

4. Segmentación de imágenes

1. Introducción

La segmentación de una imagen consiste en la división o separación de ésta en regiones de atributos similares. Es decir, la segmentación subdivide una imagen en sus partes constituyentes u objetos. El grado de subdivisión depende del problema a resolver, por lo que la segmentación deberá detenerse cuando los objetos de interés de nuestro estudio hayan sido aislados.

Los algoritmos para segmentar imágenes monocromáticas generalmente se basan en una de las dos propiedades básicas de los valores del nivel de gris: discontinuidad y similitud. En la primera categoría, la aproximación es dividir la imagen basándose en cambios abruptos en el nivel de gris. Las principales áreas de interés en esta categoría son de detección de puntos aislados, de líneas y de bordes en una imagen. La segunda categoría se basa en segmentar por niveles de umbral, en regiones de crecimiento y union y division de regiones.

En general, la segmentación autónoma es una de las tareas más difíciles del procesamiento de imágenes. Esta etapa del proceso determina el eventual éxito o fracaso del análisis.

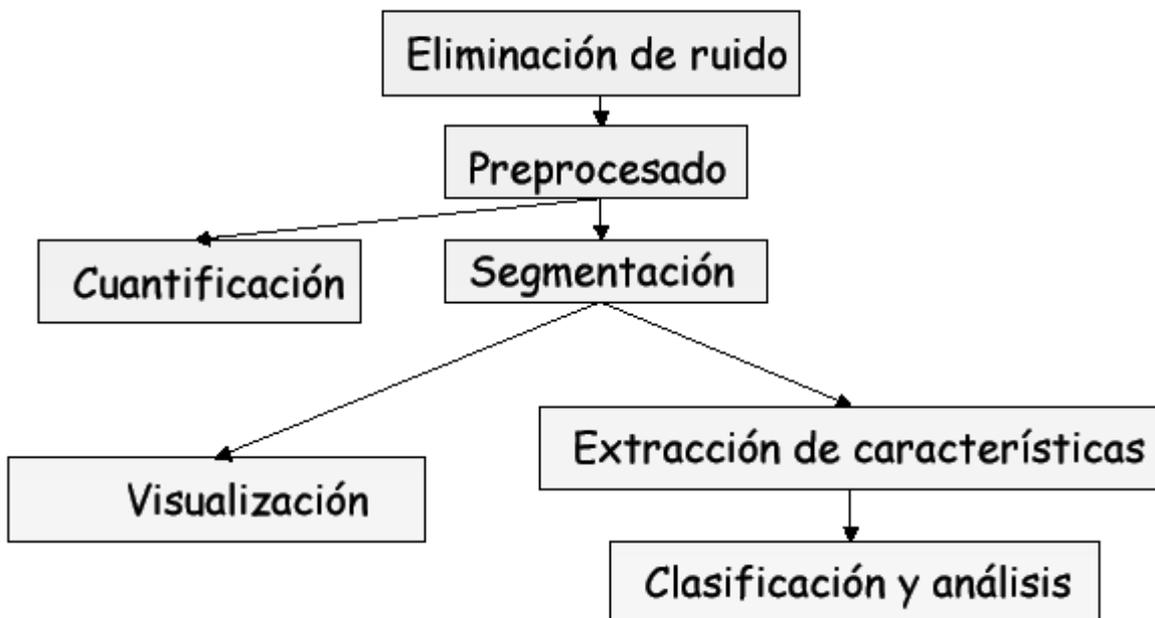


Figura 4.1: Esquema general del análisis de una imagen

1.1. Segmentación: Concepto

La segmentación es el proceso que divide una imagen en regiones u objetos cuyos píxeles poseen atributos similares, y constituye uno de los procesos más importantes de un sistema automatizado de visión, ya que permite extraer los objetos de la imagen para su posterior descripción y reconocimiento.

La segmentación es un paso imprescindible en diversos procesos de tratamiento de imagen. Entre otros, es necesaria para tomar medidas sobre una región, para realizar reconstrucciones tridimensionales de una zona de la imagen, para la clasificación o diagnóstico automático o para reducir la información de las imágenes. Si de una serie de imágenes para un determinado estudio sólo interesa una región concreta se puede segmentar y almacenar sólo las regiones para el análisis posterior.

La segmentación termina cuando los objetos que aparecen en una imagen se corresponden unívocamente con las distintas regiones disjuntas resultado del proceso de segmentación. En este caso se habla de *segmentación completa*. En caso contrario hablaremos de *segmentación parcial*.

Si se está interesado es una única región, se divide la imagen en objeto y fondo, considerando “fondo” todas las regiones de la imagen que no interesan.

Aunque con la vista, la detección de regiones pueda parecer una tarea sencilla, aparecen una serie de dificultades a la hora de realizar la segmentación de una imagen:

- ◆ **Artefactos:** Por problemas en la adquisición se puede encontrar ruido en la imagen, que distorsiona las características de las distintas regiones.
- ◆ **Volumen parcial:** Al ser las imágenes representaciones 2D, de objetos 3D, se encuentra en la imagen información que no es propiamente de ese plano y que limita la definición de las regiones.
- ◆ **Inhomogeneidad:** Los objetos no poseen siempre una iluminación uniforme, sino que muchas veces es sólo una textura lo que los diferencia de otros.

- ◆ **Forma:** Se pueden clasificar las formas de los objetos *grosso modo* en objetos con una forma más o menos regular y convexa, objetos con bordes no definidos y ciertas ramificaciones, y objetos finos y muy ramificados. Estos últimos pueden ser difíciles de segmentar ya que pueden estar entrelazados con otros objetos.
- ◆ **Ruido:** La imagen puede ser de baja calidad debido a la presencia de ruido o a una mala adquisición de la misma.

Muchas veces se emplea en la segmentación visual el conocimiento previo que se tiene sobre una determinada región. Este conocimiento previo no es sencillo de obtener ni de ponerlo en estructuras aplicables a algoritmos.

2. Técnicas de Segmentación

En cuanto a la estrategia utilizada para realizar el proceso, las técnicas de segmentación pueden clasificarse en [8,9]:

- Técnicas de segmentación basadas en el pixel.
- Técnicas de segmentación basadas en regiones.
- Técnicas de segmentación basadas en la detección de bordes y fronteras.

Las técnicas de segmentación también pueden clasificarse en función del grado de interacción en el proceso:

- Manual: El usuario realiza la segmentación con la ayuda de una herramienta informática.
- Automática: El ordenador realiza todo el proceso de forma automática.
- Semi-automática o interactiva: El ordenador realiza el proceso, pero el usuario interviene en ocasiones para definir parámetros o corregir resultados.

2.1. Técnicas de Segmentación basadas en el pixel

La segmentación basada en el píxel considera solamente el valor de éste y no ningún tipo de relación con los demás píxeles de la imagen. Dentro de estas técnicas se encuentran la umbralización [10] y los algoritmos de búsqueda de agrupaciones.

2.1.1. Umbralización

La umbralización es un proceso que permite convertir una imagen de niveles de grises o de color, en una imagen binaria, de tal forma que los objetos de interés se etiqueten con un valor distinto de los píxeles de fondo.

Esta es la forma más sencilla de segmentar una imagen en el caso de que se tenga un objeto sobre un fondo en el que contrasta. El procedimiento consiste en establecer un umbral, e ir comparando cada uno de los píxeles de la imagen con este umbral. De esta forma, se tendrá claramente distinguido fondo y objeto.

El principal problema es encontrar el umbral que nos permita obtener la segmentación correcta.

Para determinar dicho umbral se emplea el histograma que, como ya se sabe, indica el número de puntos de la imagen que poseen un determinado nivel de gris. Consiste en elegir el punto de nivel de gris que separa los valores correspondientes al objeto y al fondo. En imágenes con ruido es muy difícil encontrar ese punto.

Se pueden presentar dos casos:

- ◆ Imagen con dos niveles de gris dominantes: umbralización bimodal.

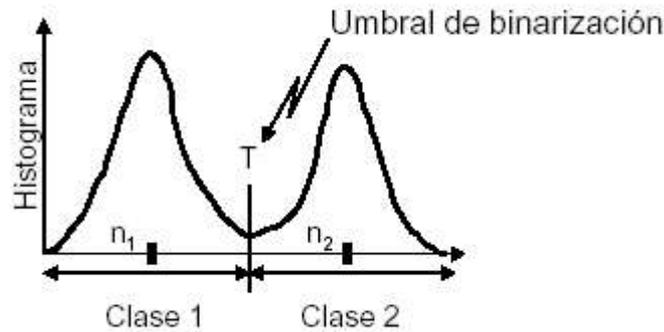


Figura 4.2: Ejemplo de umbralización en histogramas donde aparecen dos lóbulos.

- ◆ Imagen de varios niveles de grises dominantes: umbralización multinivel.

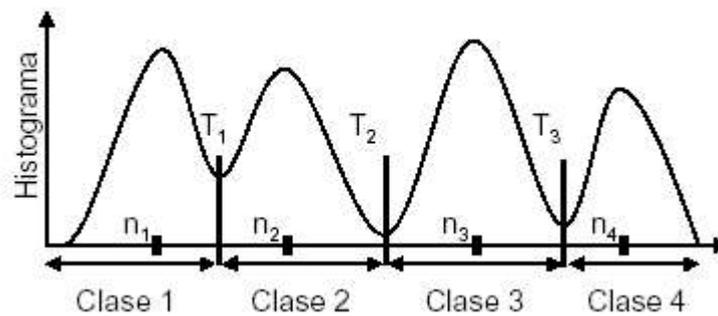


Figura 4.3: Ejemplo de multiumbralización.

2.1.1.1. Umbralización global simple

Es la técnica más simple. En este caso se tiene un único umbral T , que se va a mantener constante en toda la imagen. Es decir, este umbral sólo va a depender de la imagen que se esté tratando, y no se tendrá en cuenta la información espacial.

Esta técnica consiste en dividir la imagen en una imagen binaria, de forma que se obtenga dos partes bien diferenciadas: objeto y fondo. Si el nivel de gris del fondo se mantiene razonablemente constante y el objeto mantiene un contraste sobre el fondo, este tipo de segmentación da buenos resultados, siempre que el nivel de gris haya sido escogido apropiadamente.

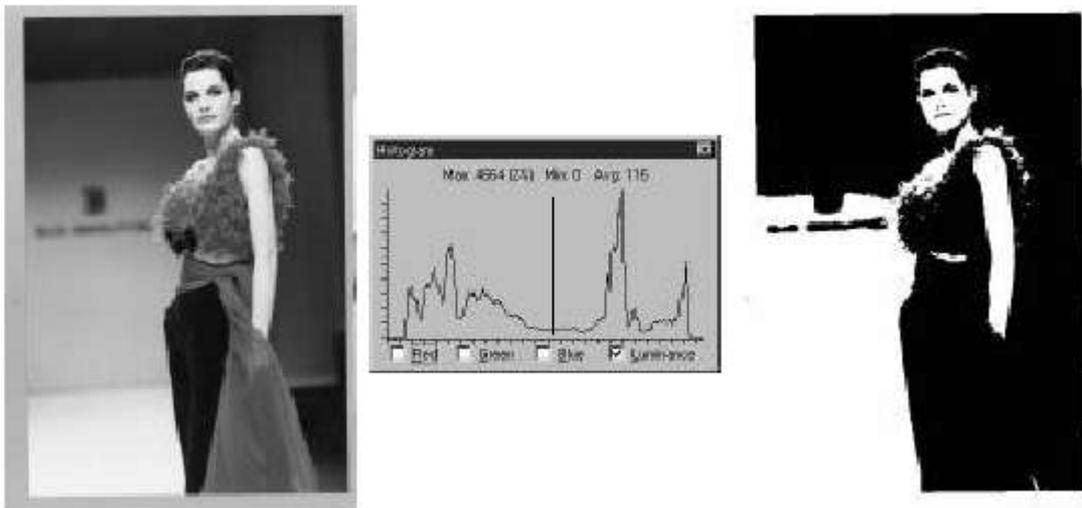


Figura 4.4: Ejemplo de umbralización global simple.

Esta técnica se utiliza principalmente en aplicaciones industriales, ya que el entorno está muy controlado y no se crean sombras que puedan alterar el histograma.

Los principales problemas de este tipo de umbralización son tres:

- Si se tienen histogramas ruidosos, es difícil encontrar los puntos significativos.
- Si se tienen objetos pequeños, estos producirán picos pequeños, que pueden ser escondidos por su entorno.
- No se tiene en cuenta la información espacial.

2.1.1.2. Umbralización de mínimo error u óptima

El objetivo de este método es calcular un umbral tal que el número de píxeles mal clasificados sea mínimo. Para este método es necesario conocer a priori las funciones de densidad de probabilidad de los píxeles del objeto (f_o) y del fondo (f_h).

Sea T un umbral, se define el error de clasificación de un píxel del objeto como:

$$\int_{-\infty}^T f_o(x) dx$$

y el error de clasificación de un objeto del fondo:

$$\int_T^{\infty} f_F(x) dx$$

Por tanto el error que se comete al definir el umbral es:

$$E(T) = p_1 \cdot \int_{-\infty}^T f_F(x) dx + (1 - p_1) \cdot \int_T^{\infty} f_F(x) dx$$

donde p_1 es la probabilidad de que un píxel pertenezca al objeto.

Si se minimiza el error: $\frac{\partial E}{\partial T} = 0$, particularizando para el valor T , se obtiene la

relación que indicará el valor óptimo del umbral:

$$p_1 \cdot f_o(T) = (1 - p_1) \cdot f_F(T)$$

El problema fundamental es que hay que conocer las funciones de densidad de probabilidad a priori.

2.1.1.3. Umbralización por el método Otsu

Es una técnica de umbralización adaptativa que permite separar dos clases entre sí. Esta técnica busca maximizar la distancia entre las dos clases, así como minimizar el área asociada a cada una de ellas. Una de las ventajas de este método es que no son necesarias las funciones de densidad de probabilidad y que el cálculo del umbral se realiza de manera totalmente automática.

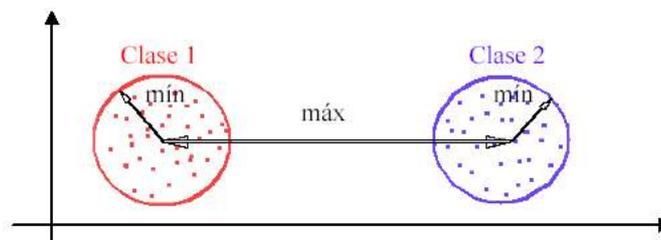


Figura 4.5: Esquema del método Otsu

Para obtener el umbral según este método, hay que seguir los siguientes pasos:

- Se calcula el histograma de la imagen (f_i) que se quiere umbralizar: $h(f_i)$
- Se calcula la media $m_1(K)$, de los píxeles que pertenecen a la clase 1, y cuyos niveles de intensidad están por debajo de un umbral K .

$$m_1(k) = \sum_{i=1}^k ih(f_i)$$

- Se calcula la media $m_2(K)$, de los píxeles que pertenecen a la clase 2, y cuyos niveles de intensidad están por encima de un umbral K .

$$m_2(k) = \sum_{i=k+1}^{f_{\max}} ih(f_i)$$

- Se calcula la probabilidad de que un píxel pertenezca a la clase 1: $P_1(K)$

$$P_1(K) = \sum_{i=0}^k h(f_i)$$

- Se calcula la probabilidad de que un píxel pertenezca a la clase 2: $P_2(K)$

$$P_2(K) = \sum_{i=k+1}^{f_{\max}} h(f_i)$$

Teniendo en cuenta que $P_1(K) + P_2(K) = 1$.

- Se calcula la varianza entre las dos clases: σ^2

$$\sigma^2 = \frac{[m_1(K) \cdot P_2(K) - m_2(K) \cdot P_1(K)]}{P_1(K) \cdot P_2(K)}$$

Finalmente hay que maximizar σ^2 respecto el umbral K .

Las principales desventajas de este método son:

1. Aunque se presupone que sólo es necesario conocer la media y la varianza, hay casos en los que esto no es cierto.
2. El método no funciona cuando se trata de clases con tamaños muy diferentes.

3. El método supone que el histograma de la imagen es bimodal, es decir que la imagen tiene dos clases. Para más clases, es necesario hacer alguna modificación para definir múltiples umbrales.

2.1.1.4. Umbralización adaptativa

Esta técnica implementa la umbralización en dos pasos. En un principio, se divide la imagen en sectores. En el histograma de cada sector, se determina como umbral el valor equidistante entre el pico del fondo y el pico del objeto. Aquellos sectores cuyos histogramas son unimodales se ignoran en los siguientes pasos.

En el primer paso, los límites del objeto son definidos utilizando un umbral constante para cada sector, diferente en cada uno de ellos si fuese necesario.

En el segundo paso, a cada objeto se le da su umbral, que está en la mitad entre los valores del interior del objeto y el nivel de gris del fondo de su sector.

2.1.1.5. Umbralización basada en varias variables.

Anteriormente se han visto las técnicas de umbralización de imágenes basadas en una única variable de intensidad. Ahora se plantea el caso de tener que calcular umbrales basándonos en varias variables. Este caso corresponde al estudio de imágenes en color.

En este caso se trabaja con histogramas tridimensionales, correspondiendo cada dimensión a un color R, G o B, ya que cada píxel posee tres valores. Ahora el problema de encontrar el umbral se complica.

2.1.2. Algoritmos de búsqueda de agrupaciones

Los métodos que pertenecen a este grupo, se caracterizan porque se basan en la propiedad de que generalmente los colores en una imagen tienden a formar agrupaciones en el histograma.

Estas técnicas calculan el histograma a partir del color de los píxeles, y posteriormente se calcula la forma de las agrupaciones. Algunos de los algoritmos más importantes son: *K-means* y *Fuzzy K-means*.

2.2 Técnicas basadas en detección de bordes y fronteras.

Consiste en localizar los puntos de la imagen donde exista una discontinuidad o borde, entendiendo por borde la frontera entre regiones con propiedades de gris, color o textura distintos. La forma más común de localizar estas discontinuidades consiste en el uso de máscaras. Si tomamos como ejemplo una máscara de tamaño 3x3 su forma general sería:

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

Figura 4.6: Máscara general 3x3

Ahora se va desplazando la máscara por la imagen, y cada píxel se sustituye por la suma de los productos de los coeficientes w_i por los niveles de grises de la imagen original:

$$R = w_1 z_1 + w_2 z_2 + w_3 z_3 + w_4 z_4 + w_5 z_5 + w_6 z_6 + w_7 z_7 + w_8 z_8 + w_9 z_9 = \sum w_i z_i$$

donde z_i corresponde con el nivel de gris que posee el píxel que se encuentra asociado al coeficiente w_i de la máscara.

Se puede distinguir tres modelos básicos en detección de discontinuidades, dependiendo de lo que se quiera localizar. Los tres métodos son:

- Detección de puntos.
- Detección de líneas.
- Detección de bordes.

2.2.1. Detección de puntos

Este método permite detectar puntos aislados de una imagen. Para ellos se utiliza una máscara cuyo punto central sea de valor muy alto, y el resto muy bajo. Si se continúa con el caso 3x3, la máscara sería de la siguiente forma:

-1	-1	-1
-1	8	-1
-1	-1	-1

Figura 4.7: Máscara para detectar puntos aislados (3x3)

Se dice que se ha detectado un punto en la localización donde está centrada la máscara, si $R > T$, donde R es el valor calculado mediante la expresión anterior, y T un umbral no negativo.

2.2.2. Detección de líneas

Ahora lo que se quiere no es detectar puntos aislados, sino una línea completa de puntos. Es decir ahora se utilizará una máscara cuya respuesta más elevada se dé en una dirección determinada. Dependiendo de la línea que se quiera detectar, se utilizará una máscara u otra.

Las más utilizadas son:

-1	-1	-1
2	2	2
-1	-1	-1

Detecta línea horizontal

-1	-1	2
-1	2	-1
2	-1	-1

Detecta líneas a 45°

-1	2	-1
-1	2	-1
-1	2	-1

Detecta líneas verticales

2	-1	-1
-1	2	-1
-1	-1	2

Detecta líneas a -45°

Si se pasan las cuatro máscaras por la imagen, se obtendría para cada punto cuatro valores de R : R_1, R_2, R_3, R_4 , respectivamente.

- ◆ Si $|R_1| > |R_i|$, con $i = 2, 3, 4$, se dice que el punto puede pertenecer a una línea horizontal.
- ◆ Si $|R_2| > |R_i|$, con $i = 1, 3, 4$, se dice que el punto puede pertenecer a una línea orientada a 45° .
- ◆ Si $|R_3| > |R_i|$, con $i = 1, 2, 4$, se dice que el punto puede pertenecer a una línea vertical.

- ◆ Si $|R_4| > |R_i|$, con $i = 1, 2, 3$, se dice que el punto puede pertenecer a una línea orientada -45° .

2.2.3. Detección de bordes

Es el método que más suele usarse para detectar discontinuidades ya que se consigue localizar el borde completo y no sólo una parte de él como podría ocurrir en el caso anterior.

Es importante señalar que para este tipo de detección se debe suponer que la región a segmentar es lo suficientemente homogénea como para que la transición entre dos regiones pueda ser determinada basándose en el cambio del nivel de gris de la discontinuidad. Esto implica que no serviría para imágenes con mucho ruido ya que la homogeneidad requerida no se conseguiría.

Para estas técnicas se utilizan principalmente operadores que se basan en la primera y/o segunda derivada (gradiente y laplaciano). Hay que tener en cuenta que al tratarse de derivadas, estas técnicas son demasiado sensibles a imágenes con ruido por lo que se necesita que las regiones se encuentren claramente diferenciadas, o que se complementen con otros algoritmos que se encarguen de realizar la unión de los píxeles detectados mediante el seguimiento del contorno.

2.3 Técnicas basadas en regiones

Las técnicas agrupadas bajo el nombre de crecimiento de regiones determinan zonas dentro de una imagen basándose en criterios de similaridad y proximidad entre los píxeles de la misma[7].

En estas técnicas, la homogeneidad (o falta de homogeneidad) entre regiones adyacentes es el criterio utilizado para unir (o dividir) regiones de la imagen. Dicha homogeneidad se puede definir a partir de criterios como el nivel de gris medio, el color, la forma, etc.

El resultado de la segmentación es una partición de la imagen en regiones homogéneas. Las técnicas basadas en regiones se utilizan con frecuencia en imágenes con

ruido.

2.3.1. Fundamentos de la segmentación basada en regiones

Una region, en una imagen, es un grupo de píxeles conectados que tienen propiedades similares o verifican una determinada característica. Es obvio que las regiones son importantes para la interpretación de las imágenes, pues pueden corresponder a objetos en la escena.

Una imagen puede contener varios objetos y además cada objeto puede contener varias regiones que corresponden a partes del mismo.

El criterio más simple de homogeneidad utiliza el nivel de gris medio de la región, algunas propiedades simples o un vector de medias para imágenes multibanda. En el método de segmentación basado en píxeles, la idea fundamental es clasificar un punto como del objeto, lo que conlleva que puntos aislados o pequeñas áreas puedan ser clasificadas como pertenecientes a la región de interés, es decir, allí no se supone la conectividad como característica importante del objeto.

Si R denota una determinada región subdividida en n regiones, las características que definen una imagen son las siguientes:

- R_i es una región independiente donde $i=1,2,3,\dots,n$.
- La intersección de dos regiones distintas es 0.
- Todos los píxeles de una determinada región deben verificar una misma condición.
- Cada región se caracteriza por una condición diferente.
- Un píxel debe pertenecer a una región, es decir, no se pueden tener píxeles aislados.

2.3.2. Métodos de generación de regiones

2.3.2.1. Crecimiento de regiones por adición de píxeles

Como su nombre indica, este procedimiento agrupa píxeles de la imagen con características similares para formar regiones.

Este tipo de algoritmo necesita que el usuario seleccione un conjunto de puntos semilla en la imagen. Estos puntos semillas servirán como puntos de comienzo del proceso de crecimiento de las regiones, y se le irán agregando píxeles adyacentes si tienen valores que cumplen cierto criterio definible de homogeneidad con los puntos “semilla”. Si ocurre esto, pasan a formar parte de esa región, y toman los mismo valores que los puntos “semillas”.

El proceso continúa hasta que no puedan añadirse más píxeles a la región. Una vez que esto sucede se elige otro punto semilla que no pertenezca a ninguna región segmentada y se repite el proceso.

Para poder realizar la agregación de píxeles similares será necesario definir el concepto de similaridad, que no tiene porque ser el mismo para todo tipo de aplicaciones. Posibles criterios, ya utilizados en algoritmos desarrollados, pueden ser que la diferencia entre el valor del pixel a agregar y el valor de la semilla, o el valor medio de la región ya formada, sea menor que un cierto umbral predeterminado.

El número final de regiones ha de ser como mucho igual al número de semillas sembradas por el usuario (puede ser menor, pues el algún paso del algoritmo se puede decidir unir dos regiones para formar una sola).

Por tanto, un pixel se añade a una determinada region si se cumple dos características:

- ◆ Es similar a los píxeles de esa region.
- ◆ Es adyacente a la región. Los posibles vecinos de un pixel, pueden verse en la figura 4.8.

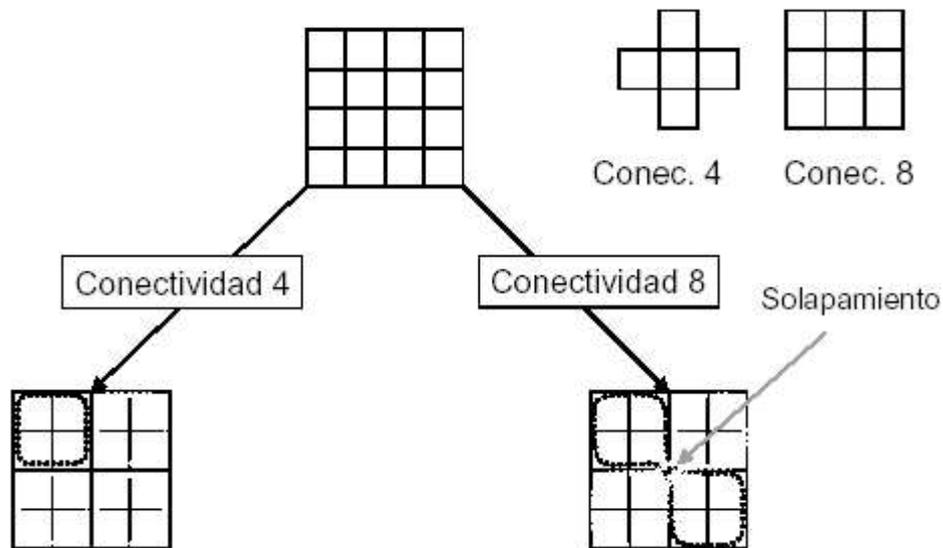


Figura 4.8: El entorno de un píxel semilla puede considerarse como una vecindad de 4 vecinos o 8

A continuación, se muestra en la figura 4.9 cómo se van recorriendo los píxeles, para ir formando las regiones. Se parte de los píxeles semillas, y a partir de ellos se comprueba si sus vecinos pertenecen a su región. A partir de cada uno de los píxeles vecinos se sigue el crecimiento. Una vez que se ha recorrido toda la imagen, ya se tienen las regiones en las que la imagen ha quedado dividida.

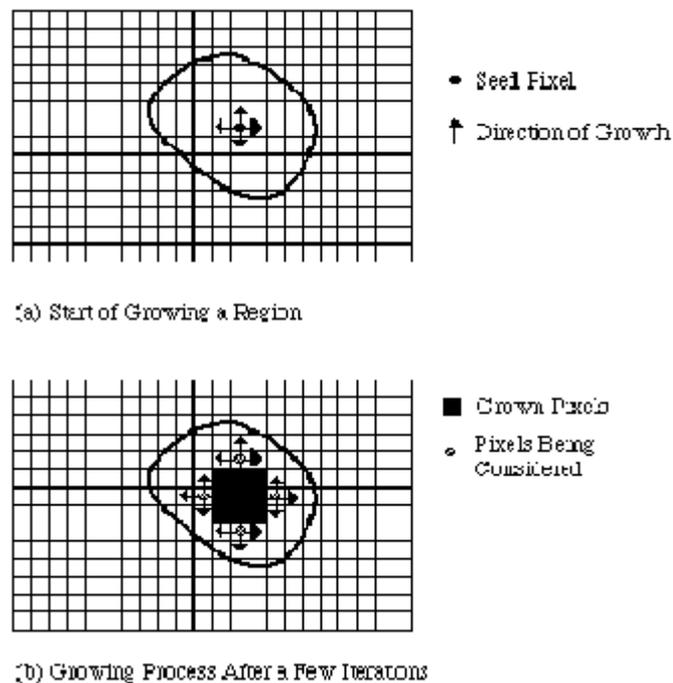


Figura 4.9: Movimiento de píxeles semilla

Por último, se muestra un ejemplo sencillo, que detalla esta técnica de segmentación:

0	0	5	6	<i>A</i>	<i>A</i>	<i>B</i>	<i>B</i>
1	1	8	7	<i>A</i>	<i>A</i>	<i>B</i>	<i>B</i>
2	0	5	6	<i>A</i>	<i>A</i>	<i>B</i>	<i>B</i>
1	1	6	5	<i>A</i>	<i>A</i>	<i>B</i>	<i>B</i>

Si las anteriores matrices expresan valores de gris en una escena, si se tienen dos semillas, una en 2 y la otra en 8, las regiones quedarán definidas por sus características similares según indica la segunda matriz. En este ejemplo se ha empleado como característica una diferencia no superior a 4 unidades en los valores.

Este algoritmo presenta dos problemas fundamentales:

- Selección de la propiedad de inicialización y selección de las semillas adecuadas.

Cuando no se conoce información de la imagen, normalmente se toman como nodos de inicio aquellos con mayor nivel de brillo. Un estudio previo del histograma es capaz de dar una leve idea de los niveles de gris más importantes y por tanto se puede emplear como fuente de información para la inicialización de los píxeles.

- Selección de la propiedad que hace finalizar el algoritmo de crecimiento.

Se podría pensar que las regiones paran cuando no hay más píxeles vecinos que verifiquen un cierto criterio; pero esto no es del todo cierto, puesto que salvo en presencia de contornos, las características suelen ser aproximadas para nodos adyacentes. Se debe emplear un criterio adicional para limitar las regiones: el gradiente de crecimiento. Cuando los crecimientos no son rápidos, las características difieren algo respecto de la condición establecida. A partir de aquí se pueden finalizar las regiones.

2.3.2.2. División de regiones

En cierta forma, es un proceso opuesto al de unión de regiones. Se parte de una única región que representa toda la imagen. Si dicha región no satisface el criterio de homogeneidad establecido, la región inicial se divide, de manera secuencial, en subregiones de las que se estudia su homogeneidad. Si una subregión está formada por puntos homogéneos, esta subregión no se subdivide, en caso contrario se sigue subdividiendo. Los criterios utilizados para dividir regiones son similares a los usados para agruparlos, y solo se diferencian en la dirección en que se aplican.

2.3.2.3. División y Fusión de regiones (Split and Merge)

Es un método que consiste en dividir la imagen en regiones uniformes. Necesita partir de una imagen de tamaño $N \times N$, siendo N un número potencia de dos, para poder dividir cada región en cuatro partes iguales.

Se basa en una estructura de datos denominada QPT (Quadric Picture Tree). Normalmente el algoritmo comienza con la hipótesis de que la imagen completa es una única región, entonces analiza la homogeneidad de la misma (mediante un cierto criterio y propiedades). Si existe homogeneidad, la imagen se encuentra ya segmentada, si no es así, entonces la región es dividida en 4 regiones. Este proceso se repite para cada una de las regiones generadas hasta que el proceso de división no puede llevarse a cabo.

Una vez que se ha llevado a cabo el proceso de división, se comprueba para cada región generada, si es posible unirla con una región adyacente (lógicamente si satisfacen el criterio de homogeneidad establecido). El proceso termina cuando no se pueden fusionar más regiones.

Por ejemplo, para una imagen como la representada en la figura 4.10, el árbol de las divisiones sería el que se encuentra en la figura 4.11, donde la propiedad elegida es que los píxeles del mismo bloque tengan el mismo color.

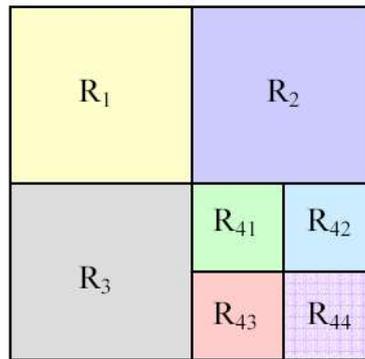


Figura 4.10: Imagen ya dividida según la condición de que los píxeles de los bloques sean del mismo color.

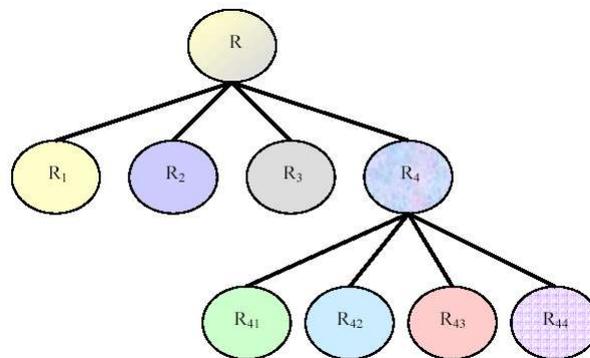


Figura 4.11: QPT de la imagen de la figura 4.10 misma intensidad o color