

Estereoscopía Automática: Visión Artificial

Schumovich, Pablo
Bonoli, Andrea

Director : Prof. Zelasco, Jose

Facultad de Ciencias Exactas y Naturales.

~ Contenido ~

1. -Introducción

2. -Imágenes Estereoscópicas

- 2.1 - Fundamentos de la Estereoscopia
- 2.2 - Percepción Visual
- 2.3 - Haz Perspectivo
- 2.4 - Rayos Perspectivos
- 2.5 - Puntos Homólogos
- 2.6 - Parámetros del Haz Perspectivo
- 2.7 - Definición de Imagen
- 2.8 - Imagen Estereoscópica
- 2.9 - Disparidad Estereoscópica
- 2.10 - Propiedades de un Pixel
 - 2.10.1 -Vecindad de un Pixel
 - 2.10.2 -Conectividad

3 - Estereoscopia Digital : Tratamiento de Imágenes

- 3.1 - Pre- Procesamiento
- 3.2 - Detección de Formas Elementales
- 3.3 - Segmentación
- 3.4 - Puesta en Correspondencia
 - 3.4.1 - Restricciones
 - 3.4.2 - Técnicas de Búsqueda de Correspondencias

4. - Solución Adoptada

- 4.1 - Detección de formas Elementales
 - 4.1.1 - Algoritmo de Sobel
 - 4.1.2 - Algoritmo de Chehikian
- 4.2 - Segmentación
 - 4.2.1 - Elementos de una imagen
- 4.4 - Puesta en Correspondencia
 - 4.4.1 - Algoritmo de Viterbi

5. - Características del Desarrollo

5.1 - Implementación

5.2 - Características

5.3 - Parámetros del sistema

5.3.1 - Detección de contornos

5.3.2 - Segmentación de imágenes

5.3.3 - Puesta en correspondencia

6. - Resultados

7. - Conclusión

Anexo I.

1 Caso Normal

2 Caso General

3 Cálculo de la rotación del referencial imagen respecto del referencial general o de las coordenadas de la perspectiva

Bibliografía

~ Abstract ~

The obtainment of three-dimensional spatial information, as of a pair of stereo images, it is one of the investigation problems in the Artificial Vision branch. Its greater complexity bases in the stereo matching problem or the corresponding problem.

It is presented an approach that constitutes a solution for this problem. The method is composed of different stages that they are: the edge detection, the enhancement through filters and the terminals extraction in the stereo matching. This last stage uses a technique based on dynamic programming (Viterbi's algorithm). The result of this treatment is the construction of a disparity matrix that permits to recapture the 3D information of the matched objects.

Finally we analyze the results to evaluate the reliability of the stereo matching obtained and we present different examples.

~ Resumen ~

La obtención de información espacial tridimensional, a partir de un par de imágenes estereoscópicas, es uno de los problemas de investigación en la rama de Visión Artificial. Su mayor complejidad radica en el apareo estereoscópico o puesta en correspondencia de objetos.

Se presenta un enfoque que constituye una solución para este problema. El método se compone de distintas etapas que son: la detección de contornos, la depuración mediante filtros y la extracción de terminales para la puesta en correspondencia. Esta última etapa utiliza una técnica basada en programación dinámica (algoritmo de Viterbi). El resultado de este tratamiento es la construcción de una matriz de disparidad que permite rescatar la información 3D de los objetos apareados.

Por último se analizan resultados para evaluar la confiabilidad del apareo obtenido y se presentan distintos ejemplos.

1

~ *Introducción* ~

Numerosas aplicaciones requieren información espacial tridimensional. En particular y especialmente aquellas ligadas a la robótica, en la conducción automática de vehículos autónomos que realizan tareas pesadas o peligrosas; en medicina, en la realización de diagnósticos médicos; en metrología industrial, en la inspección automatizada de piezas industriales y control de calidad; en aplicaciones cartográficas con imágenes digitales satelitales u otras; etc.

La técnica de visión estereoscópica permite determinar las coordenadas 3D, a partir de mediciones realizadas en imágenes estereoscópicas.

El objetivo es reconstruir un modelo numérico de una cierta escena a partir de estas imágenes para lo que se requiere resolver tres aspectos: dos geométricos y un tercero correspondiente a la identificación de puntos homólogos.

La tarea correspondiente a este último aspecto, que normalmente se realiza en aparatos clásicos, es llevada a cabo por un operador en visión 3D estereo. En consecuencia se requiere encontrar algoritmos apropiados para realizar dicho apareo en forma automática de manera de reemplazar, o en casos muy complejos y como primer paso, asistir al operador en esta tarea que es particularmente tediosa. Los problemas geométricos son análogos en ambos casos, es decir con o sin operador.

En efecto, si se define una imagen como resultado de la intersección de un haz perspectivo con un plano, es posible, conociendo ciertos parámetros de la misma, determinar

la forma del haz, primer aspecto geométrico del problema . Por otra parte, dado que el modelo numérico se puede definir como el lugar geométrico de la intersección de rayos homólogos también se debe conocer la ubicación en el espacio de cada haz, que es lo mismo que ubicar en el referencial general los referenciales locales de cada imagen. Esto implica una rotación y una traslación para cada referencial imagen , siendo este el segundo aspecto geométrico.

Por haz perspectivo entendemos al conjunto de semirrectas, rayos perspectivos, cuyo origen es el punto de vista S , y que atraviesan la superficie visible desde S , de la escena u objeto de interés.

Los rayos perspectivos de distintos haces, pero que atraviesan el mismo punto en la escena, se llaman homólogos y las perspectivas en las imágenes determinadas por rayos homólogos son puntos homólogos.

La identificación de puntos homólogos también se suele denominar apareo estereoscópico o puesta en correspondencia, y es la que permite obtener los valores de paralaje o disparidad estereoscópica, diferencia relativa de posición de estos puntos, que conjuntamente con ciertos parámetros permiten obtener las coordenadas 3D generando el modelo numérico de la escena.

Este apareo no es trivial pues la identificación de puntos no es sencilla. Los algoritmos analizan la máxima probabilidad de correspondencia según cierto criterio o parámetros, y esta mayor probabilidad será la que determine el apareo de estos puntos, dependiendo su éxito, entre otras cosas, no solo del algoritmo sino de las características de las imágenes.

Existen dos tipos de soluciones en la búsqueda de correspondencias: por áreas y por características. En ambos casos para la obtención de terminales, es decir los objetos que el algoritmo busque aparear, debemos considerar ciertas dificultades que pueden aparecer en las tomas de vistas de las imágenes, como por ejemplo las variaciones fotométricas. Estas se refieren a la aparición de ruidos, sombras o reflejos dados por la posición relativa de los sensores, dando lugar a la existencia de "supuestos" objetos que dificultan el apareo. También existen dificultades por ciertas características de la escena, estas serían las oclusiones, donde la más elocuente ocurre en los bordes de la imagen, las texturas

repetitivas que producen una alta probabilidad de correspondencia (por ejemplo una pared de azulejos), la falta de textura que produce baja densidad de apareos, o superficies particulares como los objetos translúcidos o brillantes que producen efectos confusos [8; 13].

Este trabajo ha consistido en desarrollar una serie de algoritmos conducentes a la identificación de puntos homólogos y a evaluar comparativamente resultados de los mismos variando los valores de los distintos parámetros que intervienen.

Esta realización es un aporte para el proyecto UBA In061: Estereoscopia automática. Metrología: videometría en modelos 3D. Visión artificial en medio subacuático para AUV. En esta aplicación se ha previsto la utilización de dos sensores calibrados en un vehículo, lo que produciría imágenes roto-rectificadas, es decir ambas imágenes están en el mismo plano y las líneas de imágenes o píxeles son paralelas al segmento determinado por los dos puntos de vista. Cuando las imágenes tienen estas características se dice que son conformes a la geometría epipolar.

En consecuencia este trabajo se interesa particularmente en el problema del apareo, en este caso, en que la búsqueda de puntos homólogos se realiza sabiendo que los mismos se encuentran en la misma línea de imagen. Cabe destacar que se requiere resolver el problema del apareo sin contar con esta restricción de epipolaridad de las líneas de imagen para la solución de otros problemas más complejos para el tratamiento integral de la visión 3D de AUV.

En el próximo capítulo se presentan las características problemas y definiciones de la estereoscopia. En el siguiente, por un lado, la descripción de algunos algoritmos usuales en el tratamiento de imágenes orientados a la segmentación, y por el otro, algoritmos de búsqueda de puntos homólogos. El capítulo posterior aborda concretamente aquellos algoritmos que conciernen la solución adoptada. Finalmente se mencionan aspectos y detalles de la instrumentación para terminar con la presentación de conclusiones.

2

~ Imágenes Estereoscópicas ~

2.1 - Fundamentos de la Estereoscopia

La visión estereoscópica es un proceso por el cual se obtiene una impresión “visual” de relieve, análogo al de la visión binocular, a partir de la observación de dos imágenes de un mismo objeto tomadas desde dos puntos de vista diferentes. El modelo estereoscópico se produce gracias a la respuesta perceptual que está ligada a la visión binocular. El cerebro opera la fusión de dos imágenes estereoscópicas, generando la percepción de relieve. la visión binocular hace cooperar distintos tipos de información, tal como el ángulo de convergencia, distancia focal y, lo más importante respecto a este trabajo, que es la disparidad binocular. Aunque en visión estero no se cuenta con toda esta información los mecanismos de la percepción logran sin embargo crear un modelo de relieve.

2.2 Percepción visual

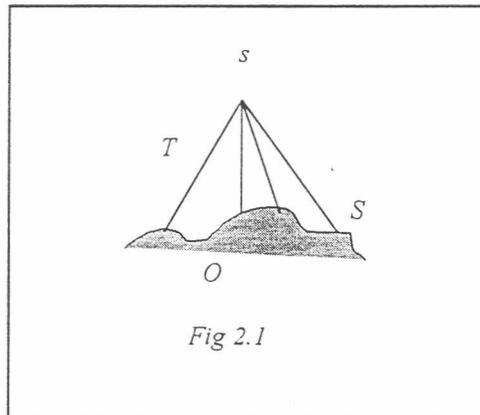
La sensibilidad del sistema visual humano depende en forma logarítmica de la intensidad de la luz que penetra en el ojo. Por lo tanto cuanto mayor sea el brillo, mayor debe ser al contraste entre los objetos para detectar cualquier diferencia.

La habilidad del ojo en discriminar entre varios niveles de brillo es una consideración importante en el procesamiento de imágenes. En el caso de la observación de una escena compleja, el sistema de visión humano no se adapta a un nivel único de intensidad, sino a un nivel medio que depende de las propiedades de esa escena.

Para obtener una imagen homogénea el ojo necesita una gama menor de niveles de intensidades diferentes que los 256 que proporcionan generalmente las imágenes digitales [10].

2.3 - Haz Perspectivo

Sea un objeto O , de superficie S , y un punto s , exterior a S , se llama haz perspectivo T , relativo a s y a S , al conjunto de semirrectas de origen s , que pasan por cada uno de los puntos de S , puntos de interés visibles de s .



Se denomina:

Punto de vista: al vértice s del haz.

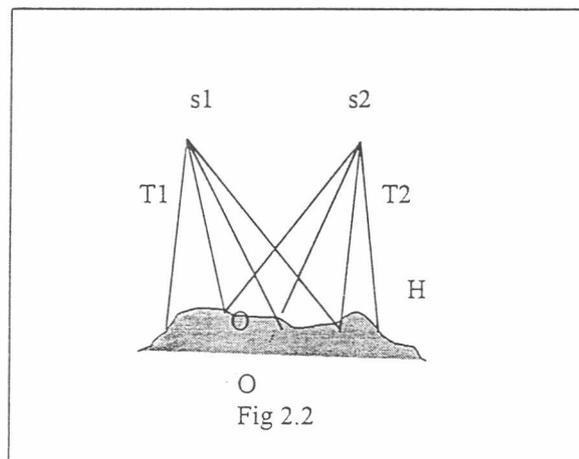
Rayo perspectivo: todo elemento del haz.

Destacamos que el conocimiento de s y de S , permiten determinar T , pero el conocimiento de s y de T , no son suficientes para conocer S , entonces la posición del punto M , intersección del rayo perspectivo T , con la superficie S del objeto O , está indeterminado.

2.4 - Rayos Perspectivos

Se supone tener dos vértices; s_1 y s_2 , de dos haces perspectivos; T_1 y T_2 , relativos a un objeto O , de superficie S , siendo los dos puntos de vista diferentes, es decir $s_1 \neq s_2$.

A cada punto de S , visible desde s_1 , corresponde un solo rayo t_1 del haz perspectivo T_1 , análogamente, s_2 , corresponde un solo rayo t_2 del haz perspectivo T_2 . Estos rayos se llaman rayos homólogos.



Por conocimiento de los haces perspectivos T_1 y T_2 se entiende: disponer de elementos que indiquen

1. La forma del haz, que esta dada por los parámetros internos del haz y no dependen del referencial general.
2. La posición de cada haz en el espacio: los parámetros externos, que dependen del referencial adoptado, y que son los de una rotación y una translación.
3. Que rayos son homólogos.

Esta información permite determinar la intersección de estos rayos homólogos, obteniendo los puntos M de la superficie S . Esta operación se denomina restitución.

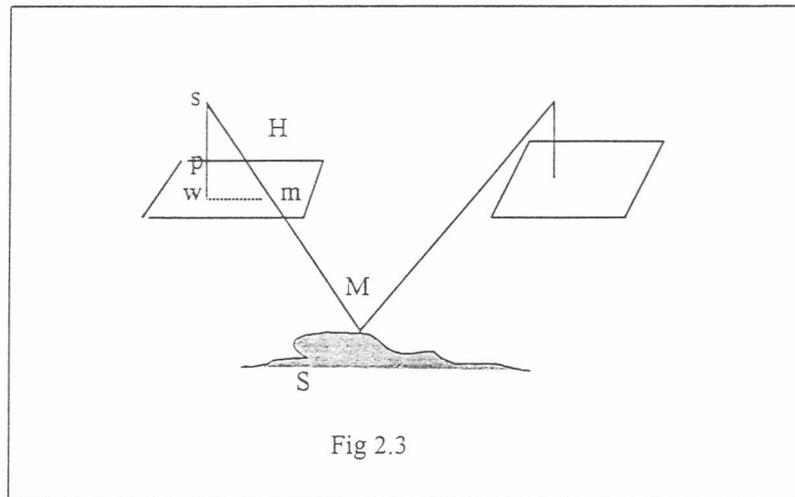
2.5 - Puntos Homólogos

Para determinar la forma del haz es conveniente recurrir a un plano perspectivo. Las

perspectivas de un mismo punto objeto se llaman puntos homólogos

2.6 - Parámetros del haz perspectivo

Para conocer al haz perspectivo T , debemos conocer sus parámetros externos e internos.



• Elementos Internos

Una perspectiva de la superficie S se determina mediante la intersección de un plano H con el haz perspectivo.

La posición del punto m , cuando un rayo atraviesa en M la superficie S , en H es la perspectiva del punto M .

Los elementos importantes de un plano perspectivo son

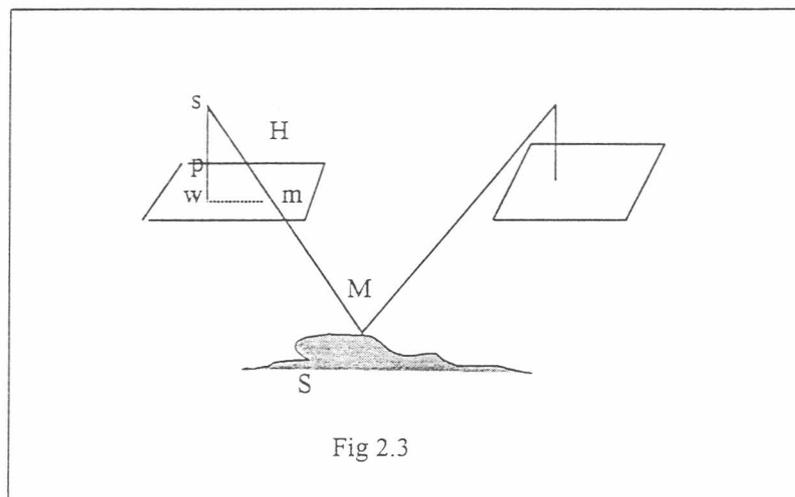
1. Eje principal : normal al plano H desde s .
2. Punto Principal : pie de dicha normal llamado w
3. Distancia Principal : segmento determinado por s y w llamado p

Si se conoce la perspectiva de la superficie de un objeto, distancia principal y posición del punto principal respecto del referencial oxy del plano H , se puede reconstruir el haz

perspectivas de un mismo punto objeto se llaman puntos homólogos

2.6 - Parámetros del haz perspectivo

Para conocer al haz perspectivo T , debemos conocer sus parámetros externos e internos.



- **Elementos Internos**

Una perspectiva de la superficie S se determina mediante la intersección de un plano H con el haz perspectivo.

La posición del punto m , cuando un rayo atraviesa en M la superficie S , en H es la perspectiva del punto M .

Los elementos importantes de un plano perspectivo son

1. Eje principal : normal al plano H desde s .
2. Punto Principal : pie de dicha normal llamado w
3. Distancia Principal : segmento determinado por s y w llamado p

Si se conoce la perspectiva de la superficie de un objeto, distancia principal y posición del punto principal respecto del referencial oxy del plano H , se puede reconstruir el haz

perspectivo.

La construcción de la perspectiva puede no ser perfecta. Esto se debe entre otras causas, a las características de distorsión geométrica de las lentes del sistema. Mas allá de la distorsión geométrica debida a las lentes, la diferencia entre la posición teórica del punto y la posición real será llamada *distorsión*.

- **Elementos externos**

Una vez elegido el referencial XYZ, el problema consiste en la orientación en el espacio del haz perspectivo, es decir, de un sólido en el espacio, lo que implica un referencial local. La ubicación del referencial local queda determinada por medio de una translación (3 parámetros) y una rotación (3 parámetros) que pueden ser conocidos por métodos directos o indirectos. Ver Anexo I .

2.7- Definición de la imagen

Una imagen se define como una aplicación de un espacio de dimensión 3, llamado espacio objeto, en un espacio de dimensión 2, llamado espacio imagen.

Un elemento del espacio objeto se denota (X, Y, Z) , y uno del espacio imagen (x, y) .
Sea I una imagen, se pueden entonces expresar las coordenadas (x, y) en función de (X, Y, Z) :

$$I = f(x, y) \quad / \quad x = F(X, Y, Z) \quad y = G(X, Y, Z)$$

La función bidimensional $f(x, y)$ depende de la intensidad de la luz, el valor de f en un punto va a ser llamado nivel de gris o intensidad de la imagen en el punto (x, y) .

Dado que la luz es una forma de energía $f(x, y)$ tiene valores no negativos y acotados.

$$0 \leq f(x, y) \leq M, \text{ para algún } M \text{ finito.}$$

2.8- Imagen estereoscópica

Dadas dos imágenes:

$$I_1 = \begin{cases} x_1 = F_1(X, Y, Z) \\ y_1 = G_1(X, Y, Z) \end{cases} \quad I_2 = \begin{cases} x_2 = F_2(X, Y, Z) \\ y_2 = G_2(X, Y, Z) \end{cases}$$

de un objeto O de superficie S de ecuación $Z=H(x,y)$, se llama imagen estereoscópica de O al conjunto de cuádruplas (x_1, y_1, x_2, y_2) cuando $(x$ y $z)$ se desplaza en toda la superficie S del objeto [14]. Si se identifica una cuádrupla (x_1, y_1, x_2, y_2) a partir de I_1 e I_2 , se define entonces dos rayos perspectivos, cuya intersección es un punto (X, Y, Z) de la superficie del objeto O .

$$(X, Y, Z)^t = f(x_1, y_1, x_2, y_2)$$

2.9- Disparidad Estereoscópica

La disparidad estereoscópica es la diferencia de posición de dos proyecciones de un mismo punto del objeto. Calculado de la siguiente manera:

Si $P_1(x_1, y_1) \in I_1$ y $P_2(x_2, y_2) \in I_2$ son homólogos, la paralaje se define como $T_1 = x_1 - x_2$. Si la columna de w_x es la misma en ambas imágenes la disparidad es igual a la paralaje. En efecto disparidad es la diferencia de columna de los puntos homólogos.

2.10- Propiedades de un pixel

2.10.1- Vecindad de un pixel

La 8-vecindad, $V_8(P)$, de un pixel P , está definida como el conjunto de todos los puntos que están alrededor de P .

Este conjunto está definido como

$$V_8(P) = \{Q / D_8(P, Q) \leq 1\},$$

donde D_8 es la distancia entre pixeles medida en unidades de pixel [6].

*	*	*
*	$P(x,y)$	*
*	*	*

2.10.2 Conectividad

Dos pixeles se dicen conexos si existe al menos una secuencia de pixeles que los unen, tal que dos pixeles consecutivos de esa secuencia satisfaga la condición de conectividad inmediata.

La conectividad inmediata entre 2 pixeles se traduce como el hecho de que sean consecutivos y tengan características comunes.

3

~ Estereoscopia Digital:

Tratamiento de Imágenes ~

El procesamiento y análisis de imágenes, es un conjunto de técnicas que permite modificar, analizar y manipular imágenes digitales, originariamente continuas o no, por medio de herramientas informáticas.

Un sistema de procesamiento de imágenes debe centrarse en tres aspectos: los medios de adquisición digitalización, los soportes de almacenamiento y la herramientas de procesamiento y de manipulación.

Estas consideraciones suponen que un sistema sea dotado de útiles que permitan atender esos objetivos. Un sistema genérico mínimo de procesamiento de imágenes, debe estar constituido de sensores para adquisición-digitalización de imágenes, una memoria para almacenar las mismas, y un conjunto de herramientas que permitan procesar y visualizar lo obtenido.

En otros aspectos, el procesamiento digital de imágenes [6], puede ser clasificado en cuanto al grado de abstracción, en tres niveles distintos; bajo, medio y alto. Donde se produce una inducción progresiva de la cantidad de información procesada, a medida que se

pasa a niveles crecientes de abstracción.

En el procesamiento de bajo nivel, la entrada de datos, son los pixels de la imagen original y los de salida, representan propiedades de la imagen, en forma de valores numéricos asociados a cada pixel. El procesamiento de nivel medio se presentan dos alternativas distintas en cuanto al tratamiento al cual se desee orientar la solución. El nivel alto realiza una interpretación del contenido de la imagen.

Respecto las alternativas del nivel medio existen dos categorías de algoritmos de segmentación para la puesta en correspondencia dependiendo de las primitivas que se utilicen [8; 13]:

- **Algoritmos basados en áreas**

Esta técnica intenta encontrar sectores similares en ambas imágenes, usando como primitiva la intensidad de los pixeles y consiste en comparar áreas de tamaño conocido entre ambas imágenes.

Los algoritmos de esta clase se aplican sobre imágenes donde la superficie varía en forma suave y continua, porque en imágenes muy texturadas o con grandes contrastes no logra resolver correctamente las ambigüedades que provienen de zonas similares ni las oclusiones.

- **Algoritmos basados en características**

Estos algoritmos manejan primitivas más elaboradas que utilizan estructuras para guardar información de interés, y permiten tratar imágenes texturadas y con grandes contrastes en forma satisfactoria, sin embargo en imágenes poco densas puede conducir a mapas de disparidad con poca información.

La técnica a implementar entonces debe decidirse considerando el tipo de imágenes a tratar.

Se presenta a continuación un esquema funcional de los algoritmos basados en características. En este esquema, si bien algunos puntos son comunes con la solución por áreas, la detección de formas elementales y la segmentación son específicos de las soluciones por características.

- **Adquisición de la imagen y digitalización**

La imagen obtenida por un sensor es transformada en una imagen digital como una tabla de valores discretos llamados pixels.

- **Pre-procesamiento**

Permite corregir cierto número de defectos aparecidos durante la adquisición de la imagen, que pueden tener como causa, las características físicas del sistema, condiciones deficientes de iluminación, o que simplemente deberse al carácter aleatorio de la señal. Dentro de esta clasificación se pueden incluir realces, modificación de histogramas, etc.

- **Detección de formas elementales**

Es la que caracteriza, identifica, y evalúa propiedades de esas formas elementales.

- **Segmentación**

La meta consiste en dividir una imagen en partes constitutivas [1; 5]. Una segmentación es efectuada por la detección de discontinuidades (contornos) o similitudes (playas) de una imagen. La mayoría del procesamiento esta basado en esta idea, los que son almacenados en forma adecuada (segmentos o primitivas).

- **Puesta en correspondencia**

Es la que establece las correspondencias que existen entre los elementos segmentados de las dos imágenes y determina cuales son las óptimas.

- **Interpretación**

Es la parte más inteligente del proceso de visión, a partir de la evaluación del modelo numérico se extraen las conclusiones de interés. Puede incluir procesos de respuesta perceptual.

A continuación haremos hincapié en los puntos que nos parecen de más relevancia a los efectos de este desarrollo, que son el pre-procesamiento, la detección de formas elementales, la segmentación y la puesta en correspondencia.

3.1 Pre-Procesamiento

Las imágenes monocromáticas están representadas por una función F , donde el valor $F(x,y)$ es una medida de niveles de gris o intensidad en el punto $P(x,y)$. Una imagen numérica resulta de un proceso de digitalización tanto en las coordenadas espaciales como en la cuantificación de niveles de gris.

Las técnicas de realce de imágenes, intentan obtener una imagen nítida, resaltando sus contornos de manera de eliminar zonas con ruido o poco contraste

Una imagen puede sufrir deformaciones debido a las imperfecciones en los dispositivos ópticos utilizados para su adquisición, fallas en el proceso de digitalización, ruido en el canal, modificación de brillo o contraste.

Se suele aplicar una serie de transformaciones a la imagen a fin de mejorar su calidad. Existen de dos tipos:

- Locales: el nuevo valor del pixel depende de su valor original exclusivamente.
- Espaciales: el nuevo valor del pixel depende de su valor original y del de los pixeles de una cierta vecindad.

Dentro de las operaciones de este tipo se encuentra el realce de detalles, que se aplica en imágenes con poca nitidez. Mediante un operador de diferenciación, se amplifica la diferencia en los tonos de grises adyacentes, resaltando zonas de poco contraste.

Se debe considerar que para aplicar esta operación, la imagen debe carecer de ruidos, dado que estos serían amplificados, por ser de niveles de brillo elevados.

- **Modificación de histograma**

El histograma de una imagen da, para un nivel de gris dado, el número de pixeles, cabe destacar que el mismo nada indica sobre la localización de los mismos.

Se define histograma de una imagen como el gráfico de una función $h(z)$, que denota la frecuencia relativa con que el nivel de gris z ocurre en la imagen, para todos los z en el rango $[0, N_g]$ de los niveles de gris de la imagen.

$h(z) = N_k/N$, donde N_k es el número de píxeles con nivel de gris Z_k , y N es el número total de píxeles en la imagen.

El histograma contiene información de interés, sobre las características de la imagen. A través de su transformación se pueden mejorar contrastes, iluminación mal controlada, etc.

3.2 Detección de formas elementales.

Estos procesos se realizan por convolución de la imagen mediante operadores de cálculo que estiman las derivadas en cada punto de la función intensidad, es decir los contornos se determinan por un salto de intensidad entre dos píxeles vecinos.

Existen diversos operadores que realizan esta tarea, entre ellos los que estiman al vector gradiente que apunta en la dirección del máximo crecimiento de la función f en el punto (x,y) , o los que calculan una aproximación digital del Laplaciano donde la función bidimensional $f(x,y)$ es una derivada de segundo orden. Estas técnicas se suelen complementar con otros filtros que eliminan bordes difusos o dobles y contornos surgidos por reflejos en la imagen.

3.3 Segmentación

Al optar por una solución por características de debe determinar el tipo de primitiva a utilizar, es decir un píxel en si mismo con su nivel de intensidad asociado no es un buen indicador, en base a esto se deben definir elementos que puedan representar una

característica dentro de una secuencia finita de sucesos posibles en la imagen

Previamente al proceso de segmentación propiamente dicho, se deben eliminar los elementos que carezcan de relevancia en el apareo, esto se refiere a primitivas con áreas pequeñas, que no aportan información de interés.

3.4- Puesta en correspondencia

Los algoritmos de apareo se distinguen según la forma de comparar sus elementos. Previamente a la presentación de las técnicas de apareo, describiremos una serie de restricciones que se deben contemplar para acotar la explosión combinatoria en el proceso de búsqueda, es decir, la comparación exhaustiva de elementos para su apareo es intratable computacionalmente y además estas restricciones sirven como una forma de verificar la consistencia del apareo.

En general se contemplan restricciones locales y globales [8].

3.3.1 Restricciones

Restricciones Locales

- **Epiporalidad**

Como ya expusieramos anteriormente, los puntos correspondientes deben estar sobre una misma línea epipolar. Esta propiedad requiere para su uso, que las líneas de pixeles coincidan con las líneas epipolares, y que las tomas respeten la geometría del caso normal (ver anexo 1).

- **Similaridad**

Los puntos correspondientes tienen características similares en las dos imágenes, si bien esta restricción es de importancia en el apareo, su aplicación debe ser lo suficientemente flexible como para contemplar distorsiones en los elementos generados por sombras y ruidos.

- **Lapso de disparidad**

Significa acotar la disparidad mínima y máxima por conocimiento de la escena. En efecto, debido al interés en la precisión de los valores XYZ del modelo, las imágenes se limitan, frecuentemente, a objetos que se encuentran en una zona limitada en profundidad. Estos límites son función de la distancia entre los sensores.

Restricciones Globales

Mencionaremos entre otras:

- **Unicidad**

A cada elemento de la imagen le corresponde a lo sumo uno de la otra imagen, en consecuencia un valor único de disparidad. Debemos destacar que la unicidad se aplica a nivel de elementos de correspondencia y no a nivel de píxeles, donde cada elemento puede estar conformado por distinta cantidad de píxeles, según la imagen.

- **Orden**

El orden posicional relativo de un elemento a lo largo de una línea epipolar, en una imagen, se preserva en la otra en ciertas aplicaciones. En ese caso el orden de aparición de elementos en una imagen y otra, debe tener el mismo orden, salvo objetos que puedan aparecer en una imagen y no en la otra.

3.4.2 Técnicas de búsqueda de correspondencia

Un recorrido de la literatura [5; 8; 13] concerniente a este tema permite determinar que las técnicas de búsqueda de correspondencias más usadas son cuatro:

- **Algoritmos basados en correlación**

Estos algoritmos se basan en la comparación de patrones entre imágenes [15] y la

forma en que se implementa es definiendo sectores o ventanas de tamaño fijo que se desplazan por una y otra imagen buscando encontrar un patrón similar. Los elementos que usa para la comparación son los píxeles tomando en cuenta solo su intensidad, y contempla restricciones como la epipolaridad y la similaridad.

Las dificultades que se presentan provienen de la ambigüedad que pueda surgir en la comparación de los patrones dado que un patrón puede aparear con varios de la otra imagen, además debe considerar la existencia de patrones repetitivos.

Estos algoritmos son inherentes a las soluciones por áreas.

- **Algoritmos basados en relajación**

Los algoritmos basados en relajación son procesos iterativos donde a cada elemento de la imagen izquierda se le asigna un nodo, a su vez este nodo tiene asignado un vector con dimensión igual a la cantidad de elementos de la imagen derecha. Como primer paso se asigna una probabilidad de correspondencia a cada vector de cada nodo y en cada iteración se ajustan en forma acumulativa las probabilidades de correspondencia, en la última iteración se seleccionan las relaciones que tienen la mayor probabilidad.

- **Algoritmo basado en programación dinámica**

El algoritmo propuesto [5; 9] basado en la técnica de programación dinámica busca un camino óptimo en un grafo de costos que relaciona objetos de la imagen izquierda con objetos de la imagen derecha. Esta alternativa corresponde a la solución adoptada y se detallará más adelante.

- **Algoritmos basados en predicción y verificación**

Estos algoritmos usan primitivas de alto nivel para describir la imagen y el apareo se realiza en dos etapas: la de predicción local que se realiza asignando correspondencias en un plano epipolar y una de verificación global que reajusta en forma recursiva los valores de correspondencia antes obtenidos pero considerando todos sus vecinos.

4

~ Solución Adoptada ~

En los capítulos precedentes se han presentado una serie de alternativas respecto a la solución del problema estereoscópico. En el corriente describiremos la solución adoptada y los algoritmos seleccionados.

A los efectos de adoptar alguna de las alternativas, se debe evaluar si conviene una solución por áreas o por características; considerando que las imágenes que se desean tratar no son específicas, se apunta a buscar una solución adecuada para una amplia gama de imágenes. Es por ello que la solución adoptada releva de algoritmos basados en características.

En función de la decisión precedente, y en una segunda instancia, el problema lleva a determinar el tipo de terminales a usar. Se han elegido los contornos, considerando estos los más adecuados debido a los buenos resultados que se obtienen al evaluar los algoritmos de detección de contornos en imágenes monoculares.

La solución se plantea mediante una serie de etapas a resolver:

- Detección de características
- Segmentación

- Puesta en correspondencia

Cada uno de éstos puntos deben resolverse según los algoritmos más eficientes, considerando además su costo computacional. Respecto al primer punto es oportuno considerar la calidad de la imagen de contornos, además de contemplar algún filtro para optimizar la imagen generada. El algoritmo de Sobel será el algoritmo adoptado en la detección de contornos, éste se vale de operadores de cálculo que se aplican sobre cada punto de la imagen, detectando cambio de intensidad entre pixeles.

Los operadores precedentes se complementan con el filtro de Chehikian. Los bordes detectados mediante el algoritmo de Sobel corresponden a zonas de contraste, entendiendo por esto regiones que abarcan varios pixeles, es conveniente, entonces, acotar la misma, a través de algún filtro que depure los bordes obtenidos.

En la segunda etapa, luego de aplicar otra serie de algoritmos que eliminan contornos con tamaño de áreas poco significativa, se procede a la segmentación, que consiste en construir los objetos para cada línea de la imagen, y almacenar los mismos con sus respectivas propiedades en forma adecuada.

Finalmente, se realiza el apareo de los objetos segmentados. Para ello se debe decidir entre los algoritmos de apareo presentados previamente. Será utilizado el algoritmo de Viterbi que pertenece a aquellos que se incluyen dentro de la técnica de programación dinámica.

La solución presentada, no es mas que un modelo viable, entre los posibles, para la obtención de un mapa de disparidad. Este consiste en una matriz de la dimensión de la imagen izquierda con el valor de disparidad correspondiente a cada punto.

En los puntos subsiguientes profundizaremos cada uno de los pasos detalladamente.

4.1 Detección de formas elementales

En esta etapa se encuentran los algoritmos inherentes a la detección de bordes. Estos algoritmos tienen su origen en el procesamiento de imágenes monoculares. Mediante

una adaptación que consta tan solo de la conservación de la información adecuada, son utilizados en el procesamiento de imágenes estereoscópicas. Esta información será utilizada en la segmentación y en la puesta en correspondencia.

El proceso inicial se realiza mediante el algoritmo de Sobel que calcula el módulo del gradiente y su dirección para cada punto (i,j) de la imagen. Este proceso detecta regiones en la imagen separadas por zonas de contraste, las que se distinguen entre si por un cambio significativo de niveles de luminosidad. En el interior de estas zonas de contraste se encuentran los contornos: líneas de partición entre regiones, compuesta por un pixel en el caso ideal.

Las zonas de contraste obtenida por el algoritmo de Sobel son generalmente superiores a un pixel dependiendo de la transición entre regiones, de los perfiles de esas zonas, de las magnitudes de los gradientes y de la dimensión de las máscaras utilizadas para el cálculo de dichos gradientes. Debido a esto se hace necesaria la utilización de algún método que analice el interior de las zonas de contraste y que conserve los pixeles representativos de las mayores variaciones entre los candidatos. El algoritmo de Chehikian será el encargado de realizar esta búsqueda, utilizando la información generada por Sobel.

4.1.1 Algoritmo de Sobel

El algoritmo de Sobel [5; 12], se basa en asociar a cada punto de la imagen $I(i,j)$ una aproximación $G(i,j)$ del vector gradiente obtenido por funciones finitas.. Los componentes G_y y G_x son las direcciones verticales y horizontales respectivamente, y son obtenidas por convolución de la imagen a través de las máscaras M_y y M_x y de dimensión 3×3 .

Lo que se busca es determinar la cercanía de un pixel a una zona de contraste. Al considerar que la distribución de $\|G\|$ teniendo en cuenta un umbral resulta bimodal, se traduce la existencia de dos tipos de pixels; aquellos cercanos a la zona de contraste -que se caracterizan por valores elevados de $\|G\|$ - y los mas alejados por valores pequeños de $\|G\|$.

- **Operadores de cálculo de los componentes del vector gradiente**

Debemos definir las máscaras M_x y M_y que por convolución de la imagen nos dan

las componentes del vector gradiente asociado a un pixel $I(i,j)$. Las máscaras de Sobel son las más apropiadas en el propósito de estudiar la orientación del vector gradiente.

- **Máscaras de Sobel**

Mediante las máscaras de Sobel se van a determinar el módulo del gradiente para la detección de contornos y la dirección utilizado en el algoritmo de Chehikian:

Módulo del gradiente en el punto (i,j) :

$$\|G(i,j)\| = (G_x^2 + G_y^2)^{1/2} \cong |G_x| + |G_y|$$

$$\text{donde } G_x = (a_{i-1j-1} + 2a_{ij-1} + a_{i+1j-1}) - (a_{i-1j+1} + 2a_{ij+1} + a_{i+1j+1})$$

$$G_y = (a_{i+1j-1} + 2a_{i+1j} + a_{i+1j+1}) - (a_{i-1j-1} + 2a_{i-1j} + a_{i-1j+1})$$

Luego se definen las máscaras como:

$$M_x = \begin{vmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{vmatrix} \quad M_y = \begin{vmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{vmatrix}$$

Dirección del gradiente en (i,j) :

$$\varphi_x = \tan^{-1}(G_y/G_x)$$

$$\text{luego } \tan(\varphi_x) = G_y/G_x$$

Por cada ventana de 3×3 se pueden determinar 8 direcciones posibles para el gradiente, utilizaremos el código de Freeman para numerarlas.

7	0	1
6	$P(x,y)$	2
5	4	3

Para determinar la dirección según Freeman se analiza el signo de las componentes del gradiente según $\Pi/8$ y $3\Pi/8$. Como los ángulos siempre se calculan sobre el eje x teniendo en cuenta los signos podemos reducir el análisis al cuadrante positivo.

Entonces si

Cod.Freeman

$$\begin{aligned} Dir(i,j) < \Pi/8 & : 2 \\ \Pi/8 \leq Dir(i,j) \leq 3\Pi/8 & : 1 \\ Dir(i,j) > 3\Pi/8 & : 0 \end{aligned}$$

Para determinar las demás direcciones debemos considerar los signos de Gx y de Gy:

Cod.Freeman

$$\begin{array}{cccc} Gx > 0 & Gx \geq 0 & Gx < 0 & Gx \leq 0 \\ Gy > 0 & Gy < 0 & Gy > 0 & Gy < 0 \\ Dir(i,j) < \Pi/8 & : 2 & 2 & 6 & 6 \\ \Pi/8 \leq Dir(i,j) \leq 3\Pi/8 & : 1 & 3 & 7 & 5 \\ Dir(i,j) > 3\Pi/8 & : 0 & 4 & 0 & 4 \end{array}$$

• **Caracterización de pixeles a una zona de contraste**

El operador precedente asocia a cada pixel $I(i,j)$ el vector $G(i,j) = (Gx, Gy)$, comparando el módulo de este vector a un umbral de decisión S, luego se atribuye o no, el pixel a una zona de contraste.

Si la comparación es favorable, se le asocia al pixel la dupla $(\|G(i,j)\|, Dir)$, caso contrario se asignará $(0,0)$. El umbral de decisión S es un parámetro variable que se comporta como la cota de valores a considerar como contorno.

4.1.2 Algoritmo de Chehikian

- Selección de pixeles del contorno

Recordemos que los pixeles pertenecientes a una zona de contraste, son aquellos que se caracterizan por el módulo $\|G\| \gg 0$, y la orientación Dir del vector gradiente que le es asociado.

El algoritmo de Chehikian [7] funciona como un filtro, por un lado, reduciendo la cantidad de pixeles que conforman un contorno (GRADIENTE MÁXIMO), y por el otro, eliminando aquellos pixeles aislados (GRADIENTE AISLADO) que generan variaciones de luz intensas y por lo tanto valores de $\|G\| \neq 0$.

Respecto a la noción de gradiente máximo, y dado que al calcular el $\|G\|$ se propaga con valores mayores a cero en pixeles cercanos a la zona de contorno, se reduce la cantidad de pixeles afectados conservando solamente los pixeles que, en la dirección del gradiente, tengan máximo valor del módulo. Es decir, se utiliza un criterio de selección basado en el pasaje por un máximo de la función módulo en la dirección del gradiente. Se detecta el pasaje por un máximo local al comparar el vector asociado al pixel tratado, con el de sus pixeles vecinos y más cercanos en la dirección del gradiente. Se debe determinar si el módulo de un gradiente es "GRADIENTE MAXIMO LOCAL". Para esto se deben evaluar los módulos de los puntos vecinos al punto (i,j) en la dirección del gradiente.

Se dispone de una tabla de conversión que nos indica el valor de los módulos en la dirección del gradiente. Los cálculos se realizan sobre una matriz de 3×3 con centro $(i,j) = (0,0)$.

$$\begin{array}{ccc}
 7 & 0 & 1 \\
 6 \left[\begin{array}{ccc} -1,-1 & -1,0 & -1,1 \\ 0,-1 & 0,0 & 0,1 \\ 1,-1 & 1,0 & 1,1 \end{array} \right] & & 2 \\
 5 & 4 & 3
 \end{array}$$

En el gráfico los valores exteriores indican las direcciones según Freeman. Por cada dirección tenemos dos módulos que llamaremos M_1 y M_2 .

<i>Dirección</i>		
0	$M_1 = a_{i-1,j}$	$M_2 = a_{i+1,j}$
1	$M_1 = a_{i-1,j+1}$	$M_2 = a_{i+1,j-1}$
2	$M_1 = a_{i,j-1}$	$M_2 = a_{i,j+1}$
3	$M_1 = a_{i-1,j-1}$	$M_2 = a_{i+1,j+1}$
4	$M_1 = a_{i-1,j}$	$M_2 = a_{i+1,j}$
5	$M_1 = a_{i-1,j+1}$	$M_2 = a_{i+1,j-1}$
6	$M_1 = a_{i,j-1}$	$M_2 = a_{i,j+1}$
7	$M_1 = a_{i-1,j-1}$	$M_2 = a_{i+1,j+1}$

$G(i,j)$ es Gradiente Máximo Local si $G(i,j) \geq M_1$ y $G(i,j) \geq M_2$

En relación con la noción de gradiente aislado y dado que los puntos aislados no conforman ningún contorno, deben ser eliminados. El pixel no precedido y no seguido de un pixel en la dirección perpendicular al gradiente, es un pixel aislado.

Para calcular si un punto (i,j) es GRADIENTE AISLADO debemos considerar los módulos pertenecientes a cada hemiplano generado por la dirección del gradiente. Consideremos como ejemplo que el módulo del gradiente en un punto (i,j) tiene dirección 5 y llamemos H_1 y H_2 a los hemisferios que se generan, debemos encontrar las componentes de H_1 y H_2 .

Considerando el gráfico precedente tenemos:

$$H_1 = \{ M_1 = a_{i-1,j-1}, M_2 = a_{i-1,j}, M_3 = a_{i,j-1} \}$$

$$H_2 = \{ M_1 = a_{i,j+1}, M_2 = a_{i+1,j+1}, M_3 = a_{i+1,j} \}$$

$G(i,j)$ es gradiente aislado si en $H_1=\{M_1=0 \text{ y } M_2=0 \text{ y } M_3=0\}$ o $H_2=\{M_1=0 \text{ y } M_2=0 \text{ y } M_3=0\}$

4.2 Segmentación

La segmentación de una línea se realiza identificando a partir de los algoritmos anteriores, los puntos consecutivos de gradiente no nulo, que conforman los objetos **contorno**, y los puntos consecutivos de gradiente nulo que conforman los objetos **playa**.

Durante estos procesos, al visualizar la línea, se observa la presencia de contornos parásitos, pertenecientes a pequeños objetos que no tienen correspondencia en la otra imagen. Un parámetro en el programa de segmentación de una línea de objetos, debe controlar y eliminar dichos contornos de gradiente débil, utilizando una regla general que evalúe el valor de la norma del gradiente en esos píxeles de contorno.

Así, los algoritmos tienen en cuenta la orientación de los contornos, y permiten la eliminación de contornos aislados.

Los contornos dobles, generados por el cálculo del gradiente, producen en la puesta en correspondencia, una multiplicidad inútil del número de objetos en cada línea.

Este problema se presenta con el siguiente patrón:

playa - contorno - pequeña playa - contorno - playa

que finalmente el algoritmo reduce a:

playa - contorno - playa

disminuyendo la posibilidad de inducir al algoritmo Viterbi a apareos falsos.

La cantidad de contornos parásitos hace a la imagen inutilizable a estos fines, la imagen debe, entonces, "limpiarse".

Uno de los métodos se basa en eliminar de la imagen, (filtrada por el algoritmo de Chehikian), todos aquellos contornos de tamaños inferiores a cierta talla, (parámetro variable es decir que se determina según la aplicación). La misma tarea se realiza para las playas que se caracterizan por tener su área menor a una cota (parámetro variable).

Las propiedades o atributos de cada objeto son: posición respecto al borde izquierdo de la imagen respectiva, longitud, nivel de gris promedio para las playas y gris inicial y final para los contornos entre otros.

La lista de objetos así obtenidos, son puestos en correspondencia con la lista de objetos de la otra imagen.

La segmentación está íntimamente ligada al problema de detección de playas y contornos.

Se definen en forma detallada los objetos contorno y playa en el punto 4.2.1.

- **Área de un elemento**

El área de una playa o un contorno se define como la cantidad de pixeles que lo conforman. La forma de determinar los puntos pertenecientes a un objeto es:

- 1) Hallar un punto perteneciente al objeto y marcarlo
- 2) Por cada $\delta_{vecino} V_i$ del punto tratado:
 - Si el tipo del punto V_i es igual al punto tratado marcarlo
 - Aplicar el punto 2) para V_i
 - sino retorna.

Los algoritmos que calculan áreas son procesos de características recursivas, debido a la gran cantidad de puntos de bifurcación que surgen al recorrer un contorno o playa. La solución recursiva es apta para calcular áreas pequeñas donde la cantidad de llamadas a la función de cálculo sea acotada. Existen dos alternativas que se presentan para evitar este problema, una es optar por una solución iterativa de búsqueda de pixeles pertenecientes al área en tratamiento y la otra planteada por este trabajo, una simulación del proceso recursivo que consiste en crear una estructura de tipo pila que contenga los puntos de bifurcación que queden pendientes de tratar.

La solución iterativa es apta para el cálculo de áreas, dado que por las características de las mismas estas se suelen encontrar en porciones de la imagen donde cada punto tiene

una alta cantidad de vecinos del mismo tipo. Este proceso consiste en seleccionar un punto no marcado del tipo a tratar y marcarlo; luego iterar sobre todos los puntos de la imagen la cantidad de veces necesarias hasta no encontrar mas puntos para marcar. Un punto de la imagen es marcado si : el punto tratado está marcado y algún punto vecino es del mismo tipo, al cumplirse esta condición el punto vecino es el marcado.

La otra solución en cambio, se la utiliza para calcular el área de los contornos. Se propone esta alternativa pues al utilizar una solución iterativa, el tiempo de procesamiento subía en forma excesiva. En este caso, se realiza un proceso similar a la solución recursiva, donde se simula el comportamiento recursivo mediante la utilización de estructuras de tipo pila.

4.2.1 Elementos de una imagen

Los elementos que se utilizarán son los **contornos** y las **playas** como indicáramos anteriormente. En este punto se define formalmente la construcción de los mismos. Se define el concepto de terminal formado por un pixel con sus atributos, para luego, partiendo de esta definición, construir los objetos correspondientes con sus respectivos parámetros. Por último se define la línea de objetos, que contempla como única condición el orden.

Cabe destacar, que los atributos antes mencionados tienen como función guardar información respecto al objeto construido, para su posterior utilización en el apareo.

- **Terminales y sus atributos**

Definición:

Llamaremos terminales a todo conjunto formado por un pixel de la imagen.

Sea $T = \{(i,j) / i \in \{1, \dots, N\}, j \in \{1, \dots, M\}\}$ si $N \times M$ es la dimensión de la imagen

Asociamos a un terminal T una lista de parámetros $P_1(T), \dots, P_k(T)$ llamados

parámetros descriptivos (atributos) del terminal T. Por otro lado, la posición de T estaría dada por las coordenadas (i,j). El mínimo terminal o elemento al cual se puede acceder en la imagen es el pixel.

- Definimos terminal de tipo CONTORNO

$C\{(i,j)\}$ de parámetros descriptivos :

$$P1(c) = \text{nro de identificación}$$

$$P2(c) = \| G(i,j) \| \text{ si } G(i,j) \text{ es el gradiente en } (i,j)$$

Regla : En el punto (i,j) existe un terminal de tipo contorno $C\{(i,j)\}$ si $G(i,j) > Sc$, donde Sc es el umbral de detección de contornos.

- Terminal de tipo PLAYA

$P\{(i,j)\}$ de parámetros descriptivos :

$$P1(P) = \text{nro de identificación}$$

$$P2(P) = 0$$

Regla: en el punto (i,j) existe un terminal de tipo “playa” si el valor de sus $\| G(i,j) \|$ no exceden el umbral Sc .

Definición:

1. Dos terminales se dicen iguales si contienen el mismo punto y sus parámetros descriptivos son iguales.
2. Dos terminales $T(l;k) = \{(l,k)\}$ y $T'(l;k') = \{(l,k')\}$ se dicen adyacentes a través de una línea l si $k' = k + 1$
3. Dos terminales se dicen del mismo tipo, si fueron definidos de la misma forma.

- **OBJETOS - Propiedades**

Definición:

Llamaremos objeto O a todo subconjunto de puntos de la imàgen.

$O = \{(i,j) / i \in D1, j \in D2\}$ son dos subconjuntos de $\{1, \dots, N\}$ si $N \times N$ es la dimensión de la imàgen.

A todo objeto O se asocia un punto de la imàgen (i_0, j_0) llamado posición de ese objeto y una lista de paràmetros $P1(O), \dots, PM(O)$ llamados paràmetros descriptivos o atributos del objeto. A continuación detallamos los atributos considerados de los objetos definidos en nuestro desarrollo.

- Objeto de tipo CONTORNO

$O_c \{(i,j)\}$ de paràmetros $P1(o) =$ nro. de identificación

$P2(o) =$ posición inicial

$P3(o) =$ punto medio

$P4(o) =$ longitud

$P5(o) =$ gris inicial

$P6(o) =$ gris final

- Objeto de tipo PLAYA

$O_p \{(i,j)\}$ de paràmetros $P1(o) =$ nro. de identificación

$P2(o) =$ posición inicial

$P3(o) =$ punto medio

$P4(o) =$ longitud

$P5(o) =$ gris promedio de sus 8-vecinos

$P6(o) =$ gris promedio de sus 8-vecinos

Todo terminal es un objeto; la definición de objeto contiene en efecto, toda aquella de terminal.

Un objeto se puede construir sin más, por una definición apropiada, o en forma más estructurada, utilizando las reglas para la construcción de terminales previamente presentadas.

Utilizaremos esto en el párrafo siguiente, para construir las estructuras de líneas de objetos mas complejas que serán después comparadas de una imagen a la otra.

4.4 Puesta en Correspondencia

La puesta en correspondencia de un par de imágenes estereo, va a consistir en determinar, por par de línea correspondiente, el conjunto de pares de puntos homólogos asociados.

Como expusieramos anteriormente, los pares de imágenes a tratar son imágenes roto-rectificadas, es decir, que cumplen con la siguiente propiedad: "Dos puntos correspondientes están sobre líneas de imágenes, es decir, las líneas de imágenes son líneas epipolares", lo que significa que sus ángulos epipolares son iguales. Se utiliza esta propiedad para reducir la complejidad de la búsqueda de correspondencia entre dos imágenes. El problema de la identificación del homólogo de un punto dado se vuelve, entonces, monodimensional.

Esta tarea se realiza mediante el algoritmo de Viterbi, que se encuentra entre las técnicas basadas en programación dinámica. Este algoritmo considera todas las correspondencias posibles, asignando un costo a cada una. Luego, elige localmente los pares de objetos donde la probabilidad de correspondencia es la más fuerte para, finalmente, seleccionar la correspondencia global más apropiada, es decir, optimal.

El objetivo es el cálculo por línea del valor de disparidad de cada punto correspondiente a cada objeto. La tabla de disparidad obtenida, que puede eventualmente densificarse por interpolación, consiste en una matriz en la que cada punto de interés tiene, en valores análogos al nivel de gris, la disparidad estereoscópica de dicho punto. Esta disparidad es calculada de la siguiente manera:

“Si $P1 (i_1, j_1)$ y $P2 (i_2, j_2)$ son homólogos, donde i, j son respectivamente fila y columna, la disparidad o relieve local en $P1$ es $d=j_1-j_2$.” De esta forma se puede calcular el modelo numérico a condición de contar con los parámetros internos correspondientes a las imágenes, la distancia entre los puntos de vista y la eventual translación que transforma la disparidad en paralaje longitudinal.

- **Descripción de una línea**

Definición:

Una línea l será descrita por un conjunto de objetos y una estructura que los contenga.

$S(l) = \{O_k, k=1, \dots, k\}$ poseyendo un orden estricto en relación a sus posiciones.

Esta restricción de orden de objetos constitutivos se respeta en ambas imágenes.

Este orden estricto, restricción de orden, será la única condición que debemos verificar para poder poner en correspondencia una línea l con otra línea estructurada. Esta definición deja una gran gama de posibilidades en la elección de la gramática de la estructura de una línea en medio de reglas de composición aplicada sobre objetos

4.4.1 Algoritmo de Viterbi

Por medio del algoritmo de Viterbi, se trata de resolver el problema de maximizar la probabilidad de la secuencia $X=\{x_0, \dots, x_k\}$. Este es un proceso de Markov, puesto que, la probabilidad de $P(x_{k+1} / x_0 \dots x_k)$ del estado x_{k+1} en el instante $k+1$ depende solamente del estado x_k en el instante k .

Se utiliza el algoritmo de Viterbi para encontrar la secuencia X óptima.

Definimos el costo acumulado en el nodo (i, j) como:

$$C(i, j) = \min_k f(C(k, j-1), c(k, j-1; i, j)) \text{ donde } c \text{ es el costo local de unir } (k, j-1) \text{ con el}$$

nodo (i,j) y $f(C,c)$ es la función que combina c y C .

El algoritmo de Viterbi consiste en aparear primitivas mediante la búsqueda del camino óptimo en un grafo, para esto se calcula un costo local de asociación de dos objetos que con el costo del camino mínimo hasta ese punto, permite seleccionar el mejor camino.

Como primer paso para poder comparar dos objetos, daremos una lista de reglas de correspondencia que permite calcular el costo de correspondencia entre dos objetos P_{1i} y P_{2j} , $i=\{1,\dots,M1\}$ y $j=\{1,\dots,M2\}$ si $M1$ y $M2$ son la cantidad de objetos de la línea derecha e izquierda respectivamente.

Llamaremos $CR(P_{1i}, P_{2j})$ al costo de correspondencia en el objeto P_{1i} y P_{2j} y $R(c)$ a la lista de reglas de correspondencia.

Definición:

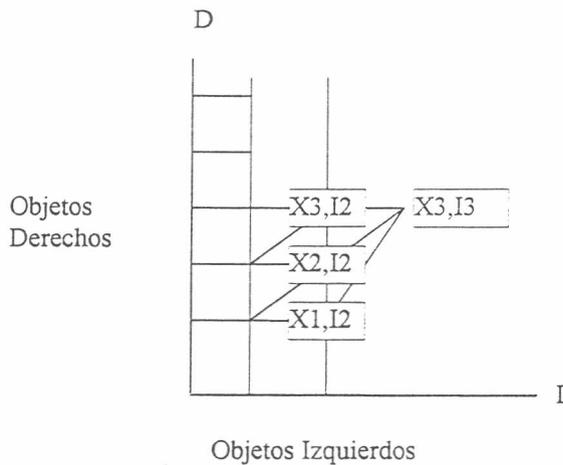
Si al aplicar la lista de correspondencias $R(c)$, dos objetos no pueden corresponderse, entonces la función de costo tomará el valor de "Costo de No Correspondencia" (CNC).

Nota:

Si $CR(P_{1i}, P_{2j}) < \text{CNC}$ el cálculo del costo estará basado en los parámetros descriptivos del objeto.

- $CR(A,B) = \text{CNC}$ si A y B son de tipos diferentes, considerar que una playa y un contorno no son comparables.
- Sea A un objeto de la imagen derecha con punto medio en la columna x y sea B un objeto de la imagen izquierda con punto medio x' luego $CR(A,B) = \text{CNC}$ si $x \leq x'$.
- **Método de recorrido del grafo**

Se utiliza un grafo de correspondencia para obtener el camino mínimo, que indiquen las correspondencias óptimas.



La construcción del grafo se inicia con el primer elemento izquierdo. Se debe calcular $C(i,1)$ para $1 \leq i \leq D$ (cantidad de objetos derechos). En este caso de inicio, no existe costo acumulado por lo que el costo de relacionar el objeto derecho i con el izquierdo 1 , coincide con el costo del camino mínimo en cada i .

Ahora, definiremos las funciones que intervienen en el cálculo del camino.

$C(i,j)$ = costo del camino hasta (i,j) .

$CR(i,j)$ = costo local de relacionar el objeto derecho i con el izquierdo j .

$$C(i,j) = \min \{ CR(i,j) * (1 - \text{param}_0) + [(\text{param}_0) * |i-k|] + C_{(x<i)}(x,j-1) \}$$

Recordemos que para el inicial: $C(i,1) = CR(i,1)$

$$CR(i,j) = [(|p_2(i) - p_2(j)| * \text{param}_1) + (|p_4(i) - p_4(j)| * \text{param}_2) + (|p_5(i) - p_5(j)| * \text{param}_3) + (|p_6(i) - p_6(j)| * \text{param}_4)] / \text{param}_5$$

donde $\text{param}_5 = \text{param}_1 + \text{param}_2 + \text{param}_3 + \text{param}_4$

Los parámetros de ponderación 'param_i' son los que permiten asignar un mayor o menor peso a las distintas características de un objeto, y los p_i son los definidos en el punto 4.3.

Supongamos que queremos calcular C(3,2), esto es lo mismo que decir que queremos obtener C(X3/ I=2). Como primer paso se calcula CR(3,2) según la fórmula previa, con este dato podemos calcular el valor de C, se asume además que param₀ tiene un valor de 0.2.

$$\text{Sea } C1 = CR(3,2) * 0.8 + 0.2 * |3-3| + C(3,1)$$

$$C2 = CR(3,2) * 0.8 + 0.2 * |3-2| + C(2,1)$$

$$C3 = CR(3,2) * 0.8 + 0.2 * |3-1| + C(1,1)$$

$$\text{Luego } C(3,2) = \min \{ C1, C2, C3 \}$$

Finalmente, cuando termine el proceso se selecciona de atrás hacia adelante los pares apareados que pertenecen al camino óptimal.

5

~ Características

del desarrollo ~

5.1 Implementación

A los efectos de visualizar las imágenes en los diferentes estados según se las haya ido procesando y sobre todo, con el objeto de manipular las distintas herramientas que se han instrumentado, se ha desarrollado una plataforma de trabajo en Visual C++ 1.0 Profesional que utiliza las Microsoft Foundation Classes (MFC) 2.0 bajo Windows 3.1. Los algoritmos específicos fueron desarrollados en lenguaje C.

Frente a la necesidad de poseer un prototipo que permitiera evaluar el comportamiento de los algoritmos seleccionados para la solución propuesta y luego de analizar diversas alternativas se optó en una primera instancia por utilizar Lenguaje C debido a que es un lenguaje de fácil portabilidad a distintos ambientes. También se tuvo en cuenta el

interés de utilizar e integrar estos algoritmos en otros subproyectos, dentro del marco del proyecto UBA In061 que cuenta con cuatro grupos que trabajan en distintos aspectos del problema de la obtención de modelos 3D a partir de pares estereo.

Luego, se planteó la necesidad de poseer una herramienta que permitiese la evaluación visual de los procesos realizados sobre las imágenes que además incluyera útiles que facilitaran la selección de imágenes, la aplicación ordenada de los algoritmos, la visualización de resultados generales de puesta en correspondencia en imágenes o tablas de disparidad, etc.

En base a lo precedente, se analizaron las diferentes opciones que brindarían una solución adecuada a las necesidades. Otro factor que resultó determinante fue la conveniencia de homogeneizar todas las plataformas desarrolladas en cada subproyecto. Todas estas circunstancias llevaron a realizar el desarrollo bajo Windows.

5.2 Particularidades

Por razones análogas a las expuestas en el punto precedente se desarrollaron los algoritmos teniendo presente el objetivo de lograr un alto nivel de modularidad de manera de conseguir que se comporten como “cajas negras” de cómoda adaptabilidad y de fácil su reusabilidad.

Dado que los procesos interactúan con la imagen, algunos transforman sus datos y otros los utilizan para generar nuevas imágenes surgió la necesidad de independizar los algoritmos de los accesos a la imagen.

Para lograr este objetivo se utilizaron estructuras que guardan la información necesaria para manejar la imagen, a su vez se definieron una serie de funciones que consisten en recuperar y grabar información a las estructuras o realizar los accesos a disco en la carga inicial. Estas funciones, que se podrían denominar de Entrada/Salida, son las únicas que deberían modificarse en caso de cambiar de ambiente.

Cabe mencionar, que para una primera evaluación de los algoritmos, se desarrolló un prototipo en DOS, lo que obligó a realizar un paginador de imágenes, debido a las limitaciones de memoria que impone este S.O. Este paginador trabaja en forma transparente

dentro de las funciones de Entrada/Salida, de tal forma que al migrar a Windows se pudo prescindir de las mismas fácilmente.

El S.O. Windows, resulta plataforma adecuada ya que satisface la función de proveer un menú, permite abrir una o varias imágenes, y además, a cada una de estas instancias se le pueden aplicar todos los algoritmos propuestos en el menú mencionado.

5.3. Funcionamiento del sistema

El sistema elabora a partir de un par de imágenes estereoscópicas una imagen de disparidad que a condición de disponer de los parámetros adecuados permite la obtención del correspondiente un modelo numérico.

Para lanzar la aplicación se debe ejecutar el programa "CORRESP.EXE". La ventana que se presenta permite como primer paso abrir el par estéreo para lo que se requiere el nombre de cada imagen y su ubicación. Obviamente es necesario poder distinguir entre imagen derecha e imagen izquierda del par, por lo que se debe adoptar algún tipo de convención que las identifique sistemáticamente.

La interfaz de este sistema provee las herramientas típicas de un ambiente Windows, con cajas de diálogos; apertura de múltiples ventanas; opciones como Cortar, Pegar y Copiar una imagen en una ventana; etc.

Posteriormente a la carga inicial, se habilitan una serie de procesos que conforman los pasos necesarios para lograr el objetivo buscado, y que consisten en:

- Detección de contornos
- Binarización
- Segmentación
- Puesta en Correspondencia

Esta subdivisión del proceso general en partes permite realizar un seguimiento de las distintas transformaciones que sufre la imagen en cada etapa, permitiendo hacer ajustes en los distintos parámetros que intervienen.

Teniendo en cuenta el origen de esta subdivisión se debe considerar

Si bien hay que tener presente que cualquier alteración en el orden de ejecución de los mismos puede llevar a obtener resultados erróneos, la independencia en el orden de aplicación permite observar los resultados al cabo de cada operación y evaluar distintas características del proceso que pueden resultar de interés.

Los tres primeros pasos de ésta división trabajan solo sobre la ventana activa. Si se desea obtener la imagen de disparidad se debe repetir el proceso para la otra imagen del par, sin tener necesidad de aplicar el mismo juego de parámetros.

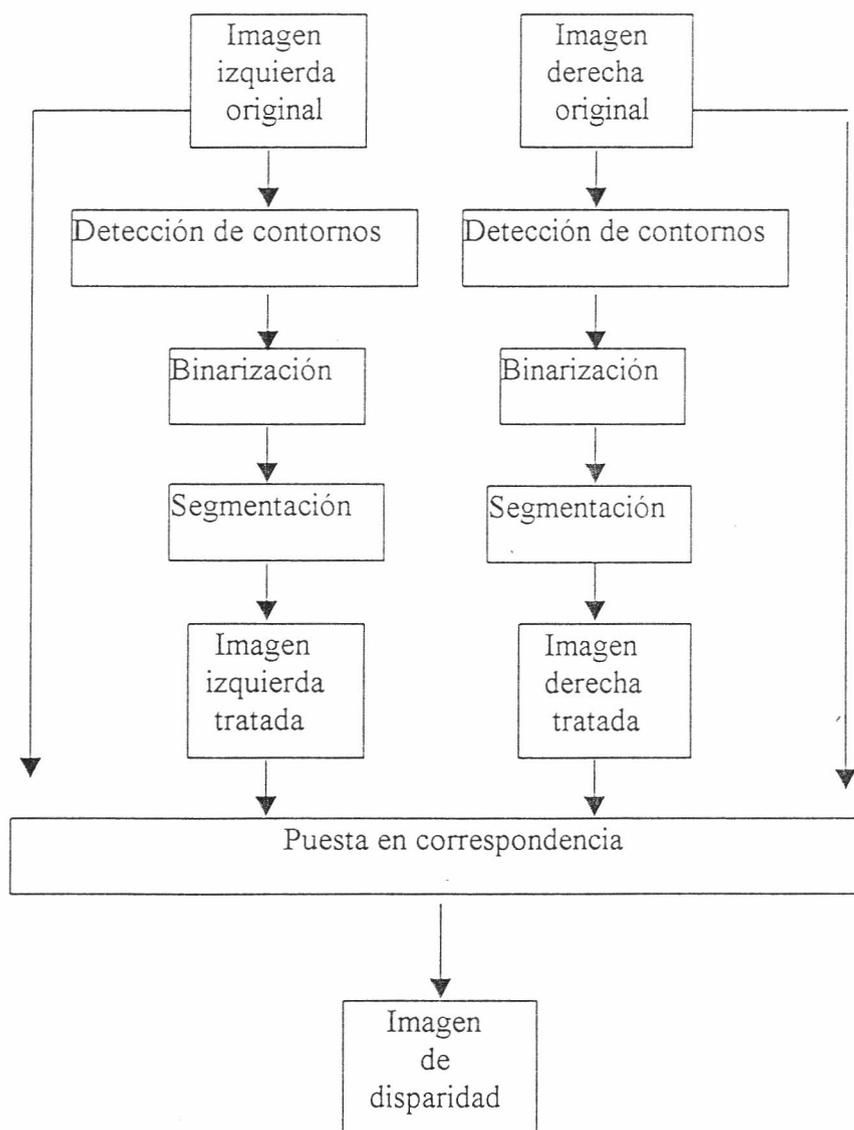


Fig. 5.1

Respecto al último paso que consiste en la Puesta en Correspondencia, el sistema necesita tener abierta el par estéreo inicial y el par correspondiente tratado según los pasos previos en el orden establecido por el menú.

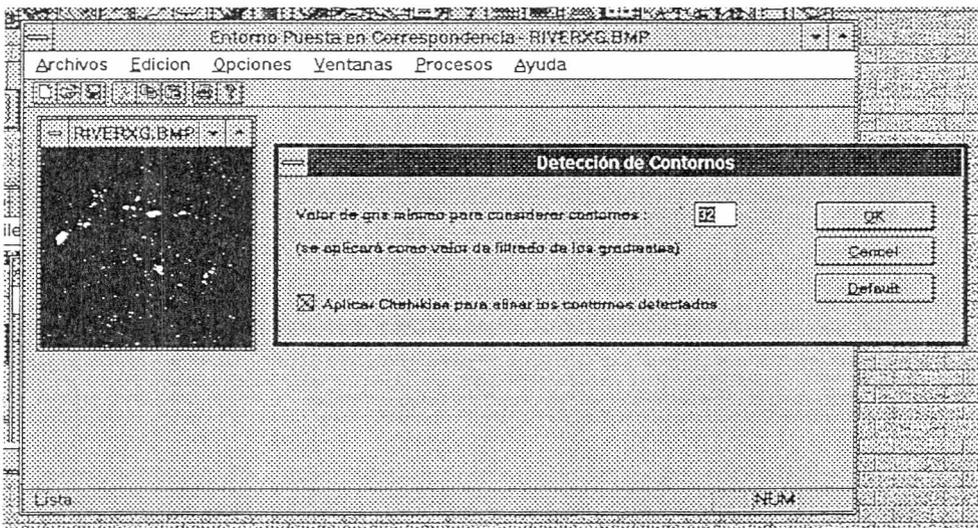
Respecto a los parámetros requeridos y a su utilización ver 5.3.1. En la figura 5.1 se muestra un esquema general del funcionamiento del sistema .

5.3.1- Parámetros utilizados

- **Detección de contornos**

Valor de gris mínimo para considerar existencia de contornos:

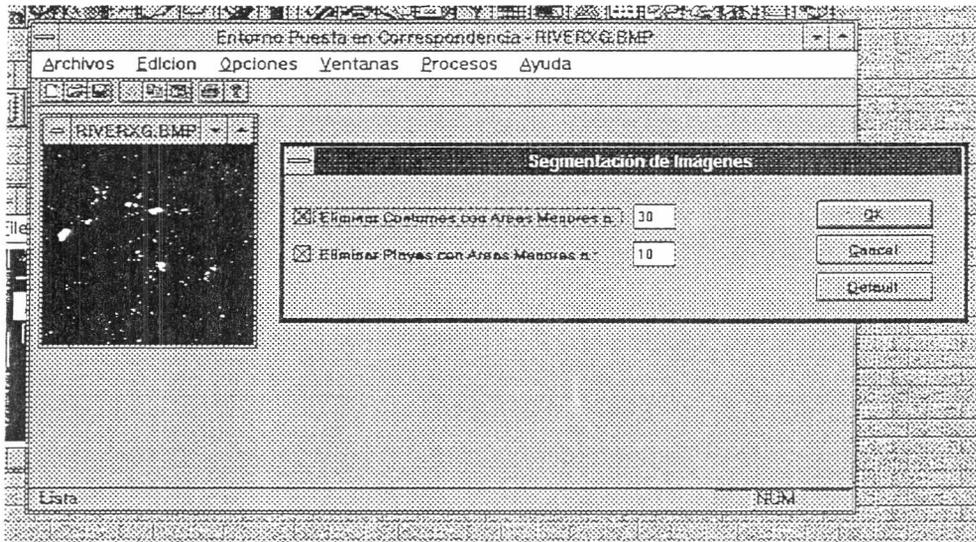
Este parámetro se utiliza para realizar una poda de los valores de módulo de gradiente que indican saltos de intensidad de poca amplitud, se debe recordar que los contornos se detectan por valores $\gg 0$.



- **Segmentación de imágenes**

En el proceso de segmentación se realizan dos procesos de poda que consisten en eliminar contornos y playas con áreas pequeñas. Los valores requeridos corresponden a la cantidad de píxeles que las componen. La dimensión del valor de poda en la eliminación de objetos es altamente dependiente del tipo de imagen, debido a ello para su determinación se debe evaluar el tamaño de los objetos encontrados, se debe tener en cuenta que valores muy

altos de poda podría eliminar información de importancia, por otro lado valores muy bajos conservan objetos de poca relevancia.



- **Puesta en correspondencia**

Valores de peso:

La asignación de valores de peso a las distintas características de un objeto influye en su preponderancia en la función de costos, es decir, si se quiere realizar un apareo donde prevalezcan la similitud en la longitud de los objetos tratados, el valor de peso de esta propiedad debe ser superior al de las demás. Los valores de estos pesos tienen un rango correspondiente a un porcentaje.

Cotas de No correspondencia :

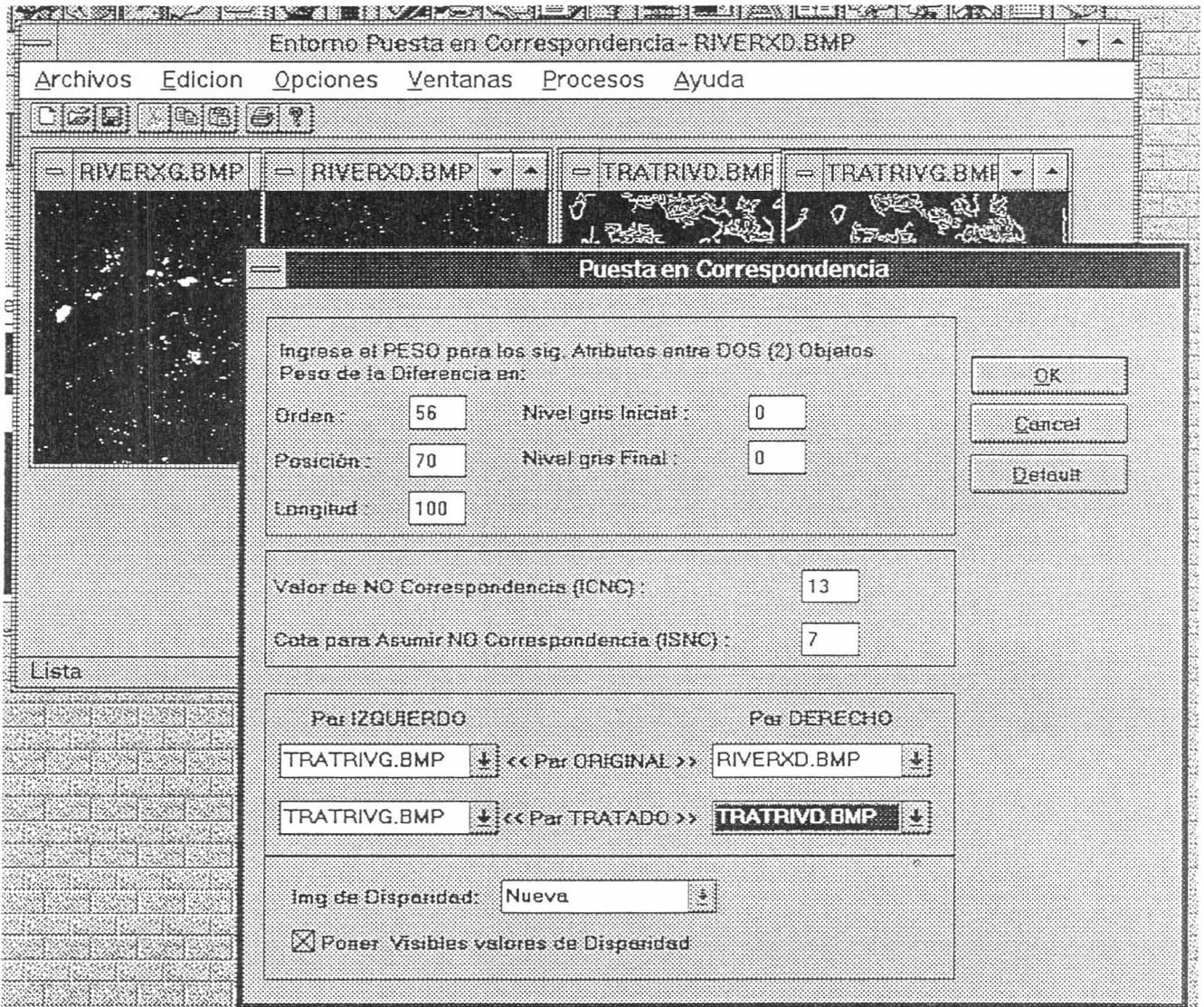
Estos valores están íntimamente ligados a la función de costos y permiten manejar mayor o menor margen de error al poner en correspondencia dos objetos.

Valor de No Correspondencia (ICNC) = valor que asigna la función al hallar una situación de No Correspondencia.

Cota para asumir No Correspondencia (ISNC) = esta cota sirve para determinar si el costo de un apareo debe asumirse de No Correspondencia.

- Poner visibles valores de disparidad:

Esta operación aumenta los valores de disparidad de una magnitud fija. Esto surge a raíz de la necesidad de hacer visibles valores de disparidad que se generan muy bajos y que, en consecuencia, no son visibles.



6

~ Resultados ~

Se han evaluado distintos aspectos relativos a los procesos que conducen a la obtención de un mapa de disparidad.

En una primera instancia se realizaron diversas pruebas referidas a la detección de bordes, variando valores del parámetro de poda que utiliza, es decir, el umbral de valor de gradiente a partir del que se admite la existencia de un borde. Los mejores resultados se obtuvieron para un rango de valores entre [20,30], teniendo en cuenta que éstos solo son válidos en imágenes nítidas, cuyo contraste esté bien definido, de otra forma este rango debería decrementarse.

Con valor de poda $\ll 20$, las imágenes generadas se caracterizan por una gran cantidad de contornos producto de sombras y/o reflejos en la imagen. Como mencionáramos previamente esta situación no se traduce en apareos confiables.

En el otro extremo, con valor de poda $\gg 30$, es decir, valores altos en este parámetro, se produce la pérdida de contornos con saltos de intensidad medios, conservando solamente aquellos donde la variación de tonos de gris es alta. Si bien en determinadas imágenes pueda resultar de utilidad, en general se denota la pérdida de información relevante.

Otra verificación realizada consistió en obtener imágenes de disparidad donde no se hubiese aplicado el filtro de Chehikian. Las imágenes de contornos que se binarizan sin utilizar este filtro previamente, se caracterizan por tener contornos de tamaño muy superior a un pixel, al realizar el apareo se obtienen mapas de disparidad de dudosa significación, además de un notorio aumento de los tiempos de procesamiento en la etapa de segmentación.

Respecto a la segmentación, el análisis se basó en analizar el resultado del apareo previo filtrado de objetos de áreas inferiores a valores pequeños ($\cong 5$ pixeles de playas o contornos). Ésto produjo una multiplicidad inútil de correspondencias. Los objetos de áreas pequeñas corresponden a defectos de la imagen que no son propios de la escena. Por otro lado, la eliminación de objetos con áreas significativas (> 50 en ambos casos) provocan la eliminación de detalles de relevancia en el apareo. Por otra parte, en la detección de contornos y a raíz del algoritmo que los calcula, se pueden producir cortes en una línea de contorno sin dejar pixeles conexos entre sí. Estas líneas pueden ser erróneamente eliminadas al considerar áreas de gran escala. El tamaño óptimo para la poda es altamente sensible al tipo de imagen tratada, dependiendo del tamaño de los objetos que contenga la escena.

Los mejores resultados, se hallaron mediante la utilización de poda de áreas de diez pixeles para contornos y veinte para playas.

Finalmente, se evaluó el comportamiento en el apareo al variar los parámetros específicos de puesta en correspondencia que se presentan en dos grupos, los de cálculo de costos, que son pesos que se aplican a los distintos atributos de los objetos tratados y los que determinan la posibilidad o no de correspondencia entre ciertos objetos.

Las propiedades que son afectadas por los pesos del primer grupo son: el orden, la posición, la longitud y los grises. Los parámetros del segundo grupo son: el ISNC que es el valor de la cota para asumir no correspondencia y el ICNC que es el valor que se asigna para los costos que superan al ISNC.

La búsqueda de los parámetros óptimos se torna compleja al prescindir de una herramienta automática de evaluación exhaustiva de parámetros. Las pruebas realizadas denotan que al variar el peso que se le asigna a las distintas propiedades no se traducen en cambios fácilmente detectables, además de ser altamente sensibles al tipo de imagen tratada.

Por otro lado, los parámetros de peso están ligados entre sí lo que dificulta aún más la evaluación.

Las pruebas se realizaron combinando diferentes juegos de parámetros que se aplicaron a distintas imágenes, para las que se evaluaron aspectos cuantitativos y cualitativos.

Para el análisis cuantitativo se calculó el porcentaje de correspondencias encontradas sobre líneas predefinidas aleatoriamente.

Para el análisis cualitativo se utilizó una herramienta para reconstruir la imagen derecha a partir de la imagen izquierda y de la imagen de disparidad; y luego se evaluó la calidad de la imagen rala obtenida.

En la tabla siguiente se muestran 12 juegos de parámetros con los que se realizaron las pruebas, donde las columnas Orden, Posición, Longitud, Gris Ini y Gris Fin indican el porcentaje de peso que se asignó a ese atributo y la columna Corresp el porcentaje de correspondencias halladas para esa prueba.

Prueba #	Orden	Posición	Longitud	Gris Ini	Gris Fin	ISNC	ICNC	Corresp ⁽¹⁾
1	0	25	25	25	25	40	60	70
2	50	25	25	25	25	40	60	76
3	0	100	0	0	0	40	60	86
4	0	0	100	0	0	40	60	92
5	0	0	0	50	50	40	60	72
6	50	25	25	25	25	20	80	51
7	50	25	25	25	25	10	15	28
8	50	25	25	25	25	85	90	70
9	0	50	50	0	0	10	15	48
10	0	0	100	0	0	10	15	77
11	50	25	25	25	25	12	15	32
12	0	50	50	0	0	12	15	76

(1) indica Porcentaje de Correspondencia

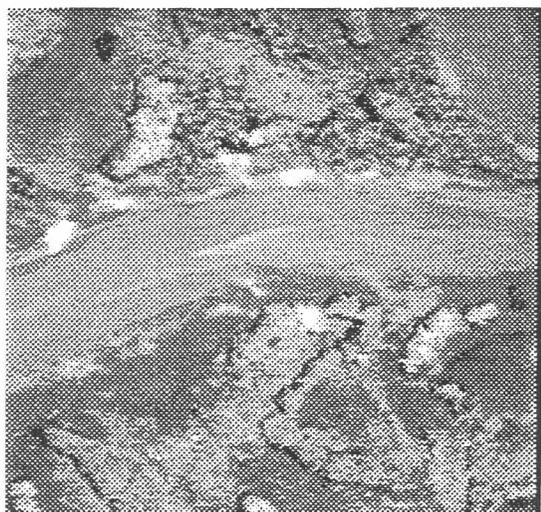
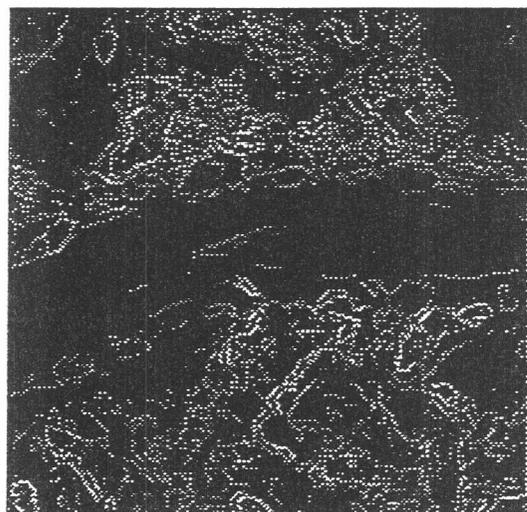
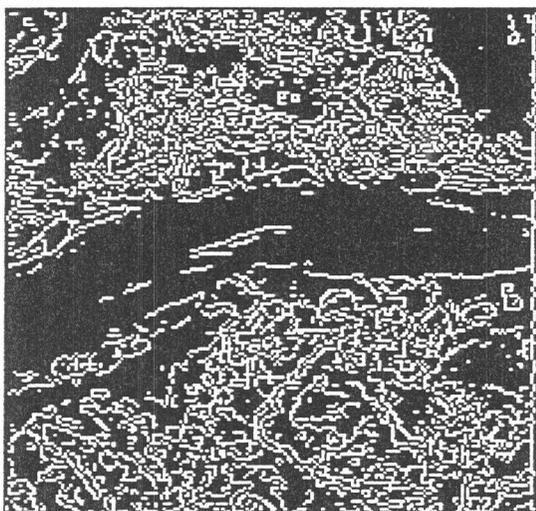


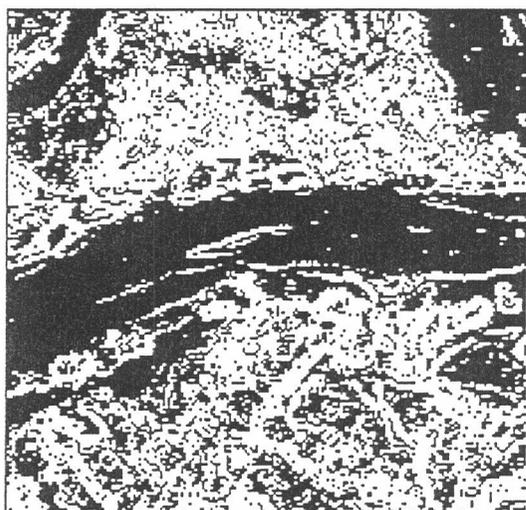
Fig. 5.1 a) Imagen derecha original



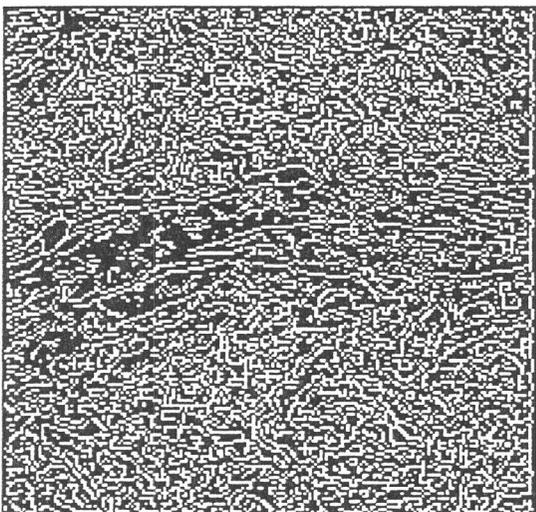
b) Detec. Contornos:25 s/filtro s/Binarizar



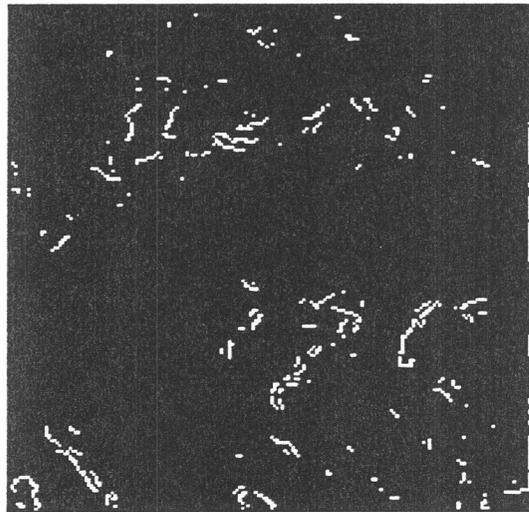
c) Detec. Cont.:25 c/Filtro Binarizada



d) Detec. Contornos:25 Binarizada



e) Detec. Cont.: 5 c/filtro Binarizada



f) Detec. Cont. :100 c/filtro Binarizada

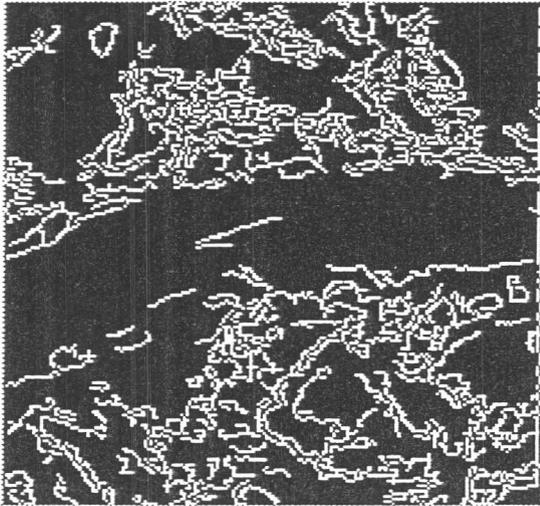
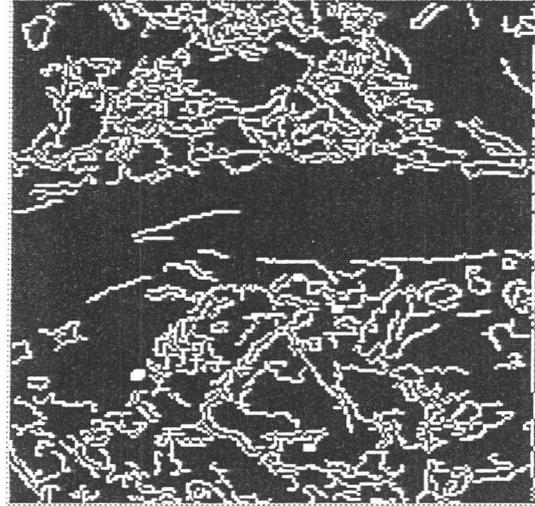
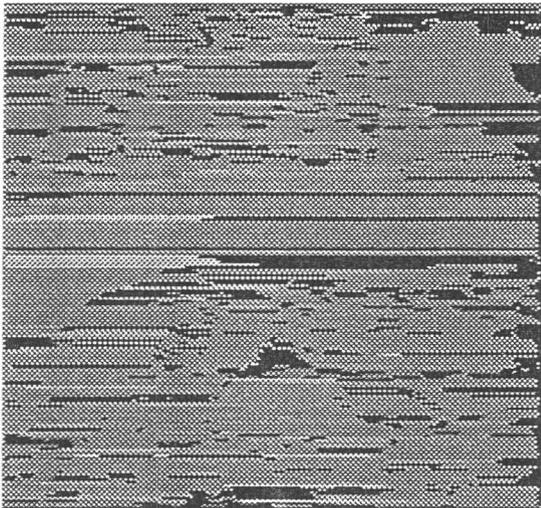


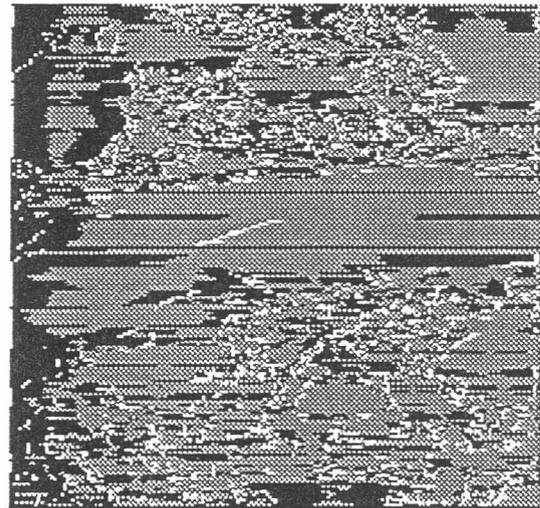
Fig. 5.2 a) Imagen derecha segmentada



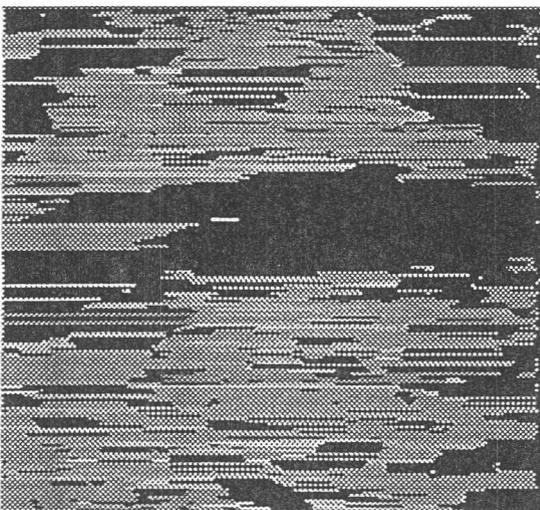
b) Imagen izquierda segmentada



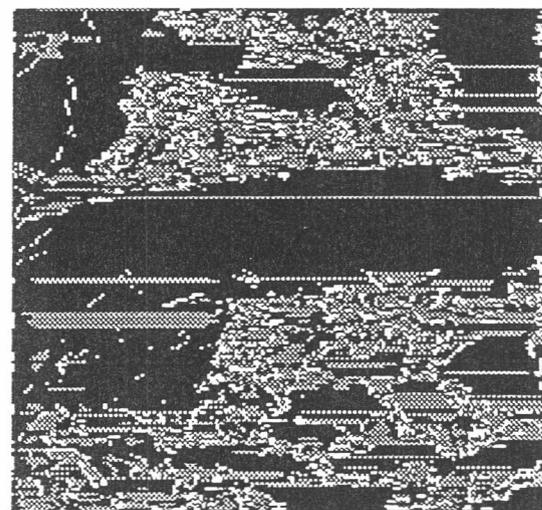
c) Imagen de disparidad (Prueba 2)



d) Imagen derecha reconstruida (Prueba 2)



e) Imagen Disparidad (Prueba 10)



f) Imagen derecha reconstruida (Prueba 10)

Las imágenes de la Fig. 6.1 muestran diversos ejemplos referentes a la detección de contornos, donde se visualizan algunos de los problemas que pueden resultar de una poda incorrecta; por ejemplo la imagen (e) resulta muy densa para un tratamiento satisfactorio en la puesta en correspondencia, en el otro extremo la imagen (f) es poco representativa de la escena original. En el ejemplo (d) donde se excluyó la aplicación del filtro de Chehikian los contornos son confusos y muy superiores a un pixel.

Respecto a los ejemplos de la Fig.6.2 se pueden ver dos imágenes de disparidad generadas de acuerdo a los valores de la tabla de resultados previamente presentada y la imagen derecha reconstruida correspondiente generada a partir de la imagen izquierda y la imagen de disparidad. En las imágenes reconstruidas los valores en grises corresponden a las playas y los blancos a contornos con correspondencias en el par estereoscópico. Los pixeles negros corresponden a objetos sin correspondencia.

7

~ Conclusiones ~

Respecto de los resultados cualitativos se analizó el comportamiento asignando peso a una propiedad por vez .

- **Atributo Longitud:**

En las imágenes tratadas, cuando aparecen contornos estos son en general de longitud uno, por lo que no se ven restringidos en la posibilidad de apareo; mientras que las playas que si pueden tener diferentes longitudes, se ven severamente afectadas si la cota ISNC es muy restrictiva. Esto se observa claramente en la prueba #10.

- **Característica Gris Inicial y Gris Final:**

Cuando las imágenes originales tienen diferentes condiciones de luminosidad esto afecta los niveles de grises en la imagen y en consecuencia otorgar más peso a los grises inicial y final afecta la posibilidad de apareo llevándolo incluso a relacionar objetos erróneos.

- **Característica Orden:**

Este parámetro ofrece versatilidad para poder considerar apareos cuando hay más objetos en una imagen que en la otra y no es recomendable que tenga un valor alto porque no genera resultados confiables.

- **Característica Posición:**

Evaluar éste parámetro por sí solo no genera resultados interesantes, y si resulta significativo como complemento a los demás.

- **Cota ISNC:**

El valor de esta cota gradúa los apareos posibles puesto que la cantidad de correspondencias es inversamente proporcional a la calidad de las mismas, es decir si es muy grande se obtienen muchas correspondencias pero de calidad dudosa, y si es demasiado pequeño los apareos pueden ser buenos pero escasos.

- **Valor ICNC:**

Este valor siempre es mayor al ISNC y sirve para reconsiderar casos de no correspondencia y retomarlos si tiene un valor cercano a la cota ISNC.

Al considerar cotas altas, (ICNC = 60, ISNC=40) se obtuvieron mapas de disparidad con una alta cantidad de apareos, en una amplia gama de imágenes. Con valores bajos (ICNC=15, ISNC=12) la cantidad de apareos se reduce notoriamente conservando correspondencia en los objetos bien definidos y poco densos. El resultado de considerar cotas muy estrictas en el apareo (ICNC = 100, ISNC = 2), produjo un mapa de disparidad muy ralo con correspondencias muy confiables pero absolutamente carente de interés para el cálculo de un modelo numérico.

La herramienta presentada para el apareo, el algoritmo de Viterbi, corresponde a una técnica de programación dinámica para búsqueda automática de puntos homólogos en imágenes estereoscópicas Este algoritmo es frecuentemente utilizado.

Al realizar distintas evaluaciones, se denota cierta dificultad para determinar los

parámetros óptimos con los que se obtuviera una mejor correspondencia, como también determinadas limitaciones.

Se podrían plantear diversas mejoras en distintas etapas del proceso, en particular en la detección de bordes. También se podrían estudiar e instrumentar ciertas simplificaciones a los efectos de reducir tiempos de proceso en el cálculo de áreas y en el algoritmo de apareo.

Queda abierta la posibilidad de integrar esta técnica con estrategias de control, Jerárquica o grueso a fino (coarse to fine), en búsqueda de un proceso global mas robusto.

~ ANEXO I ~

El apareo estereoscópico consiste en la identificación de las proyecciones de un mismo punto objeto en cada imagen, a éstas proyecciones se las llama puntos homólogos. Una vez realizado, con imágenes roto rectificadas, se obtiene la imagen de disparidad o bajo ciertas condiciones imagen de paralaje. A partir de la imagen de disparidad, conociendo el desfase relativo a la disparidad y contando con los valores de los parámetros internos fundamentalmente distancia principal, y como externo la base de toma se puede calcular sin dificultad el modelo numérico de la escena que dió lugar a la imagen estereoscópica en algún referencial general elegido.

A esto se lo llama restitución dado que consiste en restituir las coordenadas 3D de los puntos que constituyen el modelo numérico, y permite calcular valores de profundidad, distancia a los sensores, tomando valores referidos a ese referencial general elegido.

En este anexo se presenta primeramente el método de obtención de coordenadas tridimensionales para el caso llamado caso normal y que resulta de haber utilizado en el apareo estereoscópico imágenes que se encuentren roto rectificadas. El problema geométrico se simplifica notablemente dado que además de suponer que ambas imágenes tienen idéntica distancia principal, están en un mismo plano paralelo al (x,y) , un punto de vista está en el origen del referencial 3D, el otro sobre el eje x a una distancia b llamada base de toma.

Posteriormente, se da la solución al problema general, ésta permite obtener las coordenadas tridimensionales (x,y,z) de un punto, en un referencial general como funciones de las coordenadas de los puntos homólogos:

$$\begin{aligned} X &= f_x(x_1, y_1, x_2, y_2), \\ Y &= f_y(x_1, y_1, x_2, y_2), \\ Z &= f_z(x_1, y_1, x_2, y_2). \end{aligned}$$

Por último, se presenta la transformación inversa que permite obtener según el caso las coordenadas de las proyecciones de los puntos del espacio objeto en cada espacio imagen, como resultado de :

$$\begin{aligned} x &= f(X, Y, Z) \\ y &= h(X, Y, Z) \end{aligned}$$

o la rotación del referencial imagen respecto del referencial general.

Generalidades

Si se quiere localizar un sólido en el espacio se deben conocer 3 parámetros de rotación y 3 de translaciones. Cuando se trata de un haz perspectivo, las coordenadas tridimensionales del punto de vista y la dirección del eje óptico y la de los ejes coordenados del plano de la imagen perpendicular a dicho eje determinan esos seis parámetros.

Llamamos *línea base* o *base* a la distancia entre los puntos de vista, y *distancia principal* a la distancia del punto de vista al plano imagen .

1-Caso Normal

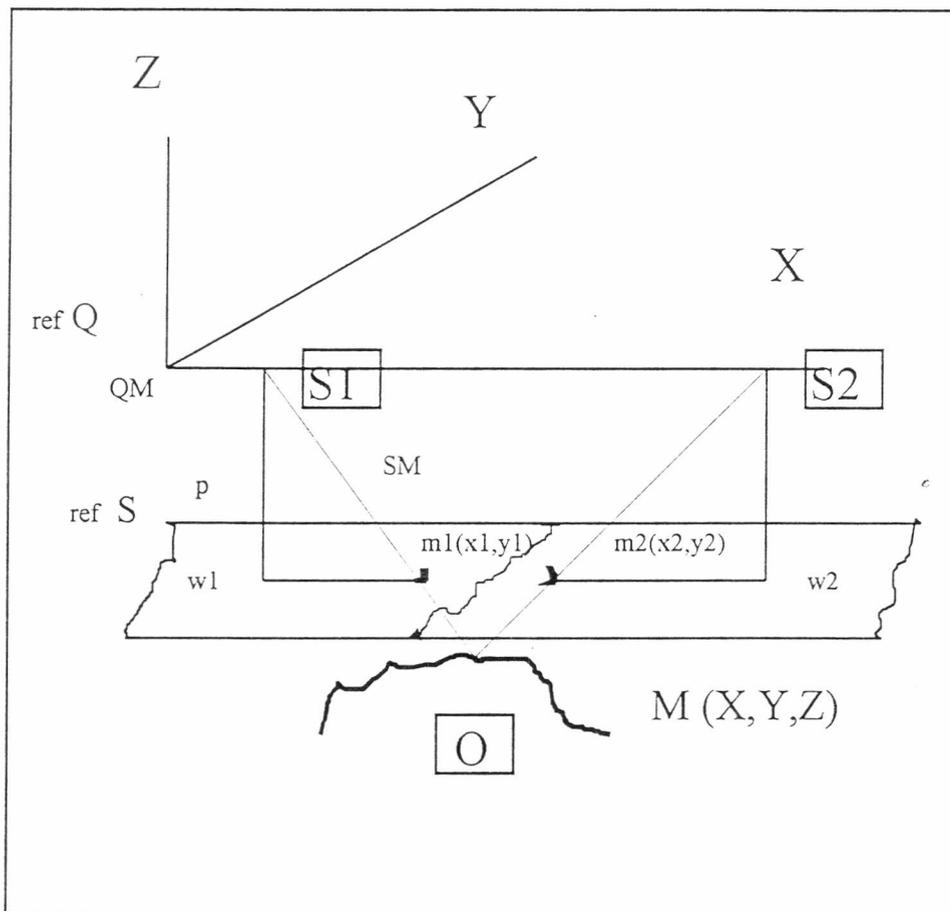
Es el caso en que el punto de vista de la imagen 1, S_1 es coincidente con el origen del referencial general, y el punto de vista de la imagen 2, S_2 está a una distancia b de S_1 sobre el eje X , si se cumple además que a una cierta distancia p sobre el eje Z se encuentra un plano que contiene las dos imágenes. Este caso es especial y se resuelve sin la necesidad de conocer las rotaciones, ya que las mismas son nulas.

Para determinar X, Y, Z , se aplica la relación de triángulos semejantes en función de las coordenadas de las imágenes

$$Z = \frac{p \cdot b}{x_1 - x_2} \quad (\text{donde } x_1 - x_2 \text{ es el paralaje longitudinal})$$

$$X = \frac{Z \cdot x_1}{p}$$

$$Y = \frac{z \cdot y_1}{p}$$



2-Caso General

Solución no canónica para la obtención de coordenadas 3D en el caso general

Supongamos un referencial general Q y el referencial de la imagen con origen S. Conocida la proyección m en el referencial S, se quieren encontrar las coordenadas de M en el referencial general Q.

Definimos : referencial Q con ejes (X,Y,Z)
 referencial S con ejes (x,y,z)

El punto M a considerar tendrá estará definido por un vector M_q en el referencial general y uno M_s en el referencial S.

Lo que se busca es conocer las coordenadas de M_q , se plantea:

$$M_q = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad M_s = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad S = \begin{pmatrix} X_s \\ Y_s \\ Z_s \end{pmatrix}$$

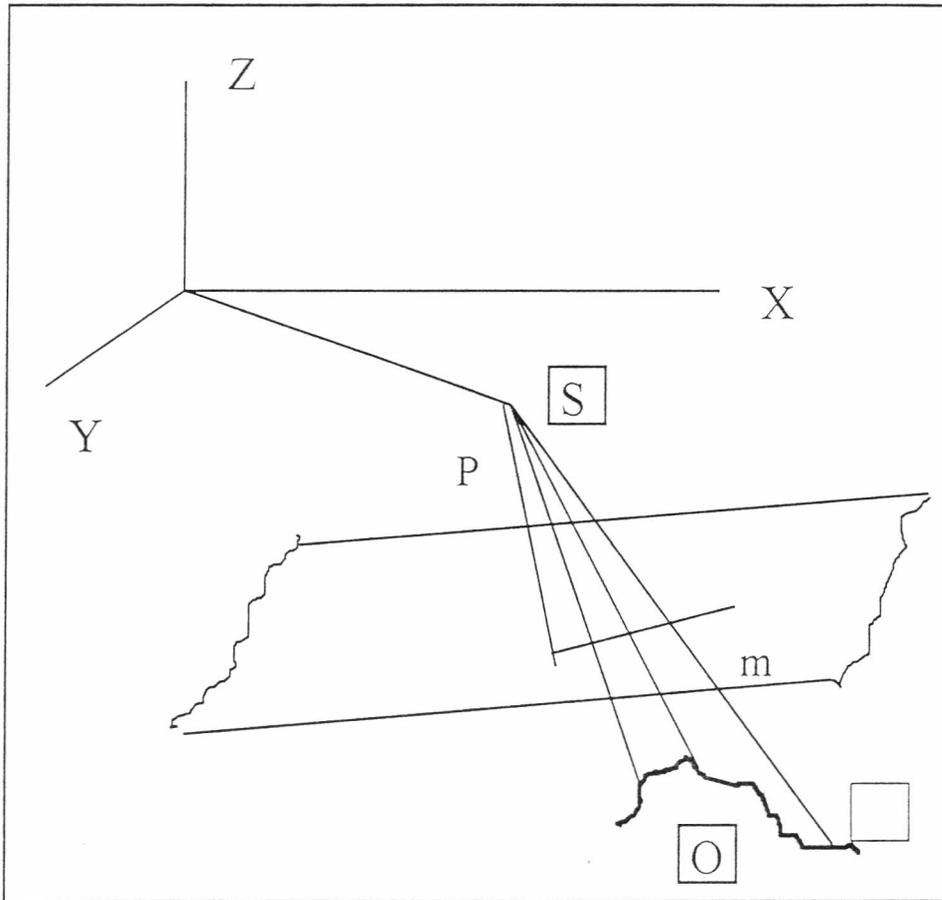
S es un vector en el espacio, en el referencial Q. Se define como la resta vectorial:

$$SM = M_q - S$$

$$M_s = R \cdot (M_q - S) \quad \text{donde R es una matriz de rotación}$$

y se puede escribir como :

$$M_s = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = R \cdot \left(\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} - \begin{pmatrix} X_s \\ Y_s \\ Z_s \end{pmatrix} \right) \quad (1)$$



Dado que R y S son valores conocidos se puede escribir:

$$\begin{pmatrix} L1 \\ L2 \\ L3 \end{pmatrix} = -R \cdot \begin{pmatrix} Xs \\ Ys \\ Zs \end{pmatrix} \quad (2)$$

reemplazando (2) en (1)

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} L_1 \\ L_2 \\ L_3 \end{pmatrix}$$

que es lo mismo que escribir:

$$x = r_{11}X + r_{12}Y + r_{13}Z + L_1$$

$$y = r_{21}X + r_{22}Y + r_{23}Z + L_2$$

$$z = r_{31}X + r_{32}Y + r_{33}Z + L_3$$

Como se ha dicho se desconoce la posición de M_q , pero se tiene como dato R y S y las siguientes igualdades:

$$\frac{u}{p} = \frac{x}{z} = U \qquad \frac{v}{p} = \frac{y}{z} = Z$$

donde u , v y p son las coordenadas de la proyección m en el referencial S , luego:

$$x = \frac{u}{p} \cdot z \qquad y = \frac{v}{p} \cdot z$$

reemplazando

$$x - \frac{x}{z} \cdot z = 0 \qquad y - \frac{y}{z} \cdot z = 0$$

sustituyendo

$$r_{11}X + r_{12}Y + r_{13}Z + L_1 - \frac{x}{z}[r_{31}X + r_{32}Y + r_{33}Z + L_3] = 0$$

$$r_{21}X + r_{22}Y + r_{23}Z + L_2 - \frac{y}{z}[r_{31}X + r_{32}Y + r_{33}Z + L_3] = 0$$

reemplazando por U, V las formulas anteriores y sabiendo que:

$$(r_{11}, r_{12}, r_{13}) = \bar{i}R \quad (r_{21}, r_{22}, r_{23}) = \bar{j}R \quad (r_{31}, r_{32}, r_{33}) = \bar{k}R$$

$$\bar{i}R.Mq - U.\bar{k}R.Mq + L_1 - U.L_3 = 0$$

$$\bar{j}R.Mq - V.\bar{k}R.Mq + L_2 - V.L_3 = 0$$

Si lo escribimos en forma matricial

$$\begin{bmatrix} \bar{i}R - U.\bar{k}R \\ \bar{j}R - V.\bar{k}R \end{bmatrix} Mq = \begin{bmatrix} -U.L_3 + L_1 \\ -V.L_3 + L_2 \end{bmatrix} = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}$$

Siendo conocidos U, V, S y la matriz R. Queda entonces un sistema con dos ecuaciones y tres incognitas que son $Mq = (X, Y, Z)$. A partir de cada imagen se puede plantear este juego de ecuaciones y en ambos juegos solo Mq es incognita con lo que el problema queda resuelto usando tres de las cuatro o aplicando cuadrados mínimos.

3-Obtención de las coordenadas de una perspectiva o cálculo de la rotación del referencial imagen respecto del referencial general

Si no existieran rotaciones alrededor de ninguno de los ejes de coordenadas, y siendo (u, p, v) las coordenadas de una perspectiva m en la imagen.

Tenemos :

$$X - X_s = F . u$$

$$Y - Y_s = F . v$$

$$Z - Z_s = F . p$$

donde (X, Y, Z) son las coordenadas de M y (X_s, Y_s, Z_s) son las coordenadas del origen del referencial imagen en el referencial general.

Si consideramos rotaciones alrededor de los ejes tenemos:

$$Mq - Ms = R^T \cdot Ms = R^T \phi \begin{vmatrix} u \\ v \\ p \end{vmatrix} \quad \text{con } F = x/u = y/v = z/p$$

$$\begin{vmatrix} X \\ Y \\ Z \end{vmatrix} - \begin{vmatrix} Xs \\ Ys \\ Zs \end{vmatrix} = \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix} \begin{vmatrix} \phi \cdot u \\ \phi \cdot v \\ \phi \cdot p \end{vmatrix}$$

y obtenemos las siguientes ecuaciones:

$$\begin{aligned} X - Xs &= \phi \cdot r_{11} \cdot u + \phi \cdot r_{12} \cdot v + \phi \cdot r_{13} \cdot p \\ Y - Ys &= \phi \cdot r_{21} \cdot u + \phi \cdot r_{22} \cdot v + \phi \cdot r_{23} \cdot p \\ Z - Zs &= \phi \cdot r_{31} \cdot u + \phi \cdot r_{32} \cdot v + \phi \cdot r_{33} \cdot p \end{aligned}$$

Según los datos que se conozcan se pueden obtener soluciones a dos problemas distintos. Si se conoce la matriz de rotación y la posición del punto (Xs, Ys, Zs) pueden establecerse los parámetros u y v , siendo p conocido, por otro lado si se conocen los últimos y las coordenadas de S se pueden obtener los coeficientes de la matriz de rotación.

~ Bibliografía ~

- [1] **Etienne Roger**
RECONSTITUTION DU RELIEF PAR SEGMENTATION ET MISE EN CORRESPONDANCE D'OBJETS
DE COUPLES D'IMAGES STEREOCOPIQUES .
Ecole Nationale des Telecommunications Paris. 1986

- [2] **A. Gerhard, H. Platzer, J. Steurer, R. Lenz**
DEPTH EXTRACTION BY STEREO TRIPLES AND FAST CORRESPONDENCE ESTIMATION
ALGORITHM.
Lehrstuhl fur Nachrichtentechnik, Technische Universitat Munchen. IEEE 1986

- [3] **Nicolas Ayache , Francis Lutzman**
FAST AND RELIABLE PASSIVE TRINOCULAR STEREOVISION.
INRIA - Domaine de Voluceau - France - IEEE 1987

- [4] **R. Volden, J. G. Balchen**
DETERMINING 3-D OBJECT COORDINATES FROM A SEQUENCE OF 2-D IMAGES.
Department of Engineering Cybernetics, Norway June 1993

- [5] **Michel Benard**
RESTITUTION AUTOMATIQUE EN STEREO PHOTOGRAMMETRIE.
Départament I.S.S.V., Laboratoire Image Nov. 1983

- [6] **Jacques Facon**
PROCESSAMENTO E ANALISE DE IMAGENS .
Centro Federal de Educacao Tecnológica do Paraná
IV Escola Brasileiro-Argentina de Informática. Julho 1993

- [7] **A. Chehikian, J. Hérault, J. E. Mendez**

UN PROCESSEUR DE DETECTION DE CONTOURS EN TEMP REEL DANS UNA IMAGE DE TELEVISION.

Laboratoire de Traitement d'Images et Reconnaissance de Formes. ANN. Télécommun, 38, nro 3-4, 1983

- [8] **Zelasco Jose, Donadío Alejandro, Mendez Daniel E.**

ESTEREOPSIS AUTOMATICA. PANORAMA DEL ESTADO DEL ARTE.

Trabajo aceptado para publicarse en 1997 en los Anales de la Sociedad Científica Argentina

- [9] **G. David Forney Jr.**

THE VITERBI ALGORITHM.

Proceedings of the IEEE, vol 61, no 3, March 1973

- [10] **Zelasco José Francisco**

UTILISATION DES METHODES DE ESTEREPHOTOGRAMETRIE A L'AMELIORATION DE LA STEREOGRAPHIE AUTOMATIQUE SUR LES IMAGES NUMERIQUES

IV Simposio Latinoamericano de Percepción Remota, 1989

- [11] **Yuichi Ohta, Masaki Watanabe, and Katsuo Ikeda**

IMPROVING DEPTH MAP BY RIGHT-ANGLED TRINOCULAR STEREO

Institute of Information Sciences and Electronics. University of Tsukuba Japan IEEE - 1986

- [12] **J.R.Parker**

PRACTICAL COMPUTER VISION USING C

August 1993

- [13] **Zhang Zheng You**

LE PROBLÈME DE LA MISE EN CORRESPONDANCE: L'ÉTAT DEL L'ART

Research Report Nro 2146, INRIA Sophia-Antipolis, Francia, Diciembre 1993

- [14] **Zelasco José Francisco**
CURSO DE ESTEREOCOPIA DIGITAL
Facultad de Ciencias Exactas, UNCPBA, 1993
- [15] **Faugeras O., Hotz B., Mathiew H.**
REAL TIME CORRELATION-BASED STEREO: ALGORITHM IMPLEMENTATION AND
APLICACIONES
Research Report 2013, INRIA, Sophia-Antipolis, Francia, 1993
- [16] **Foley - Freiser - Hughes**
COMPUTERS GRAPHICS - PRINCIPLES AND PRACTICE
Second Edition , November 1992
- [17] **Staugaard Andrew C.**
ROBOTICS AND AI - AN INTRODUCTION TO APPLIED MACHINE INTELLIGENCE
1993
- [18] **Stephen Banks**
SIGNAL PROCESSING IMAGE PROCESSING AND PATTERN RECOGNITION
Prentice Hall, 1990.