

## Capítulo 2: Métodos tradicionales de Segmentación Estéreo

---

## 2.1. Introducción

La segmentación estéreo incluye los siguientes conceptos:

- Segmentación de imágenes: En visión por computador, la segmentación de una imagen digital es el proceso que divide la imagen en regiones u objetos cuyos píxeles poseen atributos similares. Cada región segmentada tendrá un significado físico dentro de la imagen.
- Visión estéreo: La visión estereoscópica es el proceso por el cual se reconstruye un escenario 3D a partir de dos imágenes 2D, gracias al cual se puede recuperar la tercera dimensión (profundidad), perdida en las imágenes bidimensionales.

En los siguientes apartados se describen distintos métodos empleados tanto en la segmentación, como en la visión estéreo.

Adicionalmente, se realiza un estudio del estado del arte de la segmentación estéreo.

## 2.2. Segmentación de imágenes

En general, la segmentación es una de las tareas más difíciles del procesamiento de una imagen y determina el éxito o fracaso de dicho análisis. Es el proceso que permite extraerlos objetos de la imagen para su posterior descripción y reconocimiento. Numerosos artículos se encargan de intentar que los resultados sean un éxito y de hacerlo de la manera más eficiente.

Existen varios tipos de segmentaciones, entre los que podemos encontrar de forma más simple a más complicada los siguientes:

- Detección de discontinuidades
- Unión de bordes y detección de fronteras
- Umbralización
- Segmentación orientada a regiones
- Modelos deformables

### 2.2.1. Detección de discontinuidades

Existen tres tipos básicos para detectar discontinuidades del nivel de gris de una imagen. Estos tres tipos son:

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

Estos tres tipos trabajan de la misma forma, se genera una imagen donde cada píxel es el resultado ( $R$ ) de la suma de los productos de los cocientes de la máscara por los valores de la imagen cuando la máscara está centrada en la posición del píxel que se está calculando. Es decir, sigue la ecuación 1 (Ec.2.1) y un ejemplo de una matriz se puede ver en la figura 2.1 (Fig. 2.1)

$w_1$	$w_4$	$w_7$
$w_2$	$w_5$	$w_8$
$w_3$	$w_6$	$w_9$

Fig.2.1 Máscara 3x3

$$R = \sum_{i=1}^9 w_i z_i = w_1 z_1 + w_2 z_2 + \dots + w_9 z_9 \quad (\text{Ec. 2.1})$$

Donde  $z_i$  es el nivel de gris asociado con el coeficiente de la máscara  $w_i$

La máscara variará el valor de sus elementos dependiendo del tipo de discontinuidad que se quiera calcular.

### Detección de puntos

La detección de puntos en una imagen es el método más simple. Usando la máscara de la figura 2.2 (Fig.2.2) se puede decir que un punto ha sido detectado si:

$$|R| \geq T \quad (\text{Ec. 2.2})$$

Donde  $R$  es el valor calculado con la ecuación 2.1 y  $T$  es un umbral positivo.

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

Fig.2.2 Máscara para detectar puntos

Cabe destacar que  $R$  será igual a 0 si los píxeles vecinos tienen el mismo valor de gris que el píxel central.

### Detección de líneas

Este método es un poco más complicado que el anterior pero menos que el siguiente. Existen distintas máscaras dependiendo de la orientación de la línea que se desea detectar, pudiendo ser horizontal,  $+45^\circ$ , vertical y de  $-45^\circ$ . La figura 2.3 (Fig. 2.3 a,b,c y d) representa respectivamente cada máscara.

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

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

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

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

Fig 2.3 (a) Horizontal

(b)  $+45^\circ$

(c) Vertical

(d)  $-45^\circ$

Un ejemplo de la aplicación de la máscara de detección de líneas se puede ver en la figura 2.4 (Fig.2.4) donde se puede ver la imagen original y la obtenida al aplicarle la máscara de  $-45^\circ$ .



Fig.2.4 Imagen original y tras aplicarle la máscara de  $-45^\circ$

### Detección de bordes

La idea en que se basan la mayoría de las técnicas de detección de bordes es el cálculo de un operador de derivada local. Dentro de este apartado se debe diferenciar entre:

- Operadores basados en la primera derivada (o gradiente)
- Operadores basados en la segunda derivada (o Laplaciana)

La figura 2.5 (Fig.2.5) representa un borde de una función bidimensional y sus correspondientes derivadas. Se puede apreciar que la primera derivada es positiva y constante si la rampa que representa el borde aumenta de izquierda a derecha mientras que vale cero en las áreas que existe un nivel constante de gris. La segunda derivada, en cambio, es positiva en la transición asociada con el nivel de gris más oscuro y negativa en la transición asociada con el nivel de gris más claro, además valdrá cero a lo largo de la rampa que

representa el borde y en las zonas con un nivel de gris constante. En resumen se puede decir que la primera derivada proporciona la dirección del borde, mientras que la segunda derivada indica si los píxeles implicados pertenecen a la zona clara u oscura del borde.

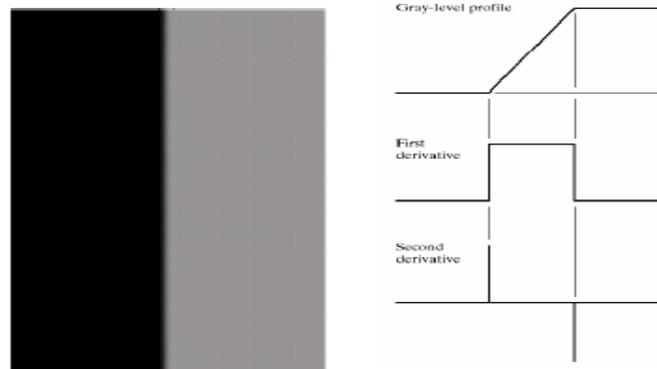


Fig.2.5 Representación de un borde con sus correspondientes derivadas

- **Operadores basados en la primera derivada (o gradiente)**

En el caso de funciones bidimensionales  $f(x,y)$ , la derivada es un vector que apunta en la dirección de máxima variación de  $f(x,y)$  y cuyo módulo es proporcional a dicha variación. Este vector, denotado como  $\nabla f(x,y)$  se denomina gradiente y se define como:

$$\nabla f(x,y) = \begin{bmatrix} \frac{\partial}{\partial x} f(x,y) \\ \frac{\partial}{\partial y} f(x,y) \end{bmatrix} = \begin{bmatrix} f_x(x,y) \\ f_y(x,y) \end{bmatrix} = \begin{bmatrix} G_F \\ G_C \end{bmatrix} \quad (\text{Ec. 2.3})$$

Donde  $G_F$  representa el gradiente de fila y  $G_C$  el gradiente de columna.

El gradiente de fila y de columna en cada punto se obtiene mediante la convolución de la imagen con las máscaras  $H_c$  y  $H_f$ , es decir:

$$\begin{aligned} G_F(i,j) &= F(i,j) \otimes H_F(i,j) \\ G_C(i,j) &= F(i,j) \otimes H_C(i,j) \end{aligned} \quad (\text{Ec. 2.4})$$

Donde  $H_c$  y  $H_f$  dependerán del operador seleccionado.

La magnitud y orientación del vector gradiente se obtiene de acuerdo a la ecuación 2.5, aunque debido a su elevado coste computacional suele usarse la aproximación de la ecuación 2.6.

$$|G(i, j)| = \sqrt{G_F^2 + G_C^2} \quad (\text{Ec. 2.5})$$

$$|G(i, j)| \approx |G_F(i, j)| + |G_C(i, j)| \quad (\text{Ec. 2.6})$$

Los distintos operadores que se suelen usar en la detección de bordes son los siguientes:

- Roberts: Buenas prestaciones en cuanto a localización pero muy sensible al ruido.
- Prewitt: Se involucran los vecinos de filas y columnas adyacentes para proporcionar mayor inmunidad al ruido. Detecta mejor los bordes horizontales y los verticales.
- Sobel: Más sensible a los bordes diagonales

La figura 2.6 (Fig.2.6) representa las distintas máscaras de los operadores antes mencionados

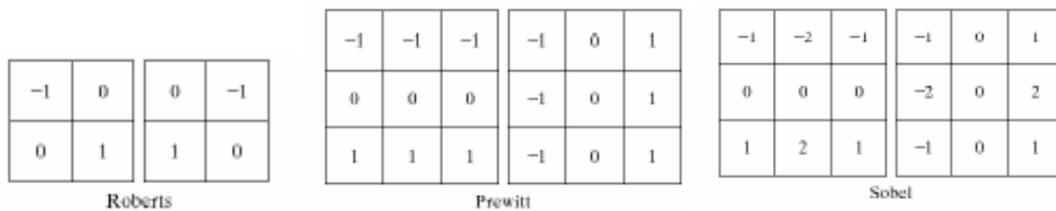


Fig.2.6: Máscaras de los operadores de Roberts, Prewitt y Sobel respectivamente

Un ejemplo del funcionamiento de estos operadores se puede ver en la siguiente figura, donde para la misma imagen se puede ver el resultado de aplicar cada operador. Las siguientes figuras (Fig.2.7 y Fig.2.8) muestran un ejemplo de esta aplicación



Fig.2.7: Imagen Original



Fig.2.8 : Resultados tras aplicar los operadores de Roberts, Prewitt y Sobel respectivamente

- **Operadores basados en la segunda derivada (o Laplaciana)**

Siempre que en la imagen se presenta un cambio de intensidades a lo largo de una determinada dirección, existirá un máximo en la primera derivada. La dirección de interés será aquella ortogonal a la dirección local de los cruces por cero. Puesto que, generalmente, esta dirección no es conocida, parece claro que la utilización de un operador independiente de la orientación como la Laplaciana presenta ventajas computacionales. Así mismo, al ser una derivada segunda, tiene la ventaja de facilitar la localización precisa del borde.

La Laplaciana del borde de una función bidimensional,  $f(x,y)$ , se define como:

$$\nabla^2 f(x,y) = \frac{\partial^2}{\partial x^2} f(x,y) + \frac{\partial^2}{\partial y^2} f(x,y) \quad (\text{Ec. 2.7})$$

La Laplaciana vale cero si  $f(x,y)$  es constante o cambia linealmente su amplitud. El cambio de signo de la función resultante indica que en ese lugar existe un cruce por cero y, por tanto, indica la presencia de un borde. Por tanto, la aproximación discreta de la Laplaciana de la imagen  $f(i,j)$  se obtiene mediante la convolución:

$$L[F(i,j)] = F(i,j) \otimes H(i,j) \quad (\text{Ec. 2.8})$$

Donde  $H(i,j)$  representa la máscara de la Laplaciana. La figura 2.9 (Fig.2.9) representa algunas de las más importantes máscaras de la Laplaciana.

0	-1	0
-1	4	-1
0	-1	0

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

Fig.2.9: Máscaras de Laplaciana de 4 y de 8 vecinos respectivamente

Un ejemplo del funcionamiento de la Laplaciana se puede ver en la siguiente figura (Fig.2.10).



Fig.2.10 Imagen original y el resultado de aplicar la Laplaciana

### 2.2.2. Unión de bordes y detección de fronteras

El método comentado en el apartado anterior genera una imagen píxeles en blanco donde estarían los puntos, líneas o los bordes de la imagen. El problema de este método es que generalmente produce píxeles sin unir, es decir, los bordes no aparecen continuos. Los siguientes métodos se proponen como solución al apartado anterior:

- Unión de bordes mediante seguimiento del contorno
- Unión de bordes mediante la transformada de Hough

#### Unión de bordes mediante seguimiento del contorno

Una de las formas de obtener el contorno de un objeto de la imagen consiste en formar ésta a partir de propiedades similares en la imagen gradiente. Para establecer la similitud entre píxeles vecinos se tienen en cuenta la magnitud y dirección del vector gradiente.

Considerando la magnitud, se dice que un píxel con coordenadas  $(x,y)$  es similar en magnitud a un píxel  $(x',y')$  si cumple la ecuación 2.9. Si consideramos la dirección, los píxeles serán similares si cumplen la ecuación 2.10.

$$|G[f(x,y)] - G[f(x',y')]| \leq T \quad (\text{Ec. 2.9})$$

$$|\alpha(x,y) - \alpha(x',y')| < A \quad (\text{Ec. 2.10})$$

Donde  $T$  y  $A$  son los umbrales para la diferencia de magnitud y de ángulo respectivamente.

La idea de esta técnica consiste en ir extendiendo la frontera en dirección perpendicular al gradiente con píxeles vecinos cuya respuesta del operador gradiente sea similar. El algoritmo termina cuando no se consiguen incorporar píxeles a la frontera, o bien cuando el punto original de partida es reencontrado.

### Unión de bordes mediante la transformada de Hough

La transformada de Hough es una herramienta que permite detectar curvas en una imagen. A la hora de aplicar la transformada de Hough a una imagen es necesario obtener previamente una imagen binaria de los píxeles que forman parte de las fronteras, es decir usar algunos de los métodos comentados anteriormente.

Para calcular la transformada de Hough de forma generalizada se debe usar la representación de una recta normal en polares, indicada en la ecuación 2.11

$$x \cos \theta + y \sin \theta = \tau \quad (\text{Ec. 2.11})$$

Donde  $x, y$  son las coordenadas del píxel y  $\vartheta$  es el ángulo que forma la normal con el eje  $x$  y  $\tau$  es la distancia de la recta al origen.

El pseudocódigo para la detección de rectas mediante la transformada de Hough es el siguiente:

- Discretizar el espacio de parámetros, estableciendo valores máximos y mínimos de  $\tau$  y  $\vartheta$ , así como el número total de valores de  $\tau$  y  $\vartheta$ .
- Generar el acumulador  $A(\tau, \vartheta)$ , poniendo todos los valores a cero.
- Para todos los puntos del borde  $(x_i, y_i)$ 
  - Calcular la dirección del vector gradiente  $\vartheta$
  - Obtener  $\tau$  de la ecuación 2.11
  - Incrementar  $A(\tau, \vartheta)$
- Para todas las celdas en el acumulador
  - Buscar los valores máximos del acumulador
  - Las coordenadas  $(\tau, \vartheta)$  dan la ecuación de la recta en la imagen

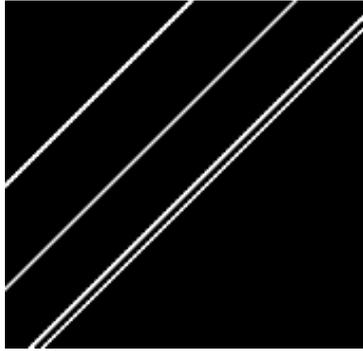
Un ejemplo de la aplicación de la transformada de Hough para detectar rectas de  $+45^\circ$  se puede ver en la figura 2.11 (Fig.2.11)



Fig.2.11 (a) Imagen original



(b) Tras aplicarle detector bordes



(c) Tras aplicar la transformada de Hough



(d) Superposición de (b) y (c)

### 2.2.3. Umbralización

Si a una imagen compuesta de un objeto claro sobre fondo oscuro (o al revés) se le calcula su histograma quedará algo parecido a la figura 2.12 (Fig.2.12 (a)). Una forma de extraer el objeto de la imagen es seleccionar un umbral de intensidad  $T$  por encima del cual se encuentran todos los píxeles pertenecientes al objeto. De este modo, cualquier punto  $(x,y)$  para el cual  $f(x,y) > T$  es un punto del objeto, mientras que si  $f(x,y) \leq T$  será un punto del fondo. En la práctica se necesitan más de un umbral para dividir la imagen (Fig. 2.12(b))

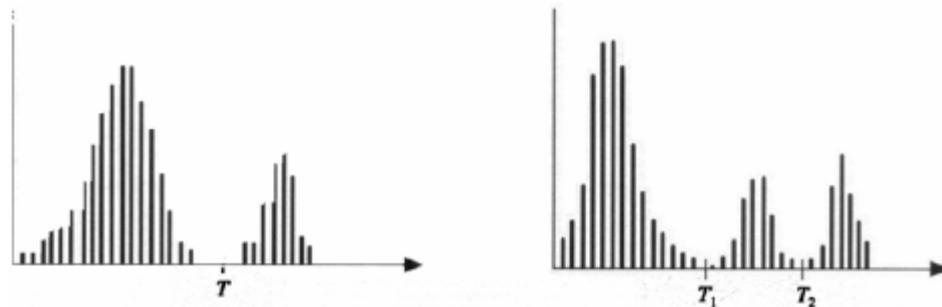


Fig.2.12 (a) Histograma con un único umbral

(b) Histograma con varios umbrales

De forma general, se puede decir que en una imagen umbralizada  $g(x,y)$  vale 1 si  $f(x,y) > T$  y vale 0 en caso contrario, es decir:

$$g(x,y) = \begin{cases} 1, & \text{si } f(x,y) > T \\ 0, & \text{si } f(x,y) \leq T \end{cases}$$

Cuando  $T$  depende solamente de  $f(x,y)$  se denomina *global*. Si  $T$  depende de  $f(x,y)$  y de alguna propiedad local de ese punto, el umbral se llama *local*. Si además depende de las coordenadas  $x$  e  $y$  se llama umbral *dinámico*.

La siguiente imagen (Fig.2.13) muestra un ejemplo de la umbralización global

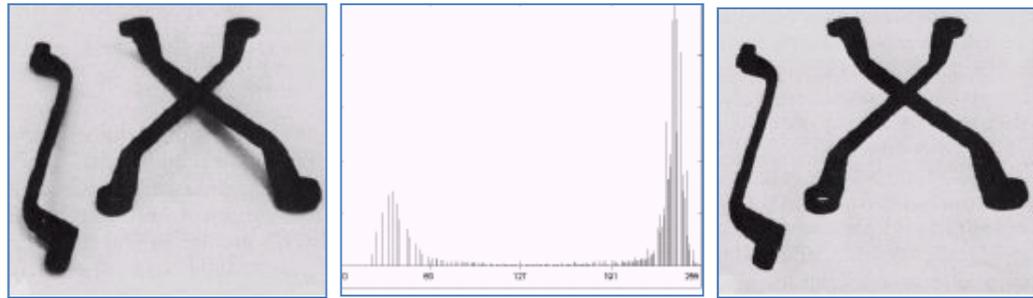


Fig.2.13: (a) Imagen original (b) Histograma bimodal (c) Imagen umbralizada

Dependiendo de la imagen que se tenga y de su histograma se usará un tipo de umbral u otro. En la elección del tipo de umbral también influirán condiciones ambientales como la iluminación, siendo el umbral dinámico al que menos le influyen las condiciones de iluminación, aunque sea el más complejo de calcular.

#### 2.2.4. Segmentación orientada a regiones

Las técnicas de este método están basadas en la búsqueda de regiones directamente.

Sea  $R$  una región que representa a la imagen completa. La segmentación de la imagen se puede ver como un proceso que divide a  $R$  en  $n$  subregiones,  $R_1, R_2, \dots, R_n$ , tal que:

$$(a) \bigcup_{i=1} R_i = R$$

(b)  $R_i$  con  $i=1,2,\dots,n$  es una región conectada

(c)  $R_i \cap R_j = 0$ , para todo  $i$  y  $j$  con  $i$  distinto de  $j$

(d)  $P(R_i)=TRUE$  para  $i=1,2,\dots,n$

(e)  $P(R_i \cup R_j)=FALSE$  para  $i$  distinto de  $j$

Donde  $P(R_i)$  es un predicado lógico definido sobre los puntos del conjunto  $R_i$ , basado en alguna medida de similitud.

La condición (a) indica que la unión de todas las regiones obtenidas tras la segmentación debe ser la imagen completa. La condición (b) requiere que los puntos de una región estén conectados. La condición (c) indica que las regiones han de ser disjuntas. La condición (d) indica que los píxeles de una región segmentada deben satisfacer unas determinadas propiedades. Finalmente, la condición (e) indica que las regiones  $R_i$  y  $R_j$  son distintas según el criterio del predicado  $P$

Siguiendo estas condiciones, se encuentran dos métodos distintos:

- Crecimiento de regiones mediante adición de píxeles
- División y fusión de regiones

### Crecimiento de regiones mediante adición de píxeles

El crecimiento de regiones es un procedimiento que agrupa píxeles o subregiones en regiones más grandes. El método más simple es la adición de píxeles, donde se parte de un conjunto de puntos semilla a los que se les van añadiendo píxeles vecinos que poseen propiedades similares. Si se usan  $n$  semillas, al final se podrá obtener una segmentación con un máximo de  $n$  regiones, además del fondo.

Esta técnica plantea dos problemas fundamentales:

- La colocación de las semillas
- La elección de las propiedades para la inclusión de píxeles en cada región
- La formulación de la regla de parada.

El método expuesto tiene el siguiente algoritmo recursivo:

- Tomar  $p$  como semilla inicial de crecimiento
- Si el criterio de selección se cumple para ' $p$ ', entonces añadir dicho punto a la región de crecimiento.
- Si  $p$  ha sido añadido, entonces para cada uno de sus 4 vecinos ( $p_i$ ) que estén fuera de la región:
  - $p = p_i$
  - Ir al paso anterior

La siguiente imagen (Fig.2.14) muestra un ejemplo de este método donde se puede ver la imagen original (a), la imagen con las semillas (b) y el resultado final(c):

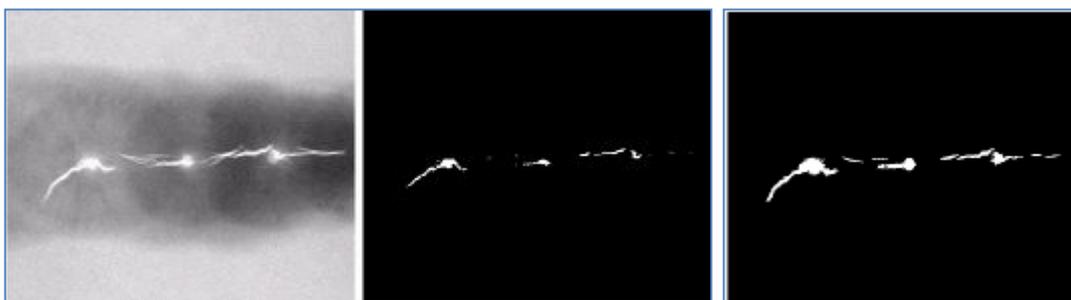


Fig.2.14 (a) Imagen original

(b) Imagen con las semillas

(c) Resultado final

### División y fusión de regiones

Esta técnica se basa en la subdivisión inicial de la imagen en un conjunto de regiones arbitrarias y disjuntas. A partir de este conjunto, se comienza un proceso iterativo de fusión y/o división de dichas regiones, sujeto a las condiciones establecidas en la introducción del apartado 2.2.4. Este método tiene el siguiente algoritmo:

1. Sea  $R_0$  la región inicial, constituida por la imagen completa;
2. Seleccionar un predicado  $P$ ;
3. Para toda la región  $R_i$ , tal que,  $P(R_i)=FALSE$  : Subdividir  $R_i$  en cuatro cuadrantes disjuntos;
4. Fusionar cualquier par de regiones adyacentes  $R_j$  y  $R_k$ , para que se verifique  $P(R_k \cup R_j)=TRUE$
5. Si existen más regiones para fusionar o dividir, entonces ir a 3, si no, parar;

Para la representación de las sucesivas subdivisiones se suele usar un árbol cuaternario como el de la siguiente imagen (Fig.2.15), donde se representa una imagen dividida y su correspondiente árbol cuaternario.

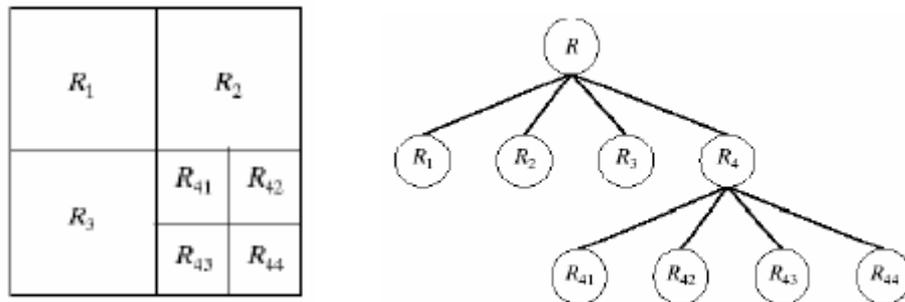


Fig.2.15: (a) Imagen dividida en subregiones (b) Árbol cuaternario correspondiente a (a)

La siguiente imagen (Fig.2.16) muestra un ejemplo de una imagen original, de su descomposición en regiones y el resultado final.

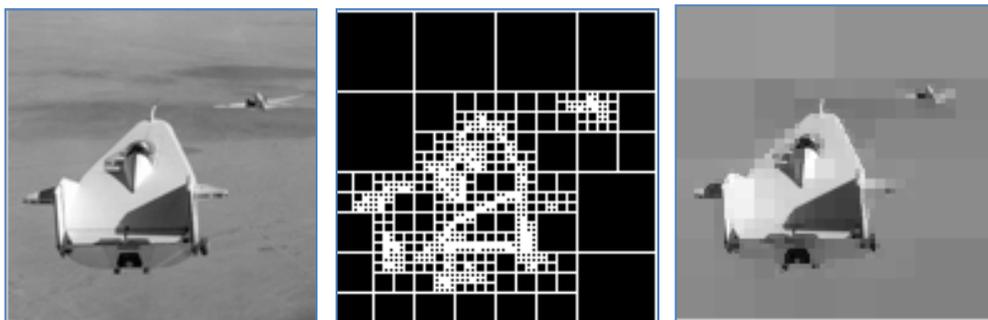


Fig.2.16: Aplicación del método División y Fusión de Regiones

### 2.2.5. Modelos deformables

Los modelos deformables o de contornos activos, conocidos como snakes debido a sus características de movimiento y evolución sobre la imagen, fueron propuestos en Kass et al., [2.1] y desde entonces han sido aplicados a una variedad de problemas de análisis de imágenes y visión computacional.

Este método, es ampliamente explicado en el Capítulo 3: Contornos Activos, de esta memoria.

## 2.3. Visión estéreo

El cerebro humano percibe la realidad tridimensional a través del sistema de visión, que consta de dos ojos desplazados ligeramente uno respecto del otro según la dirección horizontal.

Basándonos en este hecho, los sistemas de percepción artificiales poseen dos cámaras desplazadas entre sí una cierta distancia, gracias a lo cual se puede recuperar la profundidad.

La siguiente imagen (Fig.2.17.) muestra un par de imágenes estereoscópicas obtenidas mediante dos cámaras desplazadas entre sí una cierta distancia horizontal



*Fig.2.17: Par de imágenes estéreo*

Una descomposición del proceso de visión estereoscópica fue realizado por Barnard y Fischer (1982). Según esta descomposición, el proceso completo de visión estéreo contempla seis pasos principales:

1. Adquisición de imágenes
2. Modelado de la cámara (geometría del sistema)
3. Extracción de las características
4. Correspondencia de las imágenes
5. Determinación de la distancia (profundidad)
6. Interpolación, cuando sea necesaria

De estos pasos, al que se reconoce ampliamente como el más complejo de resolver, es el número 4, correspondencia de las imágenes, el cual será explicado con más detalle.

### 2.3.1. Adquisición de imágenes

Las imágenes pueden capturarse de dos formas distintas:

- Utilizando dos o más cámaras, todas ellas alineadas y separadas una cierta distancia. Esta distancia puede ser cualquiera, pero siempre deberá ser conocida con exactitud. La toma de imágenes por ambas cámaras, se realizará en el mismo instante.
- Utilizar únicamente una cámara móvil que se desplazará una cantidad determinada entre la toma de una imagen y la siguiente. Al igual que el caso anterior, el desplazamiento puede ser cualquiera, pero debe ser conocido con exactitud. La diferencia de tiempo entre la toma de las imágenes puede influir negativamente en la correspondencia entre ambas imágenes, ya que dependiendo de la escena, pueden cambiar sus condiciones ambientales y atmosféricas.

### 2.3.2. Modelado de la cámara (geometría del sistema)

El modelo más simple representado se basa en dos cámaras que poseen sus ejes ópticos paralelos, siendo la distancia que separa ambas cámaras su línea base, perpendicular a los ejes ópticos. Cualquier punto del espacio tridimensional  $P(x,y,z)$  unido a los dos centros de proyección de las cámaras define un plano, llamado *plano epipolar*. La intersección de un plano epipolar con el plano de proyección de una cámara define una línea epipolar. Para todos los puntos, cuyas proyecciones izquierdas ( $P_I$ ) estén contenidas en una misma línea epipolar en la imagen izquierda, sus proyecciones derechas ( $P_D$ ) deben estar también contenidas sobre una misma línea epipolar en la imagen derecha, y viceversa. Como se ha dicho anteriormente, el principal problema al realizar visión estereoscópica es encontrar la correspondencia entre los puntos de las imágenes.

La siguiente imagen (Fig.2.18) muestra la representación comentada.

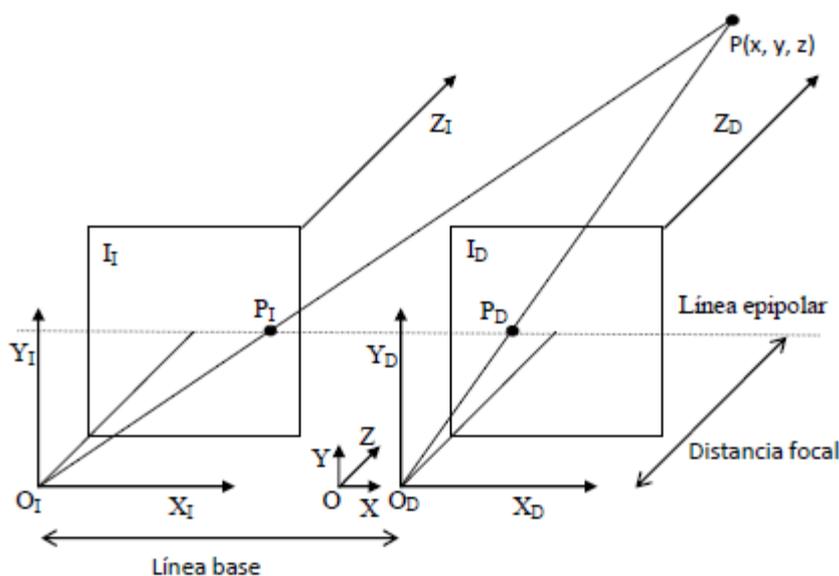


Fig.2.18: Geometría del sistema

### 2.3.3. Extracción de las características

En este paso se obtendrán datos características de las imágenes que ayuden a determinar la correspondencia entre las imágenes, como bordes, esquinas, patrones, etc. Este paso es usado sobre todo si en el paso de correspondencia entre imágenes se emplean métodos basados en características

### 2.3.4. Correspondencia de las imágenes

El paso principal en la visión estéreo es establecer la correspondencia entre la imagen izquierda y la derecha.

Existen diversos métodos para calcular dicha correspondencia, pudiéndose clasificar en los siguientes métodos:

- Métodos locales
  - Basados en área
  - Basados en característica
- Métodos globales

#### Métodos locales

Son métodos que aplican restricciones a un pequeño número de píxeles alrededor globales del pixel de estudio. Suelen ser muy eficientes pero sensitivos a las ambigüedades locales de las regiones como las oclusiones, regiones sin textura o textura repetitiva.

#### Basados en área

Para cada píxel de una imagen se calcula la correspondencia entre la distribución de disparidad de una ventana centrada en dicho píxel y una ventana del mismo tamaño centrada en el píxel a analizar de la otra imagen. El problema consiste en encontrar el punto que se ajusta de manera adecuada al primero.

La restricción epipolar, que consiste en que un punto y su correspondiente en la otra imagen se encuentran situados sobre la misma horizontal, reduce la búsqueda a un problema de búsqueda a una dimensión.

Estos métodos tratan consideran las dos imágenes como un objeto trasladado, tratando de obtener, para cada punto de la imagen, dicha traslación.

Algunos de estos métodos son los siguientes:

Sum of Absolute Differences (SAD)	$\sum_{(i,j) \in W}  I_1(i,j) - I_2(x+i, y+j) $
-----------------------------------	-------------------------------------------------

Zero-mean Sum of Absolute Differences (ZSAD)	$\sum_{(i,j) \in W}  I_1(i,j) - \bar{I}_1(i,j) - I_2(x+i,y+j) + \bar{I}_2(x+i,y+j) $
Locally scaled Sum of Absolute Differences (LSAD)	$\sum_{(i,j) \in W}  I_1(i,j) - \frac{\bar{I}_1(i,j)}{\bar{I}_2(x+i,y+j)} I_2(x+i,y+j) $
Sum of Squared Differences (SSD)	$\sum_{(i,j) \in W} (I_1(i,j) - I_2(x+i,y+j))^2$
Zero-mean Sum of Squared Differences (ZSSD)	$\sum_{(i,j) \in W} (I_1(i,j) - \bar{I}_1(i,j) - I_2(x+i,y+j) + \bar{I}_2(x+i,y+j))^2$
Locally scaled Sum of Squared Differences (LSSD)	$\sum_{(i,j) \in W} \left( I_1(i,j) - \frac{\bar{I}_1(i,j)}{\bar{I}_2(x+i,y+j)} I_2(x+i,y+j) \right)^2$
Normalized Cross Correlation (NCC)	$\frac{\sum_{(i,j) \in W} I_1(i,j) \cdot I_2(x+i,y+j)}{\sqrt{\sum_{(i,j) \in W} I_1^2(i,j) \cdot \sum_{(i,j) \in W} I_2^2(x+i,y+j)}}$
Zero-mean Normalized Cross Correlation (ZNCC)	$\frac{\sum_{(i,j) \in W} (I_1(i,j) - \bar{I}_1(i,j)) \cdot (I_2(x+i,y+j) - \bar{I}_2(x+i,y+j))}{\sqrt{\sum_{(i,j) \in W} (I_1(i,j) - \bar{I}_1(i,j))^2 \cdot \sum_{(i,j) \in W} (I_2(x+i,y+j) - \bar{I}_2(x+i,y+j))^2}}$

Tabla 2.1. Métodos de correspondencia de imágenes basados en la correlación

Dependiendo del criterio empleado, los resultados pueden ser dispares. La Suma de la diferencia de cuadrados (SSD), es más sensible a diferencias de iluminación, y la correlación cruzada normalizada (NCC) es el estándar utilizado como método estadístico para determinar las similitudes.

### Métodos basados en características

Frente a las técnicas basadas en el área en las que la correspondencia se realiza píxel a píxel, están las técnicas basadas en las características, en las que se emplean determinadas estructuras de un nivel superior al píxel para establecer correspondencias entre ellas. En concreto, se entiende por característica alguna estructura significativa de la imagen, en concreto, puntos de borde, segmentos de borde, esquinas o regiones.

Los métodos basados en características recibieron especial interés en los años ochenta debido a su eficacia. Sin embargo, debido a la necesidad de obtener mapas de profundidad densos y también debido a la mejor eficacia de los métodos basados en áreas, ha hecho que el interés declinara en los últimos años.

### Métodos globales

Son métodos que aplican restricciones a la imagen completa. Suelen ser menos sensitivos a las peculiaridades locales ya que realizan el cálculo de las disparidades de manera global, sin embargo, suelen ser computacionalmente costosos. El objetivo de los métodos globales es encontrar la disparidad que minimice la función de energía global.

$$E(d) = E_{data}(d) + \alpha E_{smooth}(d) \quad (\text{Ec 2.12})$$

Donde:

- $E_{data}(d)$  mide la semejanza entre los puntos ( $d$ ) que se hacen corresponder entre la imagen izquierda y la derecha
- $E_{smooth}(d)$  modela la suavidad dentro del mapa de disparidad, intentando asignar una disparidad similar a los puntos que se encuentren situados dentro de un mismo vecindario, imponiendo una coherencia en todas las direcciones y no sólo en la de la línea epipolar

La ecuación de optimización lleva asociadas ciertas restricciones que deben ser consideradas a la hora de resolver el problema. Las restricciones más básicas son:

- **Restricción epipolar:** Un punto y su correspondiente en la otra imagen se encuentran situados sobre la misma horizontal.
- **Restricción del orden:** Los puntos correspondientes entre dos imagen tienen que guardar siempre el mismo ordena relativo sobre la línea epipolar. Esto no se cumple en el caso de que existan objetos estrechos situados a distancias diferentes ya que no existe continuidad de superficie.
- **Límite del gradiente de disparidad:** Esta restricción impone límites en la tangente de las superficies del objeto, que deben ser bastante pequeñas, es decir, impone una restricción sobre la forma de los objetos que pueden ser reconstruidos por visión estéreo.
- **Restricción de unicidad:** Para cada punto de la imagen existe un único punto correspondiente en la otra imagen. Siempre que no haya objetos semitransparentes en la imagen, esta condición se cumple. En los casos en que existen puntos ocultos, éstos no tendrán correspondencia mientras que si no existen puntos ocultos la correspondencia será bidireccional entre las dos imágenes.
- **Restricción de continuidad:** Los objetos se consideran como superficies continuas presentando disparidades únicamente en los bordes de separación con otros objetos. La continuidad en la superficie también implica continuidad en los mapas de disparidades.

Los métodos que se han ido desarrollando para realizar la optimización de la función objetivo se dividen en deterministas y heurísticos o probabilísticos

#### 2.3.5. Determinación de la distancia (profundidad)

Una vez hecho corresponder los elementos que aparecen en la imagen izquierda con los elementos de la imagen derecha, es decir, una vez obtenida la disparidad ( $\xi$ ) de forma precisa

la determinación de la profundidad es un proceso relativamente sencillo, reduciéndose a una simple triangulación:

- Dado el punto tridimensional  $P(x,y,z)$ , las coordenadas de los puntos de proyección en las imágenes  $P_I(x_I, y_I)$  y  $P_D(x_D, y_D)$  y teniendo en cuenta las siguientes características de las cámaras  $f$ (distancia focal) y  $b$  (línea base), se obtiene las siguientes ecuaciones.

$$x_I = f \frac{x + \frac{b}{2}}{z} \quad (\text{Ec. 2.13})$$

$$x_D = f \frac{x - \frac{b}{2}}{z} \quad (\text{Ec. 2.14})$$

- Suponiendo ejes ópticos paralelos:

$$y_D = y_I = f \frac{y}{z} \quad (\text{Ec. 2.15})$$

- Se obtiene que la disparidad se calcula como:

$$\xi = x_I - x_D \quad (\text{Ec. 2.16})$$

Y por tanto:

$$z = f \frac{b}{\xi} \quad (\text{Ec. 2.17})$$

### 2.3.6. Interpolación

Este paso no resulta de aplicación siempre, se suele emplear cuando el mapa de disparidad obtenido en el paso anterior, no es denso, es decir, que no se obtiene suficiente información por unidad de superficie.

Uno de los métodos más sencillos, es considerar el mapa de disparidad como un muestreo de una función de profundidad continua, empleando un método de interpolación tradicional para hallar la función de continua que la aproxime. Algunos métodos que se pueden emplear son la interpolación de Lagrange, la interpolación de Hermite, mediante Splines o mediante Wavelets.

Mediante estos métodos, obtendremos un mapa de disparidad denso a partir de uno disperso.

## 2.4. Conclusiones

Una vez realizado el repaso de los métodos tradicionales de segmentación y los conceptos fundamentales de la visión estéreo, se obtienen las siguientes conclusiones:

- Los métodos tradicionales de detección de discontinuidades, detección de fronteras y umbralización son útiles desde el punto de vista de pretratamiento de la imagen, pero por sí solos no los más óptimos para la segmentación de objetos no rígidos que puedan variar en el tiempo tanto su forma como su posición.
- En los métodos de segmentación orientada a regiones, tienen las siguientes desventajas:
  - El método de adición de píxeles necesita la interacción del usuario para la elección de los puntos semilla, interacción que no interesa en este proyecto.
  - Los métodos de división y fusión no preservan los contornos reales, lo cual puede producir un error importante en el cálculo de la profundidad.
- El método de modelos deformables es que más se adapta a la segmentación de objetos no rígidos, sin embargo, estos métodos están desarrollados para imágenes en 2D.

Por lo comentado, se ha decidido emplear en este proyecto métodos innovadores que permitan segmentar objetos no rígidos en imágenes en 3D.