaron un método de segmentación que integra bordes y regiones con un banco de filtros de Gabor (RQM=0.902). Guo [28] usa un método Mean Shift con un banco de filtros de Wavelets de Gabor (RQM=0.861) y nuestro método basado en contornos activos sin bordes con un banco de filtros de Gabor (RQM=0.897) ofrece un resultado comparable en eficacia con los dos anteriores.

La eficacia de los métodos 1 y 3 es comparable entre si y del orden de RQM=0.90. Es de notar dos aspectos: a) la complejidad que presenta el mosaico de textura exige a los tres métodos la descripción cuantitativa de la textura mediante bancos de filtros de Gabor con varias frecuencias y direcciones azimutales para obtener descripciones de la textura con alto poder discriminativo; b) la Métrica de Calidad Real (RQM) facilita comparar la eficacia de varios métodos de segmentación cuando se enfrentan a la misma imagen.

La Fig. 8 muestra la comparación gráfica y cuantitativa de la segmentación de varias imágenes, usando el mismo método de segmentación. En este experimento se usan 5 clases diferentes de texturas para conocer el comportamiento del método de segmentación ante ellas. Observe que la primera es una foto tomada de la realidad natural, mientras que las restantes son imágenes sintetizadas artificialmente con mezclas de texturas. Todas las imágenes de este experimento poseen dos texturas, una incrustada en el interior de la otra. Aquí aplicamos nuestro algoritmo de segmentación que llamamos Método acB, destinado a segmentar dos clases de textura (bi-class). El Método acB es el precursor del Método acM, cuando este no había sido extendido para segmentar n-clases. Se basa en contornos activos sin bordes, conjuntos de nivel, una etapa de filtrado digital de componentes conexas en cada etapa de minimización del funcional y un conjunto de n-descriptores de textura. Con el método de segmentación se obtienen las componentes conexas correspondientes al objeto y su background y con ellas es posible aislar al objeto incrustado. Este efecto, la combinación de rasgos de textura extraídos de diferentes modelos teóricos que se utilizó en cada proceso de segmentación y el valor cuantitativo de la métrica RQM se puede apreciar.

Las métricas de evaluación de imágenes segmentadas RQM y TQM tienen la propiedad deseable de su generalidad para ser aplicadas en el estudio de las propiedades de diferentes algoritmos de segmentación, tanto aquellos que segmentan usando tonos, descriptores de textura como tonos + textura. Ambos son métodos empíricos de diferencias, supervisados, objetivos y se utilizan principalmente para estudiar el rendimiento de los algoritmos de segmentación teniendo en cuenta la eficacia de los resultados en la segmentación. Permiten comparar varios métodos de segmentación sobre una imagen natural, o mosaico de textura; y estudiar método ante numerosas imágenes.

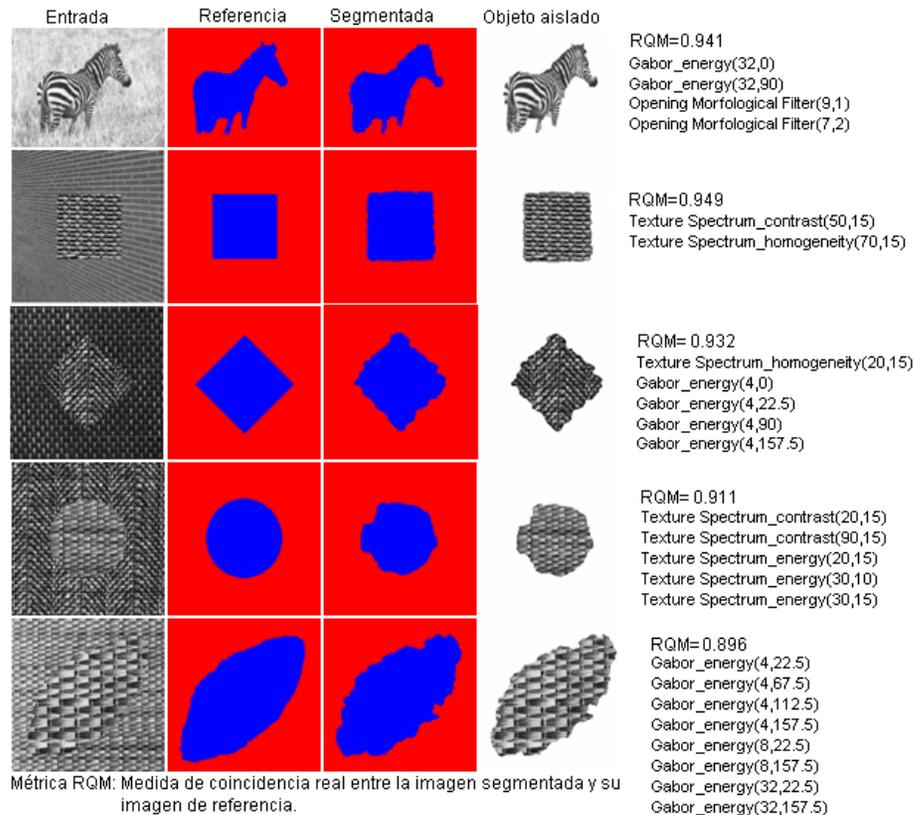


Fig. 1. Muestra la evaluación del algoritmo de segmentación de textura acB mediante nuestra métrica RQM ante distintos tipos de textura naturales y artificiales.

6 Comparación de los grupos de métodos

Los tres grupos de métodos para evaluar la segmentación de imágenes (análíticos, empíricos no supervisados (bondad) y empíricos supervisados (diferencias) que se describieron en la sección 4 tienen sus propias características y también ventajas y limitaciones. Sobre esos aspectos se discute en este epígrafe.

6.1 Generalidades para la evaluación

Una propiedad deseable en un método de evaluación es su generalidad para ser aplicado en el estudio de varias propiedades de diferentes algoritmos de segmentación. Para aplicar los métodos analíticos primero debe ser definido algún modelo formal. El comportamiento del algoritmo sobre una imagen puede analizarse matemáticamente en términos de la información que se tiene de la imagen y los parámetros del algoritmo [46]. Hay ciertas propiedades de los algoritmos de segmentación que se pueden obtener fácilmente mediante análisis, como son, la estrategia de procesamiento del algoritmo y la resolución del resultado de la segmentación [116]. Sin embargo, otras propiedades no pueden ser analizadas dado que no existe un modelo formal que las extraiga. Por ejemplo, no hay una medida cuantitativa que pueda incorporarse en los algoritmos de segmentación capaz de ofrecer a priori, conocimientos sobre la imagen [50]. Aún cuando existiera la forma de extraer diferentes tipos de conocimientos, estos son difícilmente comparables. Además, hay métodos que pueden ser aplicables sólo a ciertos algoritmos de segmentación. Por ejemplo, el método basado en la probabilidad de detección únicamente es adecuado para estudiar los detectores de bordes simples [1].

Los métodos empíricos no supervisados y supervisados, descritos en las secciones 4.2 y 4.3, se utilizan principalmente para estudiar el rendimiento de los algoritmos de segmentación teniendo en cuenta la exactitud de los resultados en la segmentación. Una razón es que hay otras propiedades de los algoritmos, como es el costo computacional, que ha sido parcialmente beneficiado por los progresos de la tecnología. Otra razón es que la exactitud de la segmentación, con mucha frecuencia, es la principal preocupación en las aplicaciones reales y es difícil estudiarla con los métodos analíticos. Teniendo en cuenta que los métodos empíricos solamente estudian una propiedad, puede interpretarse como algo limitados. Sin embargo, muchos de ellos pueden ser considerados como relativamente generales, porque son capaces de evaluar diferentes tipos de algoritmos de segmentación [120]. Los estudios que están contenidos en las referencias ([69], [75], [117], [118], [126], [127]) son algunos ejemplos en los cuales muy diferentes tipos de algoritmos han sido tratados con los métodos empíricos usados. En muchos estudios empíricos, sólo se necesitó la imagen a segmentar y la segmentada, sin importar que tipo de algoritmo se utilizó. Hay algunas excepciones cuando los métodos de segmentación están basados en medidas de forma [104], [91] dado que es necesario entregar un valor de umbral para calcular las medidas de evaluación en esos tipos de algoritmos.

6.2 Terminología y taxonomía

A continuación, se describen algunos pares de términos que se usan en la evaluación de la segmentación. Primero se explica el significado de los términos para establecer las bases de la discusión posterior.

Evaluación intra-técnica vs evaluación inter-técnica

Los procesos de evaluación de la segmentación pueden ser clasificados en dos tipos [121]:

- 1) *Evaluación intra-técnica*: caracteriza el comportamiento de un algoritmo de segmentación particular con diferentes parámetros y escenarios, cuyo propósito es obtener el mejor rendimiento ante varias imágenes.
- 2) *Evaluación inter-técnica*: ordena jerárquicamente el rendimiento de varios algoritmos de segmentación, cuyo propósito es recomendar el algoritmo más conveniente para una aplicación. La evaluación inter-técnica también puede ayudar a promover nuevos desarrollos tomando los aspectos fuertes de varios algoritmos y con ello mejorar el algoritmo precedente [132].

Método Goodness (de bondad) vs métodos de diferencias

Los métodos de evaluación empírica pueden ser clasificados en dos clases según el requisito sobre las imágenes de referencia.

- 1) *Métodos Goodness*: Pueden realizar la evaluación sin ayuda imágenes de referencia. También se les llama métodos stand-alone (específicos para...). La evaluación stand-alone es sensible al tipo de aplicación, al tipo de imagen o el contenido del video. Conduce a una evaluación más cualitativa [12] que cuantitativa.
- 2) *Métodos de Diferencias*: Necesitan de imágenes de referencia para arbitrar la calidad de la segmentación. También se les llama métodos de referencia o métodos de discrepancia. Conduce a una evaluación cuantitativa [124].

Criterio subjetivo vs criterio objetivo

En la evaluación de la segmentación es esencial y crucial, el criterio que se utiliza en el método de evaluación para juzgar el rendimiento de la segmentación. Al criterio de evaluación, también se llama métrica, métrica de rendimiento, medida, índice, y otros nombres. Los métodos empíricos goodness (de bondad) usan criterio subjetivo que refleja algunas propiedades deseables de la imagen segmentada,

mientras que los métodos empíricos de diferencia usan criterio objetivo que indica la diferencia entre la imagen segmentada y la imagen de referencia [120].

Criterio global vs criterio local

Entre los criterios objetivos, algunos de ellos son globales, los cuales no toman en consideración las diferencias locales de las regiones segmentadas. El valor de tales criterios estará determinado principalmente por un comportamiento promedio, y el efecto de una diferencia sustancial local será cancelado por un comportamiento comparativamente bueno en alguna otra parte. El criterio local tiene las características inversas [67].

Métrica simple vs métrica compuesta

En diferentes procesos de evaluación se han usado, o una métrica simple, o varias métricas. Cuando se han utilizado varias métricas, con frecuencia se ha formado una métrica compuesta. Como el comportamiento de un algoritmo está determinado por muchos factores, cuando se aplica una sola métrica, o incluso una métrica compuesta buscando obtener una evaluación integral (completa), es difícil que sea alcanzada una solución óptima porque la combinación de métricas diferentes con frecuencia es demasiado empírico para que sea eficaz. El hecho de usar varias métricas podría cubrir mejor varios aspectos del algoritmo, pero al final todavía se necesita más trabajo en esta dirección [124]. Llegar a obtener un modelo teórico que integre el efecto de varias métricas simples, podría ser más efectivo que la combinación empírica de varias medidas simples.

Evaluación de la segmentación de imágenes vs evaluación de la segmentación de video

Las imágenes de video tienen una dimensión más (el tiempo), que las imágenes estáticas 2D. La evaluación del proceso de segmentación de imágenes se realiza solamente en el dominio del espacio, mientras que la evaluación de la segmentación de video debe ser realizado en los dominios espacial y temporal. Las métricas diseñadas para medir la calidad en la segmentación de video pueden ser convertidas en métricas para medir la calidad en la segmentación de imágenes mediante la eliminación de las porciones relacionadas con el movimiento y el tiempo [124].

Cualitativo vs cuantitativo y subjetivo vs objetivo

Dos propiedades que son deseadas en un método de evaluación de algoritmos de segmentación de imágenes son la capacidad para evaluar en forma cuantitativa y sobre una base objetiva. El estudio cuantitativo de un algoritmo puede proporcionar resultados precisos que reflejan el rigor de la evaluación [132]. El estudio objetivo excluye la influencia del factor humano y proporciona consistencia y resultados sin sesgo [111]. Generalmente, los métodos analíticos están más disponibles para ser aplicados, pero ellos solamente proporcionan propiedades cualitativas de los algoritmos. Sin embargo, los métodos empíricos normalmente son cuantitativos de modo que, los valores de las medidas de calidad pueden ser calculadas numéricamente. En este grupo están los métodos goodness (de bondad) que se basan en medidas subjetivas de calidad y por ello son menos adecuados para realizar una evaluación objetiva de los algoritmos de segmentación. Los métodos de diferencias son considerados “el patrón oro” de la evaluación de algoritmos de segmentación porque pueden ser objetivos y cuantitativos [27].

6.3 Complejidad de la evaluación

La aplicación de los tres grupos de métodos, para evaluar la segmentación de imágenes, crece en complejidad progresivamente. Los métodos empíricos de evaluación normalmente son más complejos que los métodos analíticos, debido a que es necesario implementar algoritmos con sus especificidades, y además realizar un esfuerzo extra para segmentar las imágenes de prueba y calcular los valores de la medida de calidad. El costo computacional de los diferentes métodos empíricos está determinado por el

algoritmo de la medida de calidad que ellos usan. Por ejemplo, la coincidencia en la cantidad de objetos se obtiene fácilmente, mientras que la “medida de uniformidad” y la “medida de forma” necesitan más cálculos.

Entre los métodos empíricos, los métodos goodness son menos complicados que los métodos de diferencias para su aplicación. Por esta razón, los métodos goodness pueden ser usados para evaluaciones en línea [55]. Los métodos de diferencias tienen asociado un requerimiento particular para su aplicación y es, la imagen de referencia. Muchos estudios usan imágenes reales como imágenes de prueba, y obtienen las imágenes de referencias segmentándolas manualmente, por ejemplo, vea las citas [52], [88]. En la siguiente dirección está disponible la base de datos de imágenes naturales (Berkeley Segmentation Dataset)

<http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/BSDS300/html/dataset/images.html> donde, cada imagen ha sido segmentada manualmente por varias personas usando su intuición de los objetos de interés.

La aplicación de los métodos de diferencias incrementa grandemente la complejidad del proceso de evaluación de segmentación de imágenes ([120]), dada la mayor complejidad de los algoritmos (respecto a los goodness) y la necesidad de la imagen de referencia.

Respecto a las imágenes empleadas en los estudios de evaluación de algoritmos, se han estado usando imágenes reales que responden a dominios de tareas con interés particular, y por ello los resultados de esos estudios de evaluación pueden no ser apropiados para otros dominios de aplicaciones. Haralick [31] propuso que una posible alternativa efectiva es usar imágenes construidas sintéticamente. En actualizaciones posteriores se han usado imágenes extraídas de bases de datos de imágenes sintéticas con contenidos uniforme, mezclado (heterogéneo) y con texturas [88]. Los dos problemas asociados a las imágenes reales (su ground truth, dominio particular de aplicación) pueden ser superados usando imágenes sintéticas bien diseñadas. Otras ventajas de las imágenes sintéticas incluyen que son fácilmente controladas y que pueden ser reproducidas por todos los usuarios ([130], [132]).

6.4 Consideraciones sobre aplicaciones de segmentación

El adecuado uso de conocimientos extraídos del dominio de aplicación puede ayudar a obtener procesos confiables y eficaces en visión por computadora, como los refiere [94]. En este mismo sentido, para evaluar de manera efectiva un algoritmo de segmentación, es importante tener en cuenta conocimientos sobre la aplicación de segmentación para entonces aplicar un algoritmo encargado de evaluar la calidad de la segmentación.

Los tres grupos de métodos mencionados más arriba, son diferentes entre sí hasta el punto, que ellos consideran explícitamente la aplicación en que se ha usado el algoritmo de segmentación. En uno de los extremos están los métodos analíticos que no consideran para nada, la naturaleza y el objetivo de la aplicación. Los resultados de estos métodos dependen solamente del análisis de los algoritmos de segmentación en sí mismos.

Los métodos empíricos goodness (de bondad), en los que algunas propiedades deseables de las imágenes segmentadas se cuantifican mediante medidas de bondad, empiezan a apuntar hacia el problema de la aplicación como la opción en que la medida de bondad debe estar relacionada con el objetivo de la aplicación.

Los métodos empíricos de diferencias, los cuales consideran tanto las imágenes: segmentada-resultado como la de referencia-objetivo, intentan obtener la eficacia de la aplicación mediante medidas que cuantifican las diferencias entre ellas. En estos métodos, la necesidad de tener una referencia obliga a la evaluación a estar conectada con la aplicación real [27].

6.5 Evaluación y clasificación

La evaluación de la segmentación es una tarea compleja que involucra cierto número de etapas en el proceso y tiene muchos aspectos. Por ejemplo, la metodología propuesta por [59] (vea epígrafe 3) es un procedimiento de cinco pasos para evaluar cuantitativamente a un algoritmo de segmentación usando el método empírico. Similarmente, Liu [58] identificó tres elementos esenciales en un protocolo de evaluación del rendimiento del algoritmo : adquisición del ground truth, el procedimiento de macheo y la definición de las métricas cuantitativas.

Zhang [125] expone que, en una metodología de evaluación global es esencial aplicar un criterio de evaluación que sea significativo y cuantitativo (calculable dice él). Se basó en los trabajos de evaluación de segmentación en imágenes y videos especialmente desde el año 2000, y los resumió en las siguientes tablas, teniendo en cuenta solamente el criterio empírico que es objetivo y cuantitativo. Para que las tablas sean más comprensibles puede consultar los criterios listados de Zhang [122] incluidos en las Tablas 3 y 4. De este modo, las tablas que siguen ofrecen una “imagen” general sobre la evaluación de la segmentación en los últimos 10 años.

La Tabla 5 ofrece una lista de trabajos de evaluación usando las técnicas existentes. Muchos trabajos están basados en el “criterio de diferencias”, en cuyo caso el criterio perteneciente a la clase D-1 aparece más frecuentemente que otros. Pocos trabajos usaron el criterio goodness (de bondad), mientras que el criterio G-3 no fue usado en ninguno de los trabajos revisados. Sin embargo, las dos nuevas clases G-4 y D-5a añadidas en la Tabla 4, se utilizaron en algunos trabajos.

Table 5. Una lista de trabajos de evaluación usando las técnicas existentes. Es la Tabla 2 de Zhang [124].

Método #	Fuente	Criterio usado	Método #	Fuente	Criterio usado
M-1	[33]	D-5a	M-10	[34]	D-1, D-4
M-2	[121]	D-4	M-11	[8]	D-1, D-2
M-3	[4]	G-1, G-2, G-4	M-13	[80]	D-1
M-4	[106]	S-3	M-13	[86]	D-1
M-5	[9]	D-5a	M-14	[51]	G-1
M-6	[107]	D-1	M-15	[61]	S-2
M-7	[62]	D-4	M-16	[82]	D-1, D-4
M-8	[84]	G-1, G-2	M-17	[7]	D-1, D-3
M-9	[3]	D-1	M-18	[48]	D-1, S-1

La Tabla 6 ofrece una lista de trabajos de evaluación con modificación de las técnicas existentes. La mayoría de los trabajos están basados en el criterio de diferencias, en particular de la clase D-1. Varios trabajos hicieron uso de la combinación de criterios pertenecientes a diferentes clases. La clase S-1 recientemente definida también ha sido usada a través de dos trabajos.

Tabla 6. Una lista de trabajos de evaluación con criterios modificados. Es la Tabla 3 de Zhang [124].

Método #	Fuente	Criterio usado
M-19	[71]	D-1 (ROC, curva de FP vs. FN)
M-20	[10]	D-1 (con extensión espacial y temporal)
M-21	[21]	D-1 (ROC, curva de FP vs. FN)
M-22	[97]	D-1, S-1 (eficacia)
M-23	[49]	D-1 y D-2, (macheo de contornos, consistencia temporal), S-1
M-24	[123]	G-1, G-2, G-4 (usando entropía de la región)
M-25	[16]	G-1, G-2 (con extensión al color, movimiento, histogramas a color)
M-26	[65]	D-1 (ROC, curva de TP vs. FP)
M-27	[98]	D-1 (DOC, curva de TP vs. FP)
M-28	[44]	(PDR, razón de detección modificada)

La Tabla 7 ofrece una lista de trabajos de evaluación con algunas novedades en las métricas usadas. Los tres primeros trabajos consideran el problema de la combinación de diferentes criterios.

Tabla 7. Una lista de trabajos de evaluación con novedades. Es la Tabla 4 de Zhang [124].

Método #	Fuente	Novedad
M-29	[17]	Busca el "Pareto front" en un espacio de competencia multi-dimensional
M-30	[49]	Busca el "Pareto front" en un espacio de competencia 4-D
M-31	[123]	Usa mayoría pesada (WM), Bayesiano y SVM (support vector machine)
M-32	[11]	Usa métrica de relevancia contextual para machear el sistema visual humano (HVS)
M-33	[15]	Realiza la evaluación en diferentes niveles semánticos

Para comparar diferentes métodos de evaluación de segmentación, hay que considerar las técnicas y las medidas usadas en la evaluación. En este sentido se analizan los siguientes cuatro factores ([116], [122]):

1. Generalidad de la evaluación;
2. Subjetivo versus Objetivo y Cualitativo versus Cuantitativo;
3. Complejidad de la evaluación;
4. Exigencia de la evaluación en el uso de imágenes de la referencia.

Entre los cuatro factores descritos, algunos de ellos están relacionados a los grupos de métodos. Por ejemplo, la mayoría de los métodos empíricos proporcionan resultados cuantitativos. Por otra parte, subjetivo versus objetivo y la consideración de la aplicación de la segmentación están estrechamente relacionadas y pueden ser determinadas según el criterio pertenezca al grupo goodness (bondad), o al grupo de diferencias. Dado que sólo se están comparando métodos empíricos aquí, el foco de atención se pone en la generalidad y la complejidad de la evaluación. Los resultados obtenidos en tal comparación para los métodos relacionados en las Tablas 5 y 6 están dados en la Tabla 8.

Tabla 8. Comparación de los métodos reportados en la Tala 5 y Tabla 6. Es la Tabla 5 de Zhang [124].

Método #	Generalidad	Complejidad	Método #	Generalidad	Complejidad
M-1	General	Media	M-15	General	Alta (Humanos) ⁶
M-2	General	Media	M-16	General	Media/Alta
M-3	Numerosos objetos ¹	Media/Alta	M-17	Numerosos objetos ¹	Baja/Media
M-4	Estructura de árbol ²	Alta	M-18	General	Alta (Humanos) ⁶
M-5	Particular ³	Media	M-19	General	Mediia
M-6	General	Media/	M-20	General	Media/Alta
M-7	General	Baja/Media	M-21	General	Media
M-8	General	Media/Alta	M-22	General	Media
M-9	General	Media	M-23	General	Alta(Humanos) ⁶
M-10	General	Media	M-24	Numerosos objetos ¹	Media
M-11	Video ⁴	Media/Alta	M-25	Video ⁴	Alta
M-12	General	Alta	M-26	General	Media
M-13	Video ⁴	Media	M-27	General	Media
M-14	Umbralamiento ⁵	Media	M-28	Video ⁴	Media

Notas:

1. Más apropiado para el tratamiento de imágenes compuestas de regiones con numerosos objetos;
2. Solamente aplicable en algoritmos de segmentación con estructura de árbol;

3. *Para condiciones en que las cinco categorías de regiones definidas por los autores pudieran ser determinadas adecuadamente;*
4. *Sólo apropiada para evaluar la segmentación de video (principalmente porque el factor temporal es crucial);*
5. *Sólo adecuada para evaluar la técnica thresholding;*
6. *Hay factores visuales humanos que están involucrados, la complejidad se torna alta.*

7 Conclusiones y trabajos futuros

Este reporte de investigación contiene un amplio resumen de los trabajos sobre evaluación de calidad en la segmentación de imágenes. Se presenta y se describe un esquema general de los procesos de segmentación y su evaluación, los criterios de evaluación, se ha resumido una taxonomía de los métodos de evaluación, la expresión formal de las métricas consideradas importantes y se presenta un glosario con la terminología empleada en la temática de evaluación de la segmentación de imágenes.

La investigación que se ha expuesto recoge los principales resultados de todos los estados de arte que se han hecho en la temática de evaluación en segmentación de imágenes. Los más significativos son los ejecutados en la década de los 90, después del año 2000 hasta el 2006, y especialmente hasta el 2009. En este período apareció el libro de Zhang [125] donde se describen los avances experimentados en la segmentación de imágenes y videos a partir de las contribuciones de muchos autores a lo largo de unos 20 capítulos. Entre los años 2001-2006 se han presentado más de 20 trabajos sobre evaluación en segmentación de imágenes. Esto indica claramente que se ha puesto mucho más esfuerzo en la evaluación de la segmentación en los inicios del siglo 21 que en el siglo pasado.

Son claves los estados de arte de Zhang [120], [121], [122], [123], [124], [127], [128]. Este autor de la Universidad de Tsinghua, Beijing, China es considerado una autoridad internacional en la evaluación de la segmentación de imágenes. Presentó un resumen de la materia en la reunión de 2005 organizada por Christophe Rosenberger (Universidad de Orleans, Francia) sobre la evaluación del procesamiento en los sistemas de análisis de imágenes. El objetivo de los participantes, en aquella reunión fue presentar los diferentes enfoques de evaluación existentes para cuantificar la calidad en el procesamiento de imágenes. En el último de los trabajos de Yu-Jin Zhang [128] publicado por IGI Global, y denominado “evaluación de la segmentación de imágenes en este siglo” discute acerca de las técnicas de evaluación empíricas publicadas después del 2002. A partir de la revisión bibliográfica realizada fue posible destacar algunas conclusiones y trabajos futuros.

Como una contribución del autor a esta temática se han recogido en el reporte, dos medidas de evaluación de calidad denominadas RQM (medida de calidad real) y TQM (medida de calidad teórica) junto con algunos experimentos realizados donde son utilizadas para evaluar el resultado de segmentación de imágenes aplicando propiedades de la textura expresadas en términos cuantitativos haciendo uso de colecciones de descriptores de textura pertenecientes a diferentes modelos teóricos establecidos en la literatura.

7.1 Conclusiones

La evaluación de la segmentación es un proceso que se utiliza para juzgar el rendimiento de los algoritmos de segmentación. La evaluación se basa en algunos criterios de calidad previamente definidos y/o, en la verdad terreno (ground truth), con vistas a evaluar o revelar las propiedades de los algoritmos de segmentación. La meta ideal que anima a los investigadores es la siguiente: “un buen método de evaluación de segmentación de imágenes no sólo debe permitir la comparación de diferentes enfoques, sino también podría integrarse dentro del sistema de reconocimiento para seleccionar adaptativamente los parámetros apropiados de la segmentación, lo que a su vez podría mejorar la exactitud del reconocimiento [125]”. Este método no se ha alcanzado aún.

En la evaluación de la segmentación se destacan los puntos siguientes:

1. La segmentación de imágenes es el primer paso del análisis de imágenes. A su vez, el resultado del análisis de imágenes es altamente dependiente de la calidad de la segmentación. Por otro lado, *la calidad de la segmentación debe ser evaluada en el contexto de la aplicación concreta y considerando el objetivo final de la segmentación* [115]. Desde finales de la década anterior y hasta el presente los métodos de evaluación creados se han desplazado en esta dirección [122].
2. Los algoritmos de segmentación debieran ser evaluados usando criterios compuestos. No obstante, son pocos los métodos de evaluación de segmentación que emplean varios criterios, especialmente criterios de diferencias ([122], [123], [125], [128]). La idea es, usar varios criterios para obtener información de varias propiedades del algoritmo de segmentación y proponer mejoras al algoritmo.
3. La evaluación no solamente se utiliza para juzgar el rendimiento final de los algoritmos de segmentación, sino también para combinar los resultados parciales de varias segmentaciones ([88], [101]) donde se persigue la selección y/o creación de algoritmos de segmentación apropiados [122] en cierta aplicación en particular.
4. La creación de imágenes test (ground truth), donde el trazado de la frontera del objeto es delineada manualmente, producen imágenes test más realistas que las creadas mediante aproximación geométrica simple ([115], [122]).
5. En general, el análisis de imágenes médicas es una aplicación importante de las técnicas de análisis de imágenes, no solamente porque atrae la atención del desarrollo de algoritmos de segmentación, sino también porque atrae mucho la atención de su evaluación. La razón es múltiple: por la complejidad del contenido en las imágenes médicas y por la precisión que se requiere ([32], [122]).
6. Hay cuatro cortes bien definidos en el seguimiento hecho a la evaluación de la segmentación en imágenes:

Corte 1: 1973-1996: A survey on evaluation methods for image segmentation [120].

Corte 2: 1996-2001: A review of recent evaluation methods for image segmentation [122].

Corte 3: 2001-2006: Chapter 18: Unsupervised and supervised image segmentation evaluation [88].

: Chapter 19: Objective evaluation of video segmentation quality [12].

: Chapter 20: A summary of recent progresses for segmentation evaluation [124].

Incluidos en el libro “Advances in Image and Video Segmentation” (editado por Zhang, [125]).

Corte 4: 2006-2009: Image segmentation in this century [128]. IGI Global.

El seguimiento evidencia que la evaluación de la segmentación ha recibido una atención creciente, que se le ha dedicado cada vez más esfuerzos, por lo que se ha obtenido cada vez más resultados; sin embargo son necesarias nuevas ideas, nuevas metodologías y nuevas implementaciones prácticas.

7. En el ICPR2008 se presentó la métrica RQM [100]. En este reporte se introduce TQM. Las métricas RQM y TQM pertenecen al grupo de *métodos empíricos de discrepancia* que utilizan ground truth para supervisar los resultados de la segmentación de imágenes. Ambas medidas permiten: 1) conocer cuantitativamente las discrepancias pixel a pixel que hay entre las clases-regiones segmentadas y las clases-regiones segmentadas idealmente; 2) obtener un índice de calidad general de la segmentación obtenida por el algoritmo en una aplicación particular; 3) obtener el % total de coincidencias en la imagen segmentada a partir de las coincidencias entre las clases-regiones segmentadas y las clases-regiones segmentadas idealmente; y 4) las métricas tienen acoplado un método que construye un mapa binario que expresa gráficamente las coincidencias y discrepancias de la *imagen segmentada* comparada con la *imagen segmentada ideal*.
8. Las métricas RQM y TQM evalúan internamente la eficacia de algoritmos de segmentación de imágenes, y aporta un índice de calidad general sobre la segmentación obtenida, dato que permite comparar la eficacia entre dos algoritmos cuando ambos segmentan la misma imagen. Los mismos mapas binarios de coincidencias, de cada algoritmo de segmentación, podrían ser comparados gráficamente para construir otro que reflejaría la distribución espacial entre las segmentaciones de coincidencias y discrepancias mutuas.

9. En el último corte (2006-2009) se presentaron más de 20 trabajos acerca de la evaluación de segmentación en imágenes. Los métodos presentados con novedades, obedecían a los principios establecidos, cierto número de trabajos hacen modificaciones y/o mejoras a las técnicas ya propuestas previamente; y aparecieron varios trabajos con nuevas ideas. Se nota que continúa el esfuerzo en los métodos de evaluación sobre la segmentación, aunque las nuevas ideas no marcan un hito de “progreso radical” [128].
10. Por su importancia práctica han sido concebidos ambientes (framework) [113] y bases de datos (benchmark) que permiten evaluar la segmentación de imágenes [99].
11. Aunque se han obtenido progresos recientemente, parece que los resultados en el campo de evaluar la segmentación de imágenes están aún lejos de ser satisfactorios [124]. Este autor ha dicho que cierto número de factores todavía están limitando el avance en la evaluación de la segmentación y que, a su vez, también están influyendo en el mejoramiento del rendimiento de los algoritmos de segmentación, y cita:
 - No hay un modelo matemático común, o estrategia general para la evaluación.
 - Es difícil definir métricas con un “rendimiento de amplio alcance y suficientes estadísticas”.
 - Los datos de prueba usados en la evaluación, con frecuencia no son suficientemente representativos de la aplicación real.
 - Los “ground truths” apropiados son difíciles de determinar objetivamente.
 - La realización de evaluaciones comprensibles, frecuentemente están vinculadas con altos costos tanto en tiempo como en esfuerzos.

Luego, la comunidad de investigadores continúa la búsqueda de la meta ideal, de un buen método de evaluación para la segmentación: 1) que caracterice al algoritmo, 2) que facilite la comparación de algoritmos usando criterios objetivos compuestos, 3) que sea capaz de integrarse al sistema de reconocimiento, y 4) que ayude a seleccionar adaptativamente los parámetros apropiados de la segmentación, para mejorar la exactitud del reconocimiento.

7.2 Tendencias futuras

La *segmentación de imágenes* ha tenido mucho progreso en las últimas 4 décadas, aunque todavía necesita estudios y esfuerzos adicionales. Algunas direcciones de investigación pudieran ser las siguientes [127]:

- *Incorporación de Factores Humanos*: La segmentación de imágenes es un proceso que se ubica en la capa media de la ingeniería de la imagen y por esta razón está fuertemente influenciada por los factores humanos. Al parecer, la ayuda de los humanos seguirá siendo esencial en cualquier dominio de aplicación por ser conocedores del método de segmentación y de la solución deseada. La incorporación en el algoritmo de segmentación, de los conocimientos de alto nivel que poseen los humanos es un desafío en el futuro.
- *Introducción de nuevas teorías y modelos matemáticos*: La introducción de varios modelos matemáticos y teorías en la investigación sobre la segmentación de imágenes debe ser realmente eficaz. En estos años han sido creados nuevos modelos matemáticos y teorías que se han introducido en los procesos de segmentación con resultados prometedores. Debe continuarse en esta dirección.
- *Ampliación del ámbito de Aplicaciones*: La vida moderna está exigiendo nuevas aplicaciones en segmentación de imágenes, lo cual aportará nuevas áreas con capacidades potenciales para desarrollar algoritmos. Dado que no existe una teoría general sobre segmentación de imágenes, las investigaciones para segmentar imágenes diferentes en aplicaciones particulares traeran nuevos progresos en los algoritmos. Y por supuesto que también traeran progresos en la evaluación de sus rendimientos.

La *evaluación de la segmentación en imágenes* ha avanzado teóricamente, pero no significativamente. Teniendo en cuenta el desarrollo más reciente que ha tenido, a partir de los criterios que se han estado utilizando, se presentan como novedosas y probables varias direcciones para los futuros desarrollos de las investigaciones:

- *Crear métodos de evaluación que reflejen el objetivo final de la segmentación* [128]. La segmentación de imágenes es el primer paso en la cadena del análisis de una imagen; el rendimiento de la segmentación influirá en todos los procesos siguientes. Luego, parece lógico que el propósito de la segmentación, sobre todo los objetivos que persigue la tarea que se analiza, sea considerado en una evaluación de la segmentación. Por un lado, el hecho de considerar el objetivo final de la segmentación influirá en que el procedimiento de evaluación sea más objetivo y que los resultados sean más útiles. Y por otro lado, el éxito de los procesos siguientes serán un indicador convincente de la calidad obtenida en la segmentación de la imagen. Como la segmentación no es un proceso aislado, el éxito de los procesos que le siguen deben ser un indicador de su calidad.

En Cuba se desarrollan cada vez más aplicaciones de segmentación usando imágenes médicas para el diagnóstico y evaluación médica progresiva de enfermedades. Dado que la calidad de la segmentación debe ser evaluada en el contexto de la aplicación concreta; sería útil la creación de una herramienta de software (framework) con métodos de evaluación simples y estrategias para combinarlas con el objetivo de evaluar el rendimiento de los algoritmos de segmentación de imágenes. Estos a su vez deben considerar la clasificación multi-espectral con los paradigmas tono, textura y tono+textura.

- *Crear métodos de evaluación que combinen eficientemente múltiples métricas* [128]. Los resultados de la segmentación y el rendimiento de los algoritmos de segmentación están relacionados con muchos factores. Por tal razón, no es sorprendente que la estrategia que combina varios criterios para crear métricas múltiples para la evaluación, haya sido considerada desde muy temprano [108], y que continúa siendo considerada en cierto número de trabajos recientes, como muestra la tendencia que se aprecia en la Tabla 7. La táctica para combinar múltiples métricas ha evolucionado desde hacer simples sumas pesadas, hasta técnicas complicadas, como es aplicar el enfoque de aprendizaje automático [113]. En muchas aplicaciones de interés, se necesitan trabajos en el futuro que combinen múltiples métricas que incluyan eficazmente diferentes aspectos a evaluar en la segmentación de imágenes.

Referencias bibliográficas

1. Abdou, I.E. & W.K. Pratt (1979): Quantitative design and evaluation of enhancement/thresholding edge detectors, Proc. IEEE Transaction, 67:753-763.
2. Behiels, G., Maes, F., Vandermeulen, D., & Suetens, P. (2002): Evaluation of image features and search strategies for segmentation of bone structures in radiographs using active shape models. Medical Image Analysis, 6(1), 47-62.
3. Betanzos, A., Varela, A., & Martínez, C. (2000). Analysis and evaluation of hard and fuzzy clustering segmentation techniques in burned patient images. Image and Vision Computing, 18(13), 1045-1054.
4. Borsotti, M.; P. Campadelli & R. Schettini (1998): Quantitative evaluation of color image segmentation results Pattern Recognition Letters, 19(8):741-747.
5. Borisenko, V. I., Zlatotol, A. A., & Muchnik, I. B. (1987). Image segmentation (state of the art survey). Automatic Remote Control, 48, 837-879.
6. Buf, J. M. H., & Campbell, T. G.. (1990). A quantitative comparison of edge-preserving smoothing techniques. Signal Processing, 21, 289-301.
7. Carleer, A. P., Debeir, O., & Wolff, E. (2004). Comparison of very high spatial resolution satellite image segmentations. SPIE, 5238, 532-542.

8. Cavallaro, A., Gelasca, E. D., & Ebrahimi, T. (2002). Objective evaluation of segmentation quality using spatio-temporal context. In *Proceedings of the International Conference on Image Processing* (Vol. 3, pp. 301-304).
9. Chang, K. I.; K. W. Bowyer Y M. Sivagurunath. Evaluation of texture segmentation algorithms. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*. Fort Collins, Colo, USA, 1999. vol. 1: 294-299.
10. Correia, P. L., Pereira, F. M. B. (2000). Objective evaluation of relative segmentation quality. In *Proceedings of the International Conference on Image Processing* (Vol. 1, pp. 308-311).
11. Correia, P. L., & Pereira, F. M. B. (2003). Methodologies for objective evaluation of video segmentation quality. *SPIE*, 5150:1594-1600.
12. Correia, P. L. & F. Pereira (2006): Chapter 19: Objective evaluation of video segmentation quality. en: *Advances in Image and Video Segmentation*. Hershey, USA, IRM Press, pp. 394-422.
13. Cosman, P.C., R.M. Gray & R.A. Olshen (1994): Evaluating quality of compressed medical images: SNR, subjective rating, and diagnostic accuracy, *Proceedings of IEEE Transaction*, 82:919-932.
14. Davis, L. S. (1975). A survey of edge detection techniques. *Computer Graphics and Image Processing*, 4, 248-270.
15. Desurmont, X., Wijnhoven, R., & Jaspers, E., et al. (2005). Performance evaluation of realtime video content analysis systems in the CANDELA project. *SPIE*, 5671, 200-211.
16. Erdem, C. E., Sankur, B., & Tekalp, A. M. (2004). Performance measures for video object segmentation and tracking. *IEEE Image Processing*, 13(7), 937-951.
17. Everingham, M., Muller, H., & Thomas, B. (2002). Evaluating image segmentation algorithms using the Pareto Front. In *Proceedings of the European Conference on Computer Vision* (Vol. 4, pp. 34-48).
18. Fram, J.R. & E.S. Deutsch (1975): On the quantitative evaluation of edge detection schemes and their comparison with human performance, *IEEE Transaction*, C-24, 616-628.
19. Freixenet, J., Muñoz, X., Raba, D., Martí, J., & Cufí, I. (2002). Yet another survey on image segmentation: Region and boundary information integration. In *Proceedings of 7th European Conference on Computer Vision-Part III (ECCV '02)*, 2002, pp. 408-422. Copenhagen, Denmark: Springer.
20. Fu, K. S. & Mui, J.K. A survey on image segmentation. *Pattern Recognition*, 1981, Vol. 13: 3-16.
21. Gao, X., Boulton, T. E., & Coetzee, F., et al. (2000). Error analysis of background adaptation. In *Proceedings of the Computer Vision and Pattern Recognition*. Vol. 1, pp. 503-510.
22. Garbay, C.: Image structure representation and processing: a discussion of some segmentation methods in cytology. *IEEE Trans. PAMI-8*, 140-146, (1986).
23. Gerbrands, J. J.: Segmentation of noisy images. Doctoral Thesis, Delft University of Technology, Delft, The Netherlands, (1988).
24. Gil, J.L.: Segmentación de textura en imágenes combinando detectores de bordes y regiones (2007-2009). Proyecto de I+D. Código T0603-C23. Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV). MINBAS.
25. Gil, J.L.: Estado actual de la representación y análisis de textura en imágenes (2008). Reporte Técnico RT_004. Serie Azul. Reconocimiento de Patrones. La Habana. Centro de Aplicaciones de Tecnologías de Avanzada. RNPS-0551: 75p. http://www.cenatav.co.cu/doc/RTecnicos/RT%20SerieAzul_004web.pdf
26. González, R.C. & P. Wintz (2002): *Digital Image Processing*. Addison-Wesley, New York.
27. Graaf, C.N., A.S.E. Koster, K.L. Vincken & M.A. Viergeever (1994): Validation of the interleaved pyramid for the segmentation of 3D vector images, *Pattern Recognition Letters*. 15:467-475.
28. Guo, G.D., S. Li, K. Chan, H. Pan (2000): Texture Image Segmentation using Reduced Gabor Filter Set and Mean Shift Clustering. *4th Asian Conf. on Comp. Vision*, pp. 198-203,.
29. Haralick, R.M. & L.G. Shapiro, (1985). Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, 29, 100-132.
30. Haralick, R.M. & L.G. Shapiro (1992): *Computer and Robot Vision*. Addison-Wesley, New York
31. Haralick, R. M. (1994): Performance characterization in computer vision, *CVGIP-IU*, 60:245-249.
32. Haynor, D.R. (2000): Performance evaluation of image processing algorithms in medicine: a clinical perspective. *SPIE Electronic Imaging - Storage and Retrieval Methods and Applications for Multimedia*, 3979(18).
33. Hoover, A., Jean-Baptiste, G., & Jiang, X., et al. (1996): An experimental comparison of range image segmentation algorithms. *IEEE Pattern Analysis and Machine Intelligence*, 18(7):673-689.
34. Huo, Z., & Giger, M. L. (2000). Evaluation of an automated segmentation method based on performances of an automated classification method. *SPIE*, 398, 16-21.

35. Huang, Q. & B. Dom (1995): Quantitative methods of evaluating image segmentation. Proceedings of International Conference on Image Processing (ICIP '95). Washington, DC, USA, IEEE Transaction, Vol. 3: 53-56.
36. Jiang, X. (2000): An adaptive contour closure algorithm and its experimental evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1252-1265.
37. ----- (2005): Performance evaluation of image segmentation algorithms. en: Chapter 5, Handbook of Pattern Recognition and Computer Vision., 3rd edition. C.H. Chen & P.S.P.Wang, E. Singapore, pp.525-542.
38. ----- (2007): Chapter 5: Performance Evaluation of Image Segmentation Algorithms. en: CVPR-Handbook. Munster, University of Munster, Germany, 18.p.
39. ----- (2006): Distance Measures for Image Segmentation Evaluation. EURASIP Journal on Applied Signal Processing, 2006(1): 1-10.
40. Jiang, X., et al. (2000): Some further results of experimental comparison of range image segmentation algorithms. Proceedings of 15th International Conference on Pattern Recognition (ICPR '2000). Barcelona, Spain, IEEE Transaction, Vol. 4, pp. 877-881.
41. Jiang, X. & D.D. Mojon (2002): Supervised evaluation methodology for curvilinear structure detection algorithms. Proceedings of 16th International Conference on Pattern Recognition (ICPR '02), Quebec, Canada., pp. 103-106 p.
42. Jiang, X., et al. (2005): Image segmentation evaluation by techniques of comparing clusterings. Proceedings of 13th International Conference on Image Analysis and Processing (ICIAP '05), Cagliari, Italy.
43. Kanungo, T., M. J. Jaisimba, J. Palmer, & R. M. Haralick (1995): Methodology for Quantitative Performance Evaluation of Detection Algorithms. IEEE Transaction on Image Processing, 4(12):1667-1674.
44. Kim, K., Chalidabhongse, T. H., & Harwood, D., et al. (2004). Background modeling and subtraction by codebook construction. In Proceedings of the International Conference on Image Processing. Vol. 5, pp. 3061-3064.
45. Kirbas, C., & F.K.H. Quek, (2003): Vessel extraction techniques and algorithms: A survey. Proceedings of the 3rd Bioinformatics and Bioengineering Symposium (pp. 238-245). Bethesda, MD: IEEE Computer Society Press.
46. Kitchen, L.J. & J. A. Malin (1989): The effect of spatial discretization on the magnitude and direction response of simple differential edge operators on a step edge, CVGIP 47, 243-258.
47. Koprinska, I., & S. Carrato (2001): Temporal video segmentation: A survey. Signal Processing: Image Communication, 16(5): 477-500.
48. Ladak, H. M., Ding, M., Wang, Y., et al. (2004). Evaluation of algorithms for segmentation of the prostate boundary from 3D ultrasound images. SPIE, 5370, 1403-1410.
49. Li, N., Li, S., & Chen, C. (2003). Multimetric evaluation protocol for user-assisted video object extraction systems. SPIE, 5150, 20-28.
50. Liedtke, C.E., T. Gahm, F. Kappei & B. Aeikens (1987): Segmentation of microscopic cell scenes. AQCH 9, 197-211.
51. Lievers, W. B., & Pilkey, A. K. (2004). An evaluation of global thresholding techniques for the automatic image segmentation of automotive aluminum sheet alloys. Materials Science and Engineering , A. 381(1-2), 134-142.
52. Lee, S.U., S.Y. Chung & R.H. Park (1990): A comparative performance study of several global thresholding techniques for segmentation, CVGIP 52:171-190.
53. Lefèvre, S., J. Holler, & N. Vincent (2003): A review of real-time segmentation of uncompressed video sequences for content-based search and retrieval. Real-Time Imaging, 9(1):73-98.
54. Levine, M. D. & A. Nazif (1982): An experimental rule based system for testing low level segmentation strategies, in Multi-Computers and Image Processing: Algorithms and Programs, K. Preston and L. Uhr, eds., pp. 149-160. Academic Press, New York.
55. Levine, M.D. & A. M. Nazif (1985): "Dynamic measurement of computer generated image segmentations," IEEE Transactions on Pattern Analysis and Machine Intelligence 7(2):155-164.
56. Lim, Y.W. & S.U. Lee (1990): On the color Image segmentation algorithms based on the thresholding and fuzzy c-means techniques, Pattern Recognition, 23:935-952.
57. Liu, J. & Y.H. Yang (1994): "Multi-resolution color image segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 16, pp. 689-700, July.

58. Liu, W., & Dori, D. (1999). Principles of constructing a performance evaluation protocol for graphics recognition algorithms. In *Performance characterization and evaluation of computer vision algorithms* (pp. 97-106).
59. Mao, S., & Kanungo, T. (2001). Empirical performance evaluation methodology and its application to page segmentation algorithms. *IEEE Pattern Analysis and Machine Intelligence*, 23(3):242-256.
60. Mao, F. et al. (1999): "Technique for evaluation of semi-automatic segmentation methods". *SPIE*, 3661:1027-1036.
61. Marcello, J., Marques, F., & Eugenio, F. (2004): Evaluation of thresholding techniques applied to oceanographic remote sensing imagery. *SPIE*, 5573:96-103.
62. Mattana, M. F., Facon, J., & Britto, A. S. (1999). Evaluation by recognition of thresholding based segmentation techniques on Brazilian bankchecks. *SPIE*, 3572, 344-348.
63. Min, J., M. Powell, & K. W. Bowyer (2004): Automated performance evaluation of range image segmentation algorithms. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 34(1).
64. Mirmehdi, M., X. Xie & J. Suri, Eds. (2008): *Handbook of Texture Analysis*. London, Imperial College Press. 424p. ISBN-13 978-1-84816-115-3
65. Niemeijer, M., Staal, J., & Ginneken, B., et al. (2004): Comparative study of retinal vessel segmentation methods on a new publicly available database. *SPIE*, 5370:648-656.
66. Neubert, M., H. Herold & G. Meinel (2006): Evaluation of remote sensing image segmentation quality – further results and concepts. Dresden, Germany, Leibniz Institute of Ecological and Regional Development (IOER). 6p.
67. Moretti, B., Fadili, J. M., & Ruan, S., et al. (2000): Phantom-based performance evaluation: Application to brain segmentation from magnetic resonance images. *Medical Image Analysis*, 4(4):303-316.
68. Munkres, J. (1957): Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32-38.
69. Nazif, A.M. & M. D. Levine (1984): Low level image segmentation: an expert system, *IEEE Trans. PAMI*, 6:555-577.
70. Nevatia, R. (1986): Image segmentation. In *Handbook of pattern recognition and image processing*, 86, 215-231.
71. Oberti, F., Teschioni, A., & Regazzoni, C. S. (1999): ROC curves for performance evaluation of video sequences processing systems for surveillance applications. In *Proceedings of the International Conference on Image Processing* (Vol. 2, pp.949-953).
72. Olabariaga, S. D., & A.W.M. Smeulders (2001): Interaction in the segmentation of medical images: A survey. *Medical Image Analysis*, 5(2), 127-142.
73. Otsu, N. (1979): "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics* 9, pp. 62–66, January.
74. Pal, N.R. & S. K. Pal (1989): Entropic thresholding. *Signal Processing*. 16:97-108.
75. Pal, N.R. & S.K. Pal (1993): A review on image segmentation techniques. *Pattern Recognition*. 26:1277-1294.
76. Pal, N. & D. Bhandari (1993): "Image thresholding: some new techniques". *Signal Processing*. Vol. 33, pp. 139–158, August.
77. Pham, D., Xu, C., & Prince, J. (2000). Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2, 315-337.
78. Pavlidis, T. (1986). Critical survey of image analysis methods. *Proceedings 8th International Conference on Pattern Recognition*, Paris, France (pp. 502-511). New York: IEEE.
79. Peli, T., & Malah, D. (1982): A study of edge determination algorithms. *Computer Graphics and Image Processing*, 20, 1-20.
80. Prati, A., Mikic, I., & Trivedi, M. M., et al. (2003): Detecting moving shadows: Algorithms and evaluation. *IEEE Pattern Analysis and Machine Intelligence*, 25(7):918-923.
81. Pratt, W.K. (1978): *Digital Image Processing*. John Wiley and Sons, New York.
82. Renno, J. R., Orwell, J., & Jones, G. A. (2004): Evaluation of shadow classification techniques for object detection and tracking. In *Proceedings of the IEEE International Conference on Image Processing* (pp. 143-146).
83. Riseman, E. M., & Arbib, M. A. (1977). Computational techniques in the visual segmentation of static scenes. *Computer Graphics and Image Processing*, 6, 221-276.

84. Rosenberger, C., & Chehdi, K. (2000): Genetic fusion: Application to multi-components image segmentation. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (Vol. 4, pp. 2223-2226).
85. Roberts, L. G. Machine perception of three-dimensional solids: Optical and electrooptical information processing. J. T. Tippett, D. A. Berkowitz, L. C. Clapp, C. J. Koester, & A. Vanderburgh (Eds.), Cambridge, MA: MIT Press., 1965. 149-157.
86. Rosin, P. L., & Ioannidis, E. (2003): Evaluation of global image thresholding for change detection. *Pattern Recognition Letters*, 24(14), 2345-2356.
87. Rosenberger, C. (2005): Evaluation des traitements dans les systèmes d'analyse d'image. [Consultado en 2009]. Disponible en: <http://www.greyc.ensicaen.fr/scati/journees/journee151205/>
88. Rosenberger, C., S. Chabrier, H. Laurent & B. Emile (2006): Chapter XVIII Unsupervised and supervised image segmentation evaluation. en: *Advances in Image and Video Segmentation*. Hershey, USA, IRM Press. 365-392.p.
89. Rosenfeld, A. (1981): Image pattern recognition. *Proceedings of IEEE Transaction*, Vol. 69, 5:596-605.
90. Rosenfeld, A. & L. S. Davis (1979): Image segmentation and image models. *Proceedings of IEEE Transaction*, Vol. 67, pp. 764-772.
91. Sahoo, P., S. Soltani, A. Wong, & Y. Chen (1988): "A survey of thresholding techniques". *Computer Vision, Graphics, and Image Processing* 41, pp. 233-260, February.
92. Sarkar, S., & Boyer, K. L. (1993). Perceptual organization in computer vision: A review and a proposal for a classificatory structure. *IEEE Systems, Man and Cybenetics*, 23, 382-399.
93. Strasters, K. & J.J. Gerbrands (1991): Three-dimensional image segmentation using a split, merge and group approach, *Pattern Recognition Lett.* 12:307-325.
94. Shirai, Y. (1987): *Three-Dimensional Computer Vision*. Springer-Verlag, Berlin.
95. Sezgin, M. & B. Sankur (2004): Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146-165.
96. Tuceryan, M., A.K. Jain (1998): *The Handbook of Pattern Recognition and Computer Vision* (2nd Edition), by C. H. Chen, L. F. Pau, P. S. P. Wang (eds.), pp. 207-248, World Scientific Publishing Co.
97. Udupa, J. K., LeBlanc, V. R., & Schmidt, H., et al. (2002): A methodology for evaluating image segmentation algorithms. *SPIE*, 4684:266-277.
98. Udupa, J., & Zhuge, Y. (2004): Delineation operating characteristic (DOC) curve for assessing the accuracy behavior of image segmentation algorithms. *SPIE*, 5370:640-647.
99. UTIA, (2008): *The Prague Texture Segmentation Datagenerator and Benchmark*. [Consultado en 2009]. Disponible en: <http://mosaic.utia.cas.cz/> .
100. Vega-Pons, S., J.L. Gil & O.L. Vera: "Active contour algorithm for texture segmentation using a texture feature set". *Pattern Recognition*, 2008, ICPR2008. 19th International Conference on Pattern Recognition. TuBCT8.32 (1-4), 2008. ISSN: 1051-4651, ISBN: 978-1-4244-2174-9. DOI:10.1109/ICPR.2008.4761583
101. Vega-Pons, S. & Shulcloper, J. (2009): A survey of clustering ensemble algorithms. *Sended to International Journal of Pattern Recognition and Artificial Intelligence*. © World Scientific Publishing Company. Personal communication.
102. Wenyin, L., J. Zhai, D. Dori & T. Long (2002): *A System for Performance Evaluation of Arc Segmentation Algorithms CVPR '01*, Kauai, Hawaii, USA.
103. Weszka, J. S. (1978). A survey of threshold selection techniques. *Computer Graphics and Image Processing*, Vol. 7, pp. 259-265.
104. Weszka, J. & A. Rosenfeld (1978): "Threshold evaluation techniques". *IEEE Transactions on Systems, Man and Cybernetics* Vol. 8, pp. 622-629, August.
105. Wirth, M., et al. *Performance Evaluation in Image Processing*. Hindawi Publishing Corporation, *EURASIP Journal on Applied Signal Processing*, 2006, vol. 2006, (Article ID 45742): 1-3.
106. Xu, Y., Olman, V., & Uberbacher, E. C. (1998): A segmentation algorithm for noisy images: Design and evaluation. *Pattern Recognition Letters*, 19(13):1213-1224.
107. Yang, J., & Huang, S. C. (1999): Method for evaluation of different MRI segmentation approaches. *IEEE Network Simulator*, 46(6), 2259-2265.
108. Yasnoff, W. A., J. K. Mui, J.W. Bacus (1977): Error measures for scene segmentation. *Pattern Recognition*, 9: 217-231.
109. Yasnoff, W.A. & J. W. Bacus (1984): Scene segmentation algorithm development using error measures, *AOCH* 6, 45-58.

110. Yitzhaky, Y. & E. Peli (2003): A Method for Objective Edge Detection Evaluation and Detector Parameter Selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):1027-1033.
111. Young, I.T. (1988): Sampling density and quantitative microscopy, *AOCH* 10:269-275.
112. Zhang-Hui, J.E. Fritts & S.A. Goldman (2004): An Entropy-based Objective Evaluation Method for Image Segmentation. *SPIE Electronic Imaging - Storage and Retrieval Methods and Applications for Multimedia*. Washington. Vol. 5307, pp. 38-49.
113. Zhang-Hui, J.E. Fritts, S.A. Goldman (2005a): A co-evaluation framework for improving segmentation evaluation. *SPIE Electronic Imaging - Storage and Retrieval Methods and Applications for Multimedia*, 5809:420-430.
114. Zhang-Hui et al. (2005b): Meta-Evaluation of Image Segmentation Using Machine Learning. *SPIE Electronic Imaging-Storage and Retrieval Methods and Applications for Multimedia*. Washington, USA.
115. Zhang, Y.J. (1992): Segmentation evaluation using ultimate measurement accuracy *SPIE Electronic Imaging - Storage and Retrieval Methods and Applications for Multimedia*. 1657:449-460.
- 116.----- (1993a): Comparison of segmentation evaluation criteria. *Proceedings of the 2nd International Conference on Signal Processing*, 1:870-873.
- 117.----- (1993b): Image synthesis and segmentation comparison criteria. *Proceeding of the 3th International Conference YCS*, 821-824.
- 118.----- (1993c): Segmentation evaluation and comparison: a study of various algorithms, *SPIE*, 2094:801-812.
- 119.----- (1995): Influence of image segmentation over feature measurement. *Pattern Recognition Letters*. 16:201-206.
- 120.----- (1996): A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335-1346.
- 121.----- (1997): Evaluation and comparison of different segmentation algorithm. *Pattern Recognition Letters*. 18(10):963-974.
- 122.----- (2001): A review of recent evaluation methods for image segmentation. *Proceedings of the 6th International Symposium on Signal Processing and Its Applications*, Kuala Lumpur, Malaysia, *IEEE Signal Processing*. pp. 148-151.
- 123.----- (2005): The evolution of evaluation for image segmentation. [Fecha de consulta : Nov. 2009]. Disponible en: <http://www.greyc.ensicaen.fr/scati/journees/journee151205/zhang.pdf>
- 124.----- (2006a): Chapter XX: A summary of recent segmentation evaluation. en: *Advances in Image and Video Segmentation*. Hershey, USA, IRM Press, pp.423-440.
- 125.----- (2006b): *Advances in Image and Video Segmentation*. Hershey, USA, IRM Press, 473p. ISBN 1-59140-755-9 (ebook).
- 126.----- (2006c): Chapter I: An Overview of Image and Video segmentation in the last 40 years. en: *Advances in Image and Video Segmentation*. Hershey, USA, IRM Press, pp. 1-15.
- 127.----- (2009a): Image Segmentation Evaluation in the Last 40 Years. *Encyclopedia of Information Science and Technology*. Beijing, Tsinghua University, China, pp. 1818-1823.
- 128.----- (2009b): Image Segmentation Evaluation in This Century. *Encyclopedia of Information Science and Technology*. Beijing, Tsinghua University, China, pp. 1812-1817.
- 129.----- (2009c): Recent Progress in Image and Video Segmentation for CBVIR. *IGI Global. Encyclopedia of Information Science and Technology*. Beijing, Tsinghua University, China, pp. 3224-3228.
130. Zhang, Y.J. & J.J. Gerbrands (1992a): Segmentation evaluation using ultimate measurement accuracy. *SPIE*, 1657:449-460.
131. Zhang, Y.J. & J.J. Gerbrands (1992b): On the design test images for segmentation evaluation. *Proceedings of EUSIPCO-92*, Vol. 1, pp. 551-554.
132. Zhang, Y.J. & J.J. Gerbrands (1994): Objective and quantitative segmentation evaluation and comparison. *Signal Processing*. 39(3):43-54.
133. Zhang, Y.J. & H.T. Luo (2000). Optimal selection of segmentation algorithms based on performance evaluation. *Optical Engineering*, 39(6):1450-1456.
134. Zucker, S. W. (1976): Region growing: Childhood and adolescence. *Computer Graphics and Image Processing*, Vol. 5, pp. 382-399.
135. Zucker, S. W. (1977): Algorithms for image segmentation. In A. Rosenfeld & J. C. Simon (Eds.) *Nordhoff International Publisher. Digital image processing and analysis*, pp. 169-183.

Anexo

Glosario de términos¹

Análisis en el espacio de rasgos (feature space analysis): Análisis de la imagen que se realiza en el espacio de rasgos, en el cual cada punto corresponde a un valor del rasgo extraído de la imagen.

Búsqueda en grafo (graph search method): El método es un tipo particular de técnica de segmentación que combina la detección y el enlace de bordes. Representa segmentos de bordes en un grafo, de modo que los bordes significativos o fronteras de objetos, corresponden a una búsqueda en el grafo por caminos de bajo-costo.

Caracterización de la segmentación: La caracterización de la segmentación es un proceso intra-algoritmo para evaluar la segmentación. El propósito de evaluar a un algoritmo de segmentación específico es conocer cuantitativamente su comportamiento ante varias imágenes y ayudar en consecuencia, a definir los parámetros que necesita ese algoritmo ante diferentes aplicaciones buscando obtener el mejor rendimiento.

Comparación de la segmentación: La comparación de la segmentación es un proceso inter-algoritmo para evaluar la segmentación. El propósito de comparar diferentes algoritmos es ranquear sus rendimientos y proporcionar pautas para seleccionar los algoritmos adecuados según las aplicaciones, así como, para promover la creación de nuevos algoritmos tomando en cuenta eficazmente, los puntos fuertes de varios algoritmos.

Criterio compuesto: El criterio de evaluación compuesto se forma combinando varias métricas de rendimiento para cubrir varios aspectos de los algoritmos de segmentación. La combinación puede hacerse de diferentes maneras, tales como una combinación lineal, o usando un enfoque de aprendizaje automático y otras.

Criterio de evaluación: Criterio usado en el proceso de evaluación para juzgar el rendimiento de algoritmos de segmentación de imágenes. A los criterios de evaluación también se les llaman métricas de rendimiento, medida de rendimiento, o índices de rendimiento.

Corte de grafo (graph cuts): Una técnica ampliamente usada en agrupamiento y minimización de energía. Cuando hay sólo dos etiquetas, el problema de minimización de la energía puede reducirse directamente a un problema de cálculo de máximo flujo (o mínimo corte) de un grafo.

Criterio objetivo: Criterio basado en valores o cantidades determinadas objetivamente, los cuales indican las diferencias entre la imagen segmentada y la imagen de referencia. El criterio objetivo se utiliza principalmente en los métodos empíricos de diferencias para evaluar la segmentación de imágenes.

Criterio subjetivo: El criterio subjetivo está basado en el juicio humano de la percepción, el cual refleja algunas propiedades deseables de las imágenes segmentadas. El criterio subjetivo se usa en los métodos empíricos de bondad (empirical goodness methods) destinados a evaluar la segmentación de imágenes.

Detección de bordes (edge detection): la detección de bordes es el enfoque más común para la detección de discontinuidades en las imágenes, y es el paso fundamental del proceso de segmentación basado en bordes. Un borde es un concepto local. Para formar la frontera completa de un objeto, la detección de bordes debe ser seguida por un proceso de “enlace o conexión de bordes” (edge linking, en inglés).

Evaluación de la segmentación: La evaluación de la segmentación es un proceso que se utiliza para juzgar el rendimiento de los algoritmos de segmentación. La evaluación se basa en algunos

¹ Los términos del glosario fueron tomados de Yu-Jin Zhang ([120]; [127]; [128]; [129]), Zhang-Hui [114], Mirmehdi [64], y Gil [25]).

criterios de calidad previamente definidos y/o, en la verdad terreno (ground truth), con vistas a evaluar o revelar las propiedades de los algoritmos en uso.

Ingeniería de la imagen (image engineering): La ingeniería de la imagen es una disciplina que abarca el estudio de las diferentes ramas de la imagen y técnicas de video. Consiste principalmente en tres niveles: procesamiento de la imagen, análisis de la imagen, y comprensión de la imagen.

Método Analítico (analytic methods): Los métodos analíticos trabajan directamente sobre los algoritmos de segmentación y no sobre sus resultados. Ofrecen información acerca del principio del algoritmo de segmentación, su estrategia (paralelo, secuencial, o mixto), sobre sus requerimientos computacionales, complejidad del algoritmo, utilidades, etc., del algoritmo. Cuando se usan los métodos analíticos para evaluar a los algoritmos de segmentación se elimina la influencia causada por los experimentos de evaluación como hacen los métodos empíricos. Sin embargo, no todas las propiedades de los algoritmos de segmentación pueden ser obtenidas por estudios analíticos [113], [114]).

Método Empírico de Bondad (empirical goodness methods): 1) Estos métodos caracterizan a los diferentes algoritmos de segmentación mediante el simple cálculo de medidas de bondad basada en la imagen segmentada, sin el conocimiento a priori de la segmentación correcta. La aplicación de estos métodos de evaluación exige el requerimiento de usar referencias, de modo que ellos pueden ser usados en evaluación en línea. 2) Se conocen como métodos de evaluación no supervisados o métodos autosuficientes. Evalúan cuantitativamente los resultados de los algoritmos de segmentación de acuerdo a alguna caracterización que el humano ha hecho sobre las propiedades de una segmentación “ideal”. El beneficio de estos métodos es que ellos no necesitan de un conocimiento a priori de la segmentación correcta. Por ejemplo, estos métodos no necesitan ser evaluados contra una imagen de referencia segmentada manualmente. Las ideas fundamentales que se están utilizando en el diseño de los métodos son: la uniformidad intra-región, el contraste inter-región, forma de la región, y evaluación del borde [114].

Método Empírico de Diferencia (empirical discrepancy methods): Los métodos de diferencia empírica son conocidos como métodos de evaluación supervisada. Estos métodos evalúan los algoritmos de segmentación comparando la imagen segmentada resultante contra una imagen de referencia que ha sido segmentada manualmente y que a menudo se le llama “la norma de oro” [27]. Los métodos de diferencia empírica tienen la ventaja, por encima de los métodos de bondad empírica, que ellos hacen una comparación directa entre la imagen segmentada y la imagen de referencia, por lo cual se cree que ellos proporcionan una evaluación de resolución más fina y por tales motivos, son los métodos de evaluación objetiva que normalmente se usan. Sin embargo, la generación manual de una imagen de referencia es un trabajo difícil, subjetivo y altamente consumidor de tiempo. Y para la mayoría de las imágenes, especialmente las imágenes naturales, generalmente no se puede garantizar que una imagen segmentada manualmente sea mejor que otra. Además, tal método de evaluación no se puede usar dentro de un algoritmo de segmentación con el objetivo de encontrar los mejores parámetros de un algoritmo de segmentación parametrizado [114].

Modelo de Contornos Activos (active contour model): El modelo de contornos activos es una técnica secuencial para segmentación de imágenes. Dada una aproximación inicial de la frontera de un objeto, el modelo de contornos activos se usa para encontrar la frontera actual deformando la frontera inicial encerrando los rasgos de interés en la imagen. Combinando el modelo de contornos activos con una representación basada en conjuntos de nivel, la frontera inicial se deforma y desdobra en múltiples contornos para encerrar “n” regiones.

Modelo de Mezcla Gaussiana (gaussian mixture model): Modelo de agrupamiento tradicional basado en parámetros. Se asume que todas las muestras pertenecen a una distribución Gaussiana, el conjunto puede ser descrito por la suma de varias distribuciones Gaussiana.

Modelo k-Means: Modelo tradicional de agrupamiento basado en parámetros. Las “k” clases iniciales se representan por sus valores de medias, las iteraciones siguientes intentan minimizar la distancia entre las muestras con sus respectivas medias.

Operador gradiente (gradient operator): El operador gradiente es el primer tipo de operador que se usa en la detección de bordes. El gradiente de una imagen es un vector que contiene las derivadas de primer orden de la imagen lo que incluye magnitud y dirección del vector.

Primera confianza más alta (highest confidence first): Una técnica eficiente para minimizar la energía. Es un algoritmo de minimización determinístico, que es adecuado para asignar una etiqueta a los píxeles que etiqueta desconocida, porque la técnica introduce una etiqueta no comprometida.

Segmentación de la imagen: Es un proceso que consiste en subdividir la imagen en sus partes constituyentes y la extracción de las partes de interés (objetos) de la imagen. Es un paso fundamental y una tarea crucial en el análisis de la imagen.

Técnica de agrupamiento (clustering techniques): 1) Al “clustering” también se le llama aprendizaje no supervisado y es una técnica poderosa en clasificación de patrones. Es un proceso para agrupar dos o más términos basado en algún criterio que se define previamente, y así llegar a formar una colección grande. Con frecuencia, en el contexto de la segmentación de imágenes, el “clustering” se considera como la extensión multidimensional de la técnica de umbralamiento (thresholding). 2) En el contexto de segmentación de la imagen, el clustering es reunir varios píxeles o grupos de píxeles con una propiedad similar y con ellos formar una región más grande.

Técnica crecimiento de regiones (region growing techniques): El crecimiento de regiones es una técnica secuencial para la segmentación de la imagen basada en regiones. La técnica congrega píxeles en regiones cada vez más grandes sobre la base de: píxeles semillas predefinidos, un criterio de crecimiento y condiciones de parada.

Técnica de umbralamiento (thresholding techniques): Es la técnica de segmentación de la imagen más popular que existe. Comienza creando un “juego de umbrales” convenientes, y entonces la imagen es segmentada comparando las propiedades del píxel con estos umbrales.

Técnica “parteaguas” (watersheds techniques): 1) La técnica watershed está inspirada en la interpretación topográfica de “líneas divisorias” para segmentar la imagen. La técnica watershed incluye muchos conceptos de detección de bordes, detección de regiones y uso de umbrales que a menudo produce resultados estables y continuos. 2) Una transformación, o algoritmo para segmentación de imágenes. El watershed tradicionalmente está clasificado como un enfoque de segmentación basado en regiones. La idea subyacente en watershed es que la transformación viene de la geografía

Textura: 1) La textura son variaciones características de la intensidad o el color, que típicamente son originadas por por la rugosidad en la superficie de los objetos. En una textura bien definida las variaciones de la intensidad normalmente exhiben tanto regularidad como aleatoriedad.

RT_042, junio 2011

Aprobado por el Consejo Científico CENATAV

Derechos Reservados © CENATAV 2011

Editor: Lic. Lucía González Bayona

Diseño de Portada: Di. Alejandro Pérez Abraham

RNPS No. 2142

ISSN 2072-6287

Indicaciones para los Autores:

Seguir la plantilla que aparece en www.cenatav.co.cu

C E N A T A V

7ma. No. 21812 e/218 y 222, Rpto. Siboney, Playa;

La Habana. Cuba. C.P. 12200

Impreso en Cuba

