

Capítulo 4: Método de segmentación 3D basado en Contornos Activos

4.1. Introducción

En este capítulo se describe el método utilizado en la realización del Proyecto Fin de Máster.

Como se ha comentado en la sección 1.2 Objetivo del proyecto, el objetivo principal de este proyecto es desarrollar técnicas de segmentación de objetos no rígidos en 3D empleando el método de los contornos activos. Dentro de las posibles aplicaciones y utilidades, se ha seleccionado la detección de la profundidad a la que cae el agua cuando es expulsada a través de una manguera.

Por ello, el presente capítulo resume el método empleado en esta aplicación, es decir, partir de dos imágenes estéreo, se realizará una segmentación del objeto de interés, en este caso del chorro de agua, y se calculará la profundidad a la que se encuentra.

En concreto, el proyecto se ha dividido 4 partes, las cuales se resumen a continuación:

- Segmentación de la imagen: A partir de una de las imágenes del par estéreo, se realiza una segmentación cuyo objetivo es una aproximación al objeto de interés.
- Visión estéreo: Cálculo del mapa de disparidad: El objetivo de esta parte es determinar la correspondencia entre cada píxel de la imagen izquierda con los de la derecha.
- Segmentación de la imagen mediante contornos activos: Empleando la aproximación al objeto de interés calculado en la primera parte, así como la correspondencia entre los píxeles de las dos imágenes, se realiza la segmentación del objeto en ambas imágenes.
- Cálculo de la profundidad: Dado que se obtiene el objeto de interés en ambas imágenes y conocida la correspondencia, se obtiene la profundidad a la que se encuentra el objeto.

La siguiente imagen muestra la relación entre cada una de las partes mencionadas:

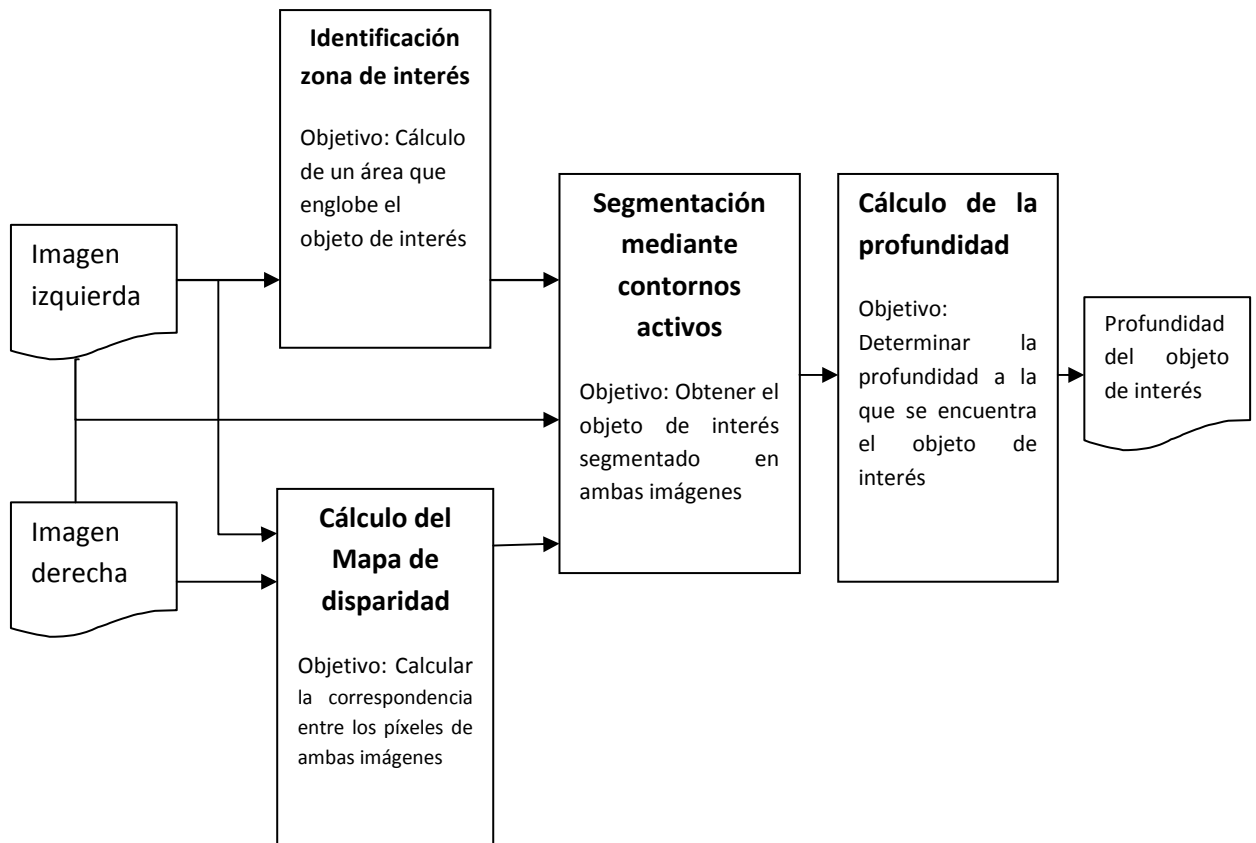


Fig. 4.1. Relación entre las partes integrantes del proyecto

4.2. Identificación zona de interés

Tal y como se ha descrito en apartados anteriores, la primera parte del proyecto consiste en cálculo de un área que englobe el objeto de interés de la imagen a procesar. En los siguientes subapartados se describe el método empleado desde la carga de la imagen a procesar hasta la segmentación del área de interés.

4.2.1. Descripción

El siguiente diagrama de flujo describe las principales áreas de la segmentación de la imagen



De forma adicional, en cada apartado se muestra, a modo de ejemplo, el resultado de la aplicación de cada área a la misma imagen, así como el desarrollo del programa para calcular el área de interés.

La imagen que se emplea como ejemplo, se muestra a continuación:



Fig. 4.1. Imagen de ejemplo

Técnicas de Preprocesamiento de la imagen

Con el fin de obtener mejores resultados, es necesario realizar técnicas de preprocesamiento de la imagen. El resultado de aplicar estas técnicas consiste en imágenes en las que se notan más los contrastes y las sombras.

La técnica usada ha consistido en una conversión del espacio de color de la imagen de RGB a $L^*a^*b^*$ y su posterior ecualización del histograma.

A continuación se muestran los detalles de esta técnica:

*Modelo $L^*a^*b^*$*

$L^*a^*b^*$ es el nombre abreviado de dos espacios de color diferentes. El más conocido es CIELAB (estrictamente CIE 1976 $L^*a^*b^*$) y el otro es Hunter Lab (estrictamente Hunter $L^*a^*b^*$). Ambos son derivados del espacio 1931 XYZ, sin embargo, CIELAB se calcula usando raíces cúbicas mientras que Hunter Lab se calcula usando raíces cuadradas.

El propósito de ambos es producir un espacio de color que sea más perceptivamente lineal que otros modelos, es decir, que un cambio de la misma cantidad en un valor de color debe producir un cambio casi de la misma importancia visual, además de describir los colores de forma visible al ojo humano.

En CIELAB, los tres parámetros que define en espacio son:

- L^* : Representa la claridad
- a^* : Representa la posición entre el magenta y el verde.
- b^* : Representa la posición entre el amarillo y el azul.

La siguiente imagen (Fig.5.1) muestra una representación de este modelo para tres valores distintos de claridad

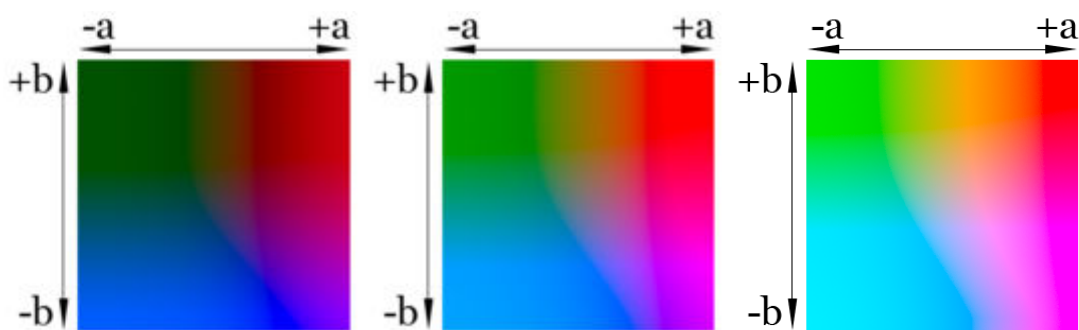


Fig. 4.2: Representación de a y b para claridad de 25,50 y 75% respectivamente

Las ecuaciones para pasar de XYZ a CIELAB son las siguientes:

$$L^* = 116f(Y/Y_n) - 16 \quad (\text{Ec.4.1})$$

$$a^* = 500[f(X / X_n) \cdot f(Y / Y_n)] \quad (\text{Ec.4.2})$$

$$b^* = 200[f(Y / Y_n) - f(Z / Z_n)] \quad (\text{Ec.4.3})$$

Donde:

$$f(t) = t^{1/3} \text{ para } t > 0.008856. \quad (\text{Ec.4.4})$$

$$f(t) = 7.787t + 16/116 \text{ en otro caso.} \quad (\text{Ec.4.5})$$

Mientras que en Hunter Lab se usan las siguientes:

$$L^* = 100\sqrt{Y / Y_n} \quad (\text{Ec.4.6})$$

$$a^* = K \cdot 100 \frac{X / X_n - Y / Y_n}{\sqrt{Y / Y_n}} \quad (\text{Ec.4.7})$$

$$b^* = K \cdot K_c \cdot 100 \frac{Y / Y_n - Z / Z_n}{\sqrt{Y / Y_n}} \quad (\text{Ec.4.8})$$

El motivo de la conversión a espacio $L^*a^*b^*$ se debe a las siguientes características:

- Modelo casi lineal, orientado al usuario.
- La componente L (claridad o luminosidad), se hace fundamental para el siguiente paso: ecualización del histograma.

Ecualización del Histograma

La ecualización del histograma de una imagen es una transformación que pretende obtener para una imagen un histograma con una distribución uniforme. Es decir, que exista el mismo número de píxeles para cada nivel del histograma de una imagen monocroma.

La función de ecualización es la siguiente:

$$v(i, j) = \frac{Fu(u(i, j)) - Fu(a)}{1 - Fu(a)} (L - 1) + 0.5 \quad (\text{Ec.4.9})$$

El resultado de la ecualización maximiza el contraste de la imagen sin perder información de tipo estructural.

El resultado de aplicar la ecualización del histograma a la imagen de ejemplo, se puede ver en la siguiente imagen:



Fig.4.3. Imagen de ejemplo ecualizada

Segmentación en espacio RGB

Una vez realizado las técnicas de preprocesamiento de la imagen, el siguiente paso consiste en realizar una primera segmentación en el rango del color objeto de la selección (color del agua).

Para ello, para cada píxel se comprueba su componente RGB, y dependiendo su valor, se le pondrá un nuevo valor dependiendo del siguiente criterio:

- Componente RGB correspondiente a color blanco o azul claro: El píxel permanece con su componente RGB original
- Componente RGB distinto de blanco o azul claro: El píxel toma el valor (0,0,0) (Negro).

Mediante este criterio se consiguen, las siguientes ventajas:

- Limitar la sección del área de interés a zonas que contengan el mismo rango de color del objeto de sección.
- Eliminación de bordes que puedan afectar a la selección del área de interés.

El resultado de aplicar este apartado a la imagen anterior, Fig.4.3, se muestra a continuación:

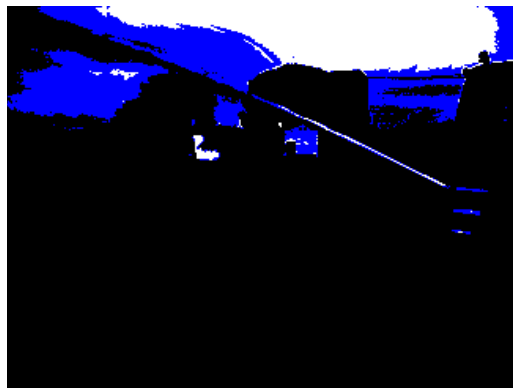


Fig. 4.4. Imagen segmentada en RGB

Detección de bordes (Canny)

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 4.5 (Fig.4.5) 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.

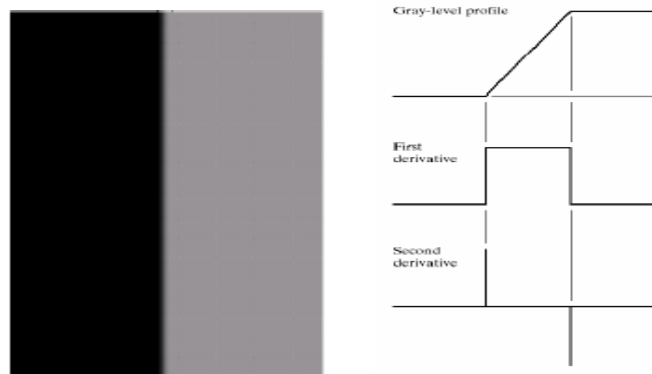


Fig.4.5. Representación de un borde con sus correspondientes derivadas

- **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:

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial}{\partial x} f(x, y) \\ \frac{\partial}{\partial y} f(x, y) \end{bmatrix} = \begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix} = \begin{bmatrix} G_F \\ G_C \end{bmatrix} \tag{Ec.4.10}$$

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:

$$\begin{aligned}G_F(i, j) &= F(i, j) \otimes H_F(i, j) \\G_c(i, j) &= F(i, j) \otimes H_c(i, j)\end{aligned}\tag{Ec.4.11}$$

Donde H_c y H_f dependerán del operador seleccionado.

La magnitud y orientación del vector gradiente se obtiene de acuerdo a la ecuación 4.12, aunque debido a su elevado coste computacional suele usarse la aproximación de la ecuación 4.13.

$$|G(i, j)| = \sqrt{G_F^2 + G_C^2}\tag{Ec.4.12}$$

$$|G(i, j)| \approx |G_F(i, j)| + |G_C(i, j)|\tag{Ec.4.13}$$

Los distintos operadores, basados en la primera derivada, 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
- Canny: Detecta tanto los bordes fuertes como los débiles, más robusto frente al ruido que otros operadores

Por la ventaja de Canny frente a los otros operadores basados en la primera derivada, se ha escogido el uso de este operador.

La detección de bordes sirve para discriminar información que no sea relevante, así como ser un paso necesario para la parte que viene a continuación.

El resultado de aplicar la detección de bordes a la imagen Fig.4.4, se muestra a continuación:



Fig. 4.6. Imagen bordes

Transformada de Hough

La transformada de Hough es una herramienta que permite detectar curvas o rectas de una determinada pendiente 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 la detección de bordes comentada 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 5.14

$$x \cos \theta + y \sin \theta = \tau \quad (\text{Ec.4.14})$$

Donde x, y son las coordenadas del píxel y ϑ es el ángulo que forma la normal con el eje x y τ 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 τ y ϑ , así como el número total de valores de τ y ϑ .
- 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 ϑ
 - Obtener τ de la ecuación 4.14
 - Incrementar $A(\tau, \vartheta)$
- Para todas las celdas en el acumulador
 - Buscar los valores máximos del acumulador
 - Las coordenadas (τ, ϑ) 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 4.7 (Fig.4.7)



Fig.4.7. (a) Imagen original



(b) Tras aplicarle detector bordes



(c) Tras aplicar la transformada de Hough



(d) Superposición de (b) y (c)

Debido a la característica parabólica del objeto de interés, una vez calculado la transformada de Hough, se han de seleccionar todos aquellos segmentos que posean un ángulo $\Theta=[30,70]$ para la parte subiente del chorro y que sean mayores de un determinado tamaño y de $\Theta=[0,20]$ para la parte bajante.

El resultado de aplicar esto a la imagen de la figura Fig. 4.6, se muestra a continuación:

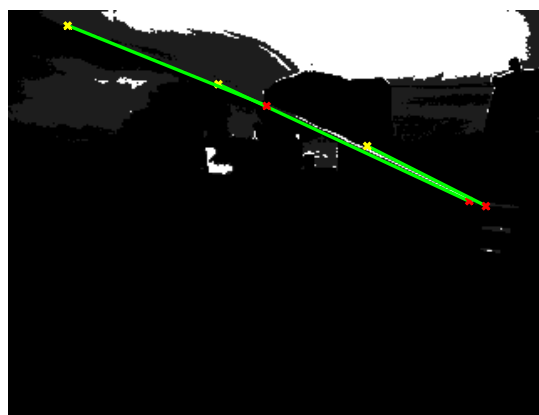


Fig. 4.8. Aplicación de la transformada de Hough

Formación del área de selección

Una vez realizada las técnicas de segmentación, se ha de formar el área de selección. Este área se forma a partir de los puntos seleccionados en el apartado anterior.

Para cada recta proporcionada en el apartado anterior, se seleccionarán puntos por arriba y por abajo para la parte subiente del chorro de agua, y por la izquierda y la derecha, en caso de que se haya detectado la parte bajante del chorro. La siguiente imagen muestra el resultado de este apartado:



Fig.4.9. Selección del área final

4.3. Visión estéreo: Cálculo del mapa de disparidad

Los siguientes subapartados describen el cálculo del mapa de disparidad con el fin de determinar la correspondencia entre dos imágenes estéreo.

4.3.1. Descripción

El objetivo de un algoritmo estéreo de cálculo de disparidad es obtener la correspondencia entre cada píxel del par de imágenes estéreo. Esta correspondencia, permite calcular a posteriori la profundidad a la que se encuentran los objetos de una imagen.

Los algoritmos de cálculo de disparidad asumen, por simplicidad, que las imágenes con que se están trabajando están rectificadas, es decir, que los planos de las imágenes de cada cámara son paralelos entre sí, y paralelos a la dirección en la cual existe el desplazamiento entre las imágenes. De esta forma, el correspondiente punto de la fila Y_l se la imagen izquierda, se encuentra en la fila $Y_r=Y_l$ de la imagen derecha.

A partir de esta afirmación, existen diversas técnicas basadas en la correlación entre el par estéreo. Las diversas técnicas de correlación en el cálculo del mapa de disparidad, son las 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)	$\frac{\sum_{(i,j) \in W} (I_1(i,j) - \bar{I}_1(i,j)) \cdot (I_2(x+i, y+j) - \bar{I}_2(x+i, y+j))}{\sqrt{\sum_{(i,j) \in W} (I_1(i,j) - \bar{I}_1(i,j))^2 \cdot \sum_{(i,j) \in W} (I_2(x+i, y+j) - \bar{I}_2(x+i, y+j))^2}}$

Cada píxel del mapa de disparidad que se obtiene, corresponde al número de píxeles que se desplaza la correspondencia entre el píxel de la imagen izquierda con el de la derecha.

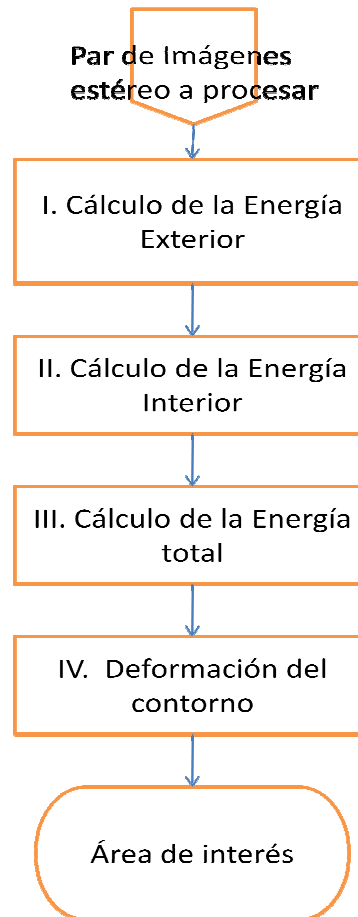
En los siguientes capítulos de este proyecto, se compara el resultado de aplicar las diversas técnicas de cálculo del mapa de disparidad citados, para el mismo par de imágenes estéreo.

4.4. Segmentación de la imagen mediante contornos activos

En los siguientes apartados, se describe el método de segmentación mediante contornos activos.

4.4.1. Descripción

El siguiente diagrama de flujo describe las principales áreas de la segmentación de las imágenes que forman el par estéreo mediante contornos activos



Cálculo de la Energía Exterior

El cálculo de la Energía Exterior, depende exclusivamente de propiedades de la imagen, y no de la forma que tenga la Snake (o contorno).

Para calcular la energía exterior se calcula el gradiente de la imagen en grises y para ello se ha usado el operador de Sobel. Definiéndose el gradiente de la imagen como:

$$G = |G_x| + |G_y| \quad (\text{Ec.4.15})$$

Donde G_x y G_y se obtienen de convolucionar la imagen por la siguiente matriz (Ec.4.16) y por su transpuesta respectivamente y tras quedarnos únicamente con la parte central de la matriz obtenida que coincide con el tamaño de la imagen:

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (\text{Ec.4.16})$$

Una vez obtenido G, se posee una matriz del tamaño de la imagen a procesar donde cada elemento de la matriz representa el gradiente del píxel de la imagen. Posteriormente esta matriz es normalizada con respecto a su valor máximo.

Se calcula una matriz de Energía Exterior por cada imagen. La correspondencia del componente de Energía Exterior de cada píxel de la imagen de la izquierda en la derecha, se obtiene a partir del mapa de disparidad comentado en apartados anteriores.

Cálculo de la Energía Interior

En esta parte se calcularán las contribuciones de E_{int} que son debidas a la forma que posee en ese instante la Snake.

Para calcular las contribuciones, se debe dividir en dos partes. La primera parte es la contribución debida a las fuerzas de tensión interna de cada nodo, que actuarán para mantener un espaciado uniforme entre los nodos del modelo, siendo α el parámetro que permite controlar la resistencia del contorno a las deformaciones de estiramiento (continuidad). La segunda componente serían las fuerzas de flexión donde el parámetro β controla la resistencia del contorno a deformaciones en curvas o recodos (curvatura).

Para calcular la primera componente, primero se calcula la distancia media de los nodos. Esta distancia media se calcula como:

$$Dis\ tan\ cia_media = \sum_{i=1}^N \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \quad (\text{Ec.4.17})$$

Posteriormente se calcula la distancia entre la posición del nodo bajo estudio (i) con respecto al nodo anterior:

$$Dis\ tan\ cia = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \quad (\text{Ec.4.18})$$

La componente de Energía interior debido a la continuidad estará dada por:

$$|Dis\ tan\ cia_media - Dis\ tan\ cia| \quad (\text{Ec.4.19})$$

Es necesario comentar que esta componente no se calcula únicamente para el nodo i , sino que se calcula para este nodo y su vecindad 3×3 , formando con ello una tabla de 9 componentes correspondientes a la energía de continuidad de cada elemento con respecto al nodo anterior (en posición, no en tiempo). A vez, se calcula el componente de Energía Interior debido a la continuidad tanto para el nodo de la imagen de la izquierda como para su correspondiente nodo en la imagen de la derecha. La correspondencia del componente de Energía Interior debido a la continuidad de cada nodo de la imagen de la izquierda en la derecha, se obtiene a partir del mapa de disparidad comentado en apartados anteriores.

Una vez calculada la tabla de la Energía de continuidad se normaliza con respecto a su valor máximo.

Para calcular la segunda componente, se realizan las siguientes operaciones:

$$V1 = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \quad (\text{Ec.4.20})$$

$$V2 = \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2} \quad (\text{Ec.4.21})$$

$$Cx = \frac{x_i - x_{i-1}}{V1} - \frac{x_{i+1} - x_i}{V2} \quad (\text{Ec.4.22})$$

$$Cy = \frac{y_i - y_{i-1}}{V1} - \frac{y_{i+1} - y_i}{V2} \quad (\text{Ec.4.23})$$

Por último la componente de Energía interna debido a la curvatura estará dada por:

$$Cx^2 + Cy^2 \quad (\text{Ec.4.24})$$

Al igual que en el caso anterior se calculará para el nodo i y para su vecindad, tanto para el nodo de la imagen de la izquierda como para su correspondiente nodo en la imagen de la derecha. La correspondencia del componente de Energía Interior debido a curvatura de cada nodo de la imagen de la izquierda en la derecha, se obtiene a partir del mapa de disparidad comentado en apartados anteriores.

La tabla creada de esta forma también será normalizada con respecto a su valor máximo.

Cálculo de la Energía Total

Dado que ya se han obtenido todas las componentes de energía necesarias, se calcula la energía total como:

$$E_{total} = \alpha E_{cont} + \beta E_{cur} + \gamma E_{ext} \quad (\text{Ec.4.25})$$

Siendo E_x la suma de la componente de Energía X de la imagen izquierda más la de la imagen derecha.

Deformación del contorno

Dado que la Energía Total se ha obtenido para la vecindad 3x3 del nodo, la nueva posición del nodo será la que minimice esta energía, pudiendo ser nuevamente la antigua posición. El algoritmo convergerá en el momento en que la posición de cada nodo que forma el contorno, se encuentre en la posición que minimice la energía.

Una vez converja el contorno, se habrá realizado la segmentación del objeto de interés.

4.5. Cálculo de la profundidad

Una vez obtenido el contorno del objeto de interés y el mapa de disparidad, la profundidad de cada nodo que forma el contorno del objeto, se calcula mediante la siguiente ecuación:

$$Profundidad = \frac{Base \times Distancia Focal}{Tamaño del Píxel \times Desplazamiento del Píxel} \quad (\text{Ec.4.26})$$

Donde *Base*, *Distancia Focal* y *Tamaño del Píxel* corresponden con parámetros internos de la cámara con la que se ha obtenido el par de imágenes estéreo.

Desplazamiento del Píxel corresponde al valor calculado en el mapa de disparidad para el nodo del contorno al que se le está calculando su profundidad.