UNIVERSITAT JAUME•I

Escuela Superior de Tecnología y Ciencias Experimentales

Departamento de Informática

# Técnicas de Visión Estereoscópica basadas en Regiones para la Obtención de Representaciones 3D de la Escena

## TESIS DOCTORAL

presentada por

M. Ángeles LÓPEZ MALO

dirigida por

Filiberto PLA BAÑÓN

Castellón, Septiembre de 1999

**A Pepe**

Resumen

La visión estereoscópica engloba un amplio abanico de técnicas para la correspondencia y la reconstrucción tridimensional. Las técnicas basadas en regiones son escasas en la literatura comparadas con las técnicas basadas en segmentos de los bordes, o las técnicas basadas en áreas. Esto es debido fundamentalmente a que los resultados de los métodos de segmentación varían mucho de una imagen a otra, lo que hace más difícil el emparejamiento de regiones.

En esta tesis se abordan estos problemas generados por las diferencias en la segmentación de ambas imágenes del par estéreo mediante dos técnicas distintas. La primera es una técnica clásica basada en grafos, de la que existen numerosas aplicaciones en la literatura, a la que se ha añadido un preproceso, también basado en grafos, para resolver los problemas de la segmentación. La segunda es una nueva técnica de correspondencia en la que sólo se segmenta una de las dos imágenes, y se buscan las correspondencias en la otra imagen basándose en la minimización de una función de energía y aplicando ciertas restricciones sobre la profundidad de las regiones. En esta técnica, las correspondencias, la reconstrucción tridimensional y las oclusiones se obtienen en un solo proceso cooperativo.

**Palabras clave:** visión estereoscópica, segmentación, correspondencia, reconstrucción 3D, detección de oclusiones, correlación, grafo de asociación, *clique* máximo, minimización de una función de energía.

# Abstract

Stereoscopic vision includes a wide range of techniques for correspondence and three-dimensional reconstruction. Region-based techniques have been hardly addressed in the literature compared with edge-based techniques or area-based techniques. This is due mainly to the fact that the results of segmentation methods vary significantly from one image to another, which makes the matching of regions more difficult.

In this thesis, the problems generated by the differences in segmentation from both images of the stereo pair are faced by means of two different techniques. First, we use a classic graph-based technique, widely used in the literature, where we add a new pre-process stage, which is a graph-based step, in order to solve the segmentation problems.

Second, we present a novel matching technique where only one of the images is segmented, and correspondences are searched in the other image by means of the minimization of an energy function and the application of some constraints on the depth of the regions. In this technique, correspondences, three-dimensional reconstruction and occlusions are obtained in an only cooperative process.

**Keywords:** stereoscopic vision, segmentation, correspondence, 3D reconstruction, occlusion detection, correlation, association graph, maximal clique, minimization of an energy function.

# Agradecimientos

Quisiera poder agradecer a todas las personas que me han ayudado, acompañado, atendido y animado a lo largo de los años que ha llevado la elaboración de esta tesis doctoral en la misma medida en que yo he recibido su apoyo, pero el material y el tiempo con el que cuento es ínfimo comparado con todo lo que tengo que decir. Así pues, sirva esta página como pequeño e incompleto resumen de mi gratitud a todas estas personas.

En primer lugar, a mis padres, por su apoyo y por ser el seno donde nació mi afición por el estudio. A mi hermana Marisa, por ser mi mejor amiga. A Laura y Alberto, y a toda mi familia, por enseñarme tantas cosas.

A mis amigos, Isabel, Pedro, Maribel, Rafa, Luis, Gloria, Antonio, Mar, Nuria, Jose, Naly, Germán, Feli, Inma, Merche, Carlos, Germán, Belén, Cristina, Miguel, Mª José, Gloria V., Juan, Susan, Pablo, Christine, Jose Manuel, Víctor... y un largo etcétera, por las tertulias, por los paseos, por sus consejos, por sus ánimos, por ser quienes son, en definitiva, por su amistad.

A Filiberto Pla, por ser mi director de tesis, por sus valiosos consejos y por guiarme en el arduo aprendizaje de la investigación. A mis compañeros del Grupo de Visión por Ordenador, por el trabajo en grupo y por hacer más fácil el día a día.

A todos los miembros del Departamento de Informática de la Universitat Jaume I, en los que siempre he encontrado colaboración y compañerismo. A Tico Cervera y Vicent Palmer, del Departamento de Matemáticas, por su paciencia y sus respuestas a mis preguntas. A Angel P. del Pobil, investigador principal del proyecto TAP95-0710, por proporcionarme el marco en el que se inscribe el trabajo de esta tesis. A la CYCIT, Ministerio de Educación y Ciencia, por financiar el proyecto.

A los miembros del proyecto ROBOTVIS del INRIA, Sophia Antipolis (Francia), por su cálida acogida durante 8 meses repartidos en los años 1996 y 1997, por comprender mis dificultades con el idioma, por las valiosas charlas científicas, y por proprocionarme un extraordinario ambiente de trabajo. A Soraya, Stephane, Robert, Mª Cristina, Jerome, Marie Cecile, Andrés, Zorabelia, César, Elisabeth, Tania, y tantos otros, por su amistad. Al Dr. Olivier Faugeras, director de ROBOTVIS, por aceptarme en su proyecto, por facilitarme el acceso a la extensísima biblioteca y el material necesario, y por sus valiosas sugerencias. A la Fundació Caixa Castelló, por financiar las estancias.

A los desarrolladores de TarjetJr, por proporcionarme un software que, aunque todavía en desarrollo, podría llegar a ser una plataforma de intercambio científico en visión por ordenador, y especialmente a Peter Vanroose, por buscar soluciones a los múltiples problemas de instalación. A los Drs. Rachid Deriche, Luc Robert y Theo Papadopoulos, por su accesibilidad y sus interesantísimas conversaciones. Y en especial, al Dr. Michael Brady, por contagiarme su inconsumible entusiasmo por la investigación.

A todos los investigadores e investigadoras que han contribuido a que la visión estereoscópica sea hoy una línea abierta de investigación, por sus contribuciones a la ciencia, y en particular, por mostrarme en sus escritos este fantástico mundo.

Y finalmente, a Pepe, a quien dedico este trabajo, por absolutamente todo.

Angeles López,
Septiembre 1999.

# Índice General

# Índice de Figuras

# Índice de Algoritmos

# Índice de Tablas

# Notación y Lista de Símbolos

## Notación

| | |
|---|---|
| $[a, b]$ | Intervalo continuo de valores |
| $[a..b]$ | Intervalo discreto de valores |
| $\lvert D\rvert$ | Si $D$ es un intervalo (continuo o discreto), $\lvert D\rvert$ indica su tamaño. |
| | Si $D$ es un conjunto de elementos, $\lvert D\rvert$ indica el cardinal del conjunto. |
| | Si $D$ es un escalar, $\lvert D\rvert$ indica su valor absoluto. |

## Sistemas de coordenadas

| | |
|---|---|
| $(X_w, Y_w, Z_w)$ | Sistema de coordenadas del mundo |
| $(X_c, Y_c, Z_c)$ | Sistema de coordenadas de la cámara |
| $(U, V)$ | Sistema de coordenadas del plano imagen |
| $(u, v)$ | Sistema de coordenadas normalizado de la imagen |
| $(u_0, v_0)$ | Origen del sistema de coordenadas del plano imagen |

## Símbolos

Las siguientes tablas presentan los símbolos y variables más importantes de cada capítulo:

| **En todos los capítulos** | |
|---|---|
| $I_1, I_2$ | Imágenes del par estereoscópico |
| $I'_k$ | Primera derivada de la imagen $I_k$ (gradiente de $I_k$) |
| $M$ | Punto en 3D |
| $m, m_i, m', m'_i$ | Punto en 2D |
| $\tilde{M}, \tilde{m}$ | Puntos $M$ y $m$ en coordenadas homogéneas |
| $P$, $P'$ | Matrices de proyección |
| $F$ | Matriz fundamental |
| $l'_m$ | Línea epipolar de un pixel $m$ de una imagen en la otra imagen |

| **En todos los capítulos** | |
|---|---|
| $C, C'$ | Centros ópticos de las cámaras |
| $f$ | Longitud focal |
| $b$ | Distancia entre las dos imágenes de un sistema binocular paralelo |
| $d$ | Disparidad |
| $\mathcal{D}$ | Intervalo de disparidades posibles |
| $d_{min}, d_{max}$ | Disparidad mínima y máxima, respectivamente |
| $Z_{min}, Z_{max}$ | Profundidad mínima y máxima, respectivamente |

| **Capítulo 2** | |
|---|---|
| $W(m)$ | Vecindario del pixel $m$ |
| $N$ | Número de pixeles del vecindario |
| $\overline{I_k(m)}$ | Intensidad media del vecindario de $m$, de la imagen $I_k$ |
| $\sigma_k(m)$ | Desviación típica de la intensidad del vecindario de $m$, imagen $I_k$ |
| $C_{ccc}(m, m')$ | Correlación entre el vecindario de $m$ y el de $m'$ por el criterio *ccc* |

| **Capítulo 3** | |
|---|---|
| $L_i, R_j$ | Regiones de la imagen izquierda y derecha, respectivamente |
| $N, M$ | Número de regiones en la imagen izquierda y derecha, respectivamente |
| $S(L_i, R_j)$ | Similitud entre dos regiones |
| $B(L_i, R_j; L_h, R_k)$ | Compatibilidad entre dos emparejamientos de regiones |
| $A(L_i)$ | Número de pixeles de la región $L_i$ |
| $T_s$ | Umbral de similitud mínima para considerar un par de regiones como una correspondencia posible |
| $L_{(p,q,\ldots,t)}$ | Unión de las regiones $L_p$, $L_q$, ... ,$L_t$. |
| $L_{i,j,d}, R_{i,j,d}$ | Zonas de intersección de las regiones $L_i$, $R_j$ a disparidad $d$ en la imagen izquierda y en la imagen derecha, respectivamente |
| $\mu_{L_{i,j,d}}, \mu_{R_{i,j,d}}$ | Media de intensidad de $L_{i,j,d}$ y $R_{i,j,d}$ |
| $\sigma_{L_{i,j,d}}, \sigma_{R_{i,j,d}}$ | Desviación típica de la intensidad de $L_{i,j,d}$ y $R_{i,j,d}$ |
| $C_{ccc}(L_i, R_j, d)$ | Correlación entre regiones $L_i$ y $R_j$ a disparidad $d$, criterio *ccc* |
| $C_{ij}(d)$ | Correlación entre regiones $L_i$ y $R_j$ a disparidad $d$ normalizada |
| $N_{ij}(d), N^!_{ij}(d)$ | Coeficiente de coincidencia de $L_i$ y $R_j$ a disparidad $d$ |
| $\hat{d}_{ij}$ | Mejor disparidad entre las regiones $L_i$ y $R_j$ |
| $K$ | Número de nodos para la búsqueda del mejor *clique* máximo en el algoritmo subóptimo |

| **Capítulos 4 y 5** | |
|---|---|
| $\mathcal{Z}$ | Mapa de profundidades |
| $m, m_i$ | Pixel de la imagen de referencia, $m = (u, v)$ |
| $m', m'_i$ | Pixel correspondiente a $m, m_i$ |
| $R, R_i$ | Región de la imagen de referencia |
| $R', R'_i$ | Región correspondiente a $R, R_i$ |
| $f_{12}$ | Función para calcular $m'$ a partir de $m$ y su profundidad |
| $g_{12}$ | Función para calcular $R'$ a partir de $R$ y su profundidad |
| $F_z, F_p, F_q$ | Derivadas parciales de $F$ respecto a $z, p, q$ |
| $\mu_k(R)$ | Intensidad media de la región $R$ de la imagen $I_k$ |
| $\sigma_k(R)$ | Desviación típica de la intensidad de la región $R$, imagen $I_k$ |
| $r_{ij}, r'_{ij}$ | Elementos de la matriz de rotación |
| $t_k, t'_k$ | Elementos de los vectores de traslación, $k = x, y, z$ |

# Capítulo 1

# Introducción

El trabajo de investigación que se presenta en esta tesis se ha llevado a cabo en el *Grupo de Visión por Ordenador* de la *Universitat Jaume I* de Castelló, y es producto de una colaboración con el *Grupo de Robótica Inteligente* de la misma universidad, en un proyecto financiado por la CICYT (Ministerio de Educación y Ciencia), TAP95-0710, titulado *Movimiento inteligente de un brazo robot con incertidumbre basado en visión artificial y sensores para tareas de extracción, desplazamiento e inserción.*

Este proyecto aborda el estudio, desarrollo e implementación de un sistema para la planificación del movimiento fino y grueso de un robot manipulador para tareas de extracción, desplazamiento e inserción de herramientas, aplicable a una célula de fabricación flexible. El sistema carece de un conocimiento completo previo de su entorno, que es adquirido mediante el uso de visión artificial –para el movimiento grueso– y sensores de fuerza –para el movimiento fino–. Para la planificación del movimiento grueso, se obtiene un mapa 3D del entorno mediante visión estereoscópica que junto con las posiciones de partida y objetivo del sistema, se integrará en un planificador del movimiento del brazo robot.

Aproximar el mundo que rodea al robot por medio de una representación poliédrica de los objetos de la escena, permitiría obtener de una forma general información de diferentes tipos de entorno, sin conocimiento a priori del tipo de objetos presentes, pudiendo así proporcionar al planificador del robot una información cuantitativa y cualitativa válida para la planificación del movimiento.

La mayor parte de las técnicas de planificación del movimiento (al menos para robots manipuladores) parten de un modelo geométrico previo, lo que a veces resulta irreal. Se pretende proporcionar un modelo cuantitativo espacial del entorno, para determinar la zona del espacio libre y poder planificar el movimiento grueso del robot sin colisiones. Por tanto, se desea obtener un mapa 3D del entorno de un robot, teniendo en cuenta las condiciones habituales de trabajo del mismo. La visión estereoscópica sirve para construir este tipo de mapa, ya que proporciona información tridimensional sobre la escena de una forma pasiva, sin interactuar con el entorno.

## 1.1 Objetivos de la presente tesis

La presente tesis doctoral se enmarca dentro de una de las áreas de la visión por ordenador, la visión estereoscópica, que es la base de muchas aplicaciones prácticas de visión por ordenador (inspección industrial, navegación de robots, manipulación de imágenes de satélite y de imágenes médicas, etc.) en las que es necesario calcular la distancia desde el observador a los objetos de la escena a partir de dos o más imágenes adquiridas con una o varias cámaras.

Las recientes investigaciones hacen de la visión estereoscópica una herramienta cada vez más útil tanto en sistemas estáticos (dos o más cámaras fijas) como en los dinámicos (generalmente dos cámaras en movimiento). En esta tesis nos centramos en la configuración mínima para la visión estereoscópica que consiste en un sistema estático de dos cámaras. Este tipo de sistemas permite obtener cierta información tridimensional utilizando el principio de la triangulación. Dado un elemento característico en una de las dos imágenes, y dada su proyección en la otra imagen (su elemento correspondiente) se puede calcular la distancia del observador al objeto representado por dicho elemento mediante una sencilla técnica de triangulación a partir de las coordenadas imagen de ambos.

A diferencia del caso más general de una única cámara en movimiento que captura secuencias de imágenes, los puntos de vista en los sistemas estereoscópicos no son arbitrarios. La geometría de una cabeza estereoscópica determina, a menudo con precisión, la transformación geométrica entre ambas cámaras. Esta información resulta muy útil para buscar cómo se corresponden entre sí los elementos que aparecen en las dos imágenes, y es aquí donde radica entonces la dificultad: en la búsqueda en una imagen, del elemento homólogo a un elemento dado en la otra imagen.

Se han desarrollado muchas técnicas para tratar este problema, conocido como el *problema de la correspondencia*, que difieren en la geometría de las imágenes, en las primitivas, en las restricciones empleadas, o en la estrategia de la correspondencia, como se puede ver en las numerosas revisiones de las técnicas de visión estereoscópica [Barnard y Fishler, 1982] [Dhond y Aggarwal, 1989] [Faugeras *et al.*, 1992] [Lane y Thacker, 1994] [Jones, 1997].

La selección del método de correspondencia todavía está muy ligado al área de aplicación de la misma, debido a la variedad de las metodologías empleadas y a las características de cada una de ellas. En el área de aplicación de la robótica, las escenas suelen ser interiores construidos por el hombre, de modo que los objetos suelen estar formados por superficies lisas, generalmente planos, y con escasa textura. Este tipo de objetos pueden ser fácilmente aproximados por planos u otro tipo de superficies simples, a diferencia de las escenas de exteriores donde resulta más difícil aproximar la forma de objetos (por ejemplo la vegetación o las ramas de un árbol). Por eso hemos seleccionado las regiones (obtenidas mediante algún método de segmentación) como primitivas de la correspondencia, ya que representan la proyección de las superficies tridimensionales en la imagen.

Un problema inherente en la reconstrucción tridimensional es la detección de las discontinuidades en la profundidad y las oclusiones, no solamente por su importancia en la creación de un mapa 3D de la escena, sino también porque a menudo suponen un problema importante en las técnicas de correspondencia.

El objetivo principal de esta tesis doctoral es la recuperación de la información tridimensional sobre la escena utilizando técnicas de visión estereoscópica binocular basadas en regiones. Para llevar a cabo este objetivo es necesario abordar los siguientes problemas:

- Búsqueda de la correspondencia entre regiones de un par estereoscópico de imágenes.
- Detección de las discontinuidades en la profundidad y de las oclusiones.
- Reconstrucción basada en regiones para la obtención de una representación poliédrica de la escena tridimensional.

Para abordar estos problemas, se han explorado dos líneas de trabajo diferentes: una técnica clásica basada en grafos, y una técnica nueva basada en minimización de una función de energía. En la primera técnica, se parte del trabajo previo realizado por otros autores y se resuelven algunos problemas que producen los errores de segmentación en la correspondencia. En la segunda técnica, se pretende evitar las diferencias en la segmentación entre ambas imágenes, realizando la segmentación de una imagen de referencia y buscando correspondencias en la otra imagen sin segmentar. En esta técnica, se aborda la detección de las oclusiones y la reconstrucción de la escena tridimensional como partes integrantes del proceso de correspondencia, de modo que se obtiene un modelo poliédrico de la escena sin necesidad de calcular el mapa de disparidades intermedio y segmentar éste a posteriori.

## 1.2 Organización de este documento

Esta tesis está organizada en los siguientes capítulos:

- En el capítulo 2, se expone un resumen el estado del arte en el campo de la visión estereoscópica. Debido al elevado y variado número de técnicas que intentan resolver el problema de la visión estereoscópica, no se ha pretendido revisarlas todas exhaustivamente, pero sí profundizar en aquellas técnicas más afines a los objetivos propuestos. Al final del capítulo se puede encontrar una discusión sobre las líneas en las que se enmarca el trabajo que se presenta en esta tesis.

- En el capítulo 3, se aborda el problema de la correspondencia de regiones basada en grafos de asociación, donde se realizan aportaciones al trabajo existente dirigidas a la resolución de los problemas que generan los errores en la segmentación y la aplicación de técnicas para reducir el tiempo de computación en la búsqueda de este tipo de soluciones.

- En el capítulo 4, se propone una nueva técnica de correspondencia de regiones basada en la minimización de una función de energía, que a su vez se basa en la correlación de las regiones correspondientes. Este método permite realizar la correspondencia y la reconstrucción tridimensional en un solo proceso, y obtener directamente una representación poliédrica de la escena.

- En el capítulo 5 se extiende el método para integrar la detección de oclusiones en el proceso de correspondencia en un proceso cooperativo que encuentra las oclusiones y las profundidades al mismo tiempo.

- En el capítulo 6 se resumen las conclusiones a las que se ha llegado tras la realización de este trabajo, enumerando las principales aportaciones del mismo al campo de la visión estereoscópica así como una discusión sobre las líneas de trabajo futuro.

Se ha incluido asimismo un apéndice sobre la geometría epipolar, donde se hace una breve exposición de algunos conceptos básicos que se manejan a lo largo de este documento.

# Capítulo 2

# Estado del arte

La visión estereoscópica es el conjunto de técnicas que tratan de la recuperación de la información tridimensional a partir de dos o más vistas de una escena.

En este proceso se pueden distinguir los siguientes pasos:

***Calibración:*** cálculo de los parámetros que relacionan las coordenadas de un punto de la escena con sus coordenadas en la imagen.

***Rectificación:*** transformación de la geometría del sistema estereoscópico, dados los parámetros de calibración, para simplificar tanto como sea posible la búsqueda de la correspondencia.

***Correspondencia:*** búsqueda del punto correspondiente en una imagen para cada punto característico de la otra imagen. Este proceso proporciona lo que se denomina un *mapa de disparidades*, preferiblemente un mapa denso.

***Reconstrucción:*** recuperación de la escena 3D, es decir, el cálculo de las profundidades de los puntos de la imagen, a partir de las disparidades de los mismos de modo que se obtiene un *mapa de profundidades.*

A menudo estos pasos se pueden unir o se pueden eliminar, como en el caso de los sistemas que no necesitan alguno o ninguno de los parámetros de calibración [Hartley *et al.*, 1992] [Hollinghurst y Cipolla, 1994] o que los estiman durante la correspondencia [Deriche *et al.*, 1994], o algunos métodos que realizan la correspondencia en un solo paso, sin rectificar las imágenes previamente y obteniendo directamente la estructura tridimensional [Robert y Deriche, 1996] [Super y Klarquist, 1997].

En los últimos años, gran parte de la investigación en visión estereoscópica se ha enfocado hacia el problema de la detección de oclusiones. En el pasado, los algoritmos de estereovisión tenían problemas en presencia de oclusiones y de discontinuidades en la profundidad, y la localización de las mismas solía tratarse generalmente como parte de la reconstrucción de la escena tridimensional [Gamble y Poggio, 1987].

Sin embargo, las investigaciones recientes muestran que las oclusiones pueden ayudar en el cálculo de la correspondencia. Esto supone un gran paso en la búsqueda de la correspondencia al considerar las oclusiones y discontinuidades en la profundidad no un problema a evitar, sino una fuente de información a tener en cuenta [Geiger *et al.*, 1992]. Sea cual sea el momento de abordarlas, hay dos problemas principales en lo concerniente a las oclusiones:

- ¿cómo identificar los puntos (o elementos) ocultos?
- ¿qué hacer con ellos?, es decir, ¿cómo determinar su profundidad?

En esta tesis nos centramos en la etapa de correspondencia, presentando dos técnicas de correspondencia basadas en regiones. En este capítulo se pretende revisar brevemente las técnicas de correspondencia más conocidas. En la sección 2.1 se introduce brevemente la geometría de un sistema de visión estereoscópico. En la sección 2.2 se revisan las principales metodologías de correspondencia existentes en la literatura, a modo de resumen del extenso conjunto de métodos que forman la visión estereoscópica. Finalmente, en la sección 2.3 se resaltan las líneas de trabajo en las que se enmarca esta tesis, delimitando las áreas afines al trabajo que se detalla en los siguientes capítulos.

## 2.1 Geometría de un sistema de visión estereoscópico

La calibración consiste en obtener la relación entre cualquier punto del mundo y su proyección en el plano imagen. Esta relación consiste en una serie de parámetros intrínsecos y extrínsecos, que son necesarios para realizar medidas en la imagen en muchas aplicaciones de visión. Se han desarrollado muchos algoritmos de calibración para sistemas monoculares, como la conocida técnica de Tsai [1987]. También se han desarrollado técnicas que extraen los parámetros de calibración a partir de una secuencia de imágenes [Robert, 1995] [Chang *et al.*, 1993].

En el caso de la visión estereoscópica, cada punto del mundo tiene dos proyecciones. La relación entre la geometría de ambas proyecciones se llama *geometría epipolar*. Muchas aplicaciones no necesitan una calibración total, sino solamente la recuperación de la geometría epipolar, la cual proporciona la estructura 3D en términos de geometría proyectiva. La estructura proyectiva no proporciona información métrica completa, pero contiene mucha información que a menudo es suficiente para aplicaciones como navegación y reconocimiento de objetos [Rothwell *et al.*, 1995] [Robert *et al.*, 1997] [Beardsley *et al.*, 1994]. En otro caso, existen métodos que recuperan una estructura afín [Faugeras, 1995] [Quan, 1993] o una estructura euclidiana a partir de la calibración proyectiva [Hartley *et al.*, 1992] [Mohr *et al.*, 1993] [Robert y Faugeras, 1995] [Devernay y Faugeras, 1996].

En nuestro caso, nos interesa la geometría de la proyección de la escena en los sistemas binoculares, que describimos someramente en esta sección y con más detalle en el apéndice A. Asímismo, en esta sección se sientan las bases de la notación que se ha seguido a lo largo de esta tesis, y se dan algunas definiciones básicas. En los siguientes capítulos se asume que se conocen los parámetros de calibración, o que las imágenes están rectificadas.

### 2.1.1 El modelo de cámara

Una imagen es la proyección de la escena tridimensional sobre el *plano imagen* de una cámara, también llamado *plano retínico*. Por tanto, el primer paso para calibrar los parámetros que permitan realizar mediciones en la imagen es determinar el modelo de la cámara que se esté usando.

El modelo de descripción de la cámara más conocido es el modelo de cámara *pinhole*, que asume una proyección perspectiva sin distorsión de la lente (figura 2.1). El punto $C$ es el llamado *centro óptico* de la cámara y el eje de la vista se denomina *eje óptico*. La distancia entre el centro óptico, $C$, y el plano imagen (que intersecta con el eje óptico en $c$) se denomina *longitud focal*, o simplemente, *focal*. Nos basaremos en este modelo de cámara que es el más utilizado en los sistemas de visión por ordenador.

Se han hecho muchos esfuerzos en la calibración de la distorsión de la lente, como son los trabajos de Nomura *et al.* [1992], Devernay y Faugeras [1995] y Zhang [1996c], que permiten la eliminación de esta distorsión. Esto permite que cualquier cámara pueda considerarse como una cámara *pinhole* mediante la aplicación de la inversa de la función de distorsión a los elementos de la imagen.

### 2.1.2 Sistemas de coordenadas y matriz de proyección

Para calcular los parámetros de la proyección, consideraremos los siguientes sistemas de coordenadas:

- Sistema de coordenadas del *mundo*: $(X_w, Y_w, Z_w)$ centrado en el origen $O$.
- Sistema de coordenadas de la *cámara*: $(X_c, Y_c, Z_c)$, centrado en el *centro óptico*, $C$, y donde el eje $Z_c$ es paralelo al *eje óptico*.
- Sistema de coordenadas del *plano imagen*: $(U, V)$, centrado en $c$, donde ambos ejes son paralelos a los ejes $X_c$ y $Y_c$.
- Sistema de coordenadas *normalizado*: $(u, v)$, que se encuentra también en el plano imagen, pero centrado generalmente en una de las esquinas de la imagen. Difiere del sistema de coordenadas del plano imagen en el origen y en las unidades de los ejes $u$ y $v$.

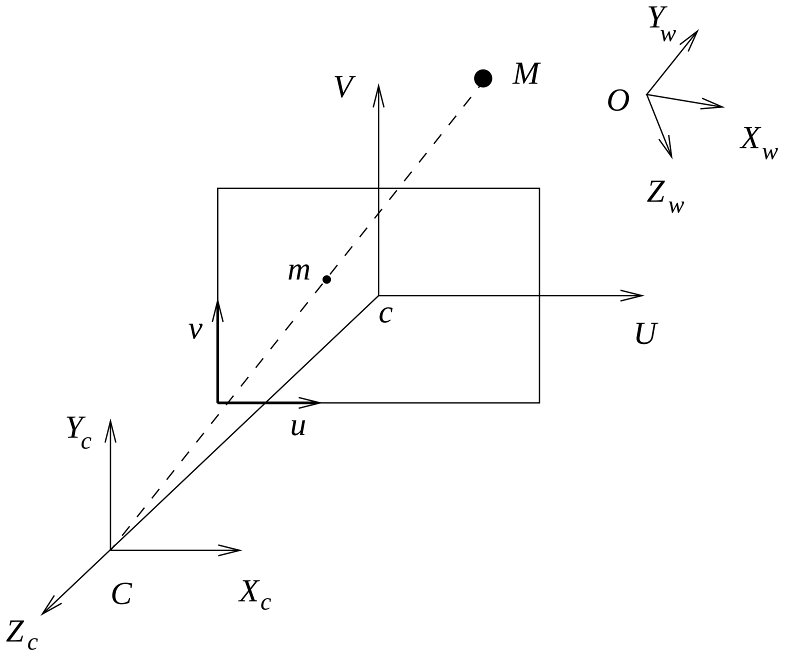



**Figura 2.1:** El modelo de proyección en perspectiva *pinhole*

Sea un punto $M$ en la escena y su proyección en el plano imagen $m$, en coordenadas del mundo y normalizadas, respectivamente:

$$M = [x, y, z]^T \quad \text{y} \quad m = [u, v]^T$$

Denotamos un punto en coordenadas homogéneas como

$$\tilde{M} = [x, y, z, 1]^T \quad \text{y} \quad \tilde{m} = [u, v, 1]^T$$

La calibración consiste en obtener la relación entre un punto del mundo, $M$, y su proyección normalizada en el plano imagen, $m$. Esta relación se puede expresar en términos de geometría proyectiva mediante la ecuación

$$s\tilde{m} = P\tilde{M} \tag{2.1}$$

donde $s$ es un factor de escala arbitrario y $P$ es una matriz 3×4 llamada *matriz de proyección* (más detalles en la sección A.1.1). La matriz de proyección se puede descomponer en los parámetros intrínsecos y los parámetros extrínsecos de la calibración, como se puede ver con detalle en la sección A.1.2.

### 2.1.3 Geometría epipolar

La geometría epipolar es la relación entre las proyecciones de una misma escena en dos planos imagen de dos cámaras. Dadas las matrices de ambas proyecciones, $P$ y $P'$, un punto $M$ del espacio tridimensional es proyectado en dos puntos diferentes, $m$ y $m'$, en la primera y la segunda imagen, respectivamente,

$$s\tilde{m} = P\tilde{M} \tag{2.2}$$
$$s'\tilde{m}' = P'\tilde{M} \tag{2.3}$$

de modo que todos los puntos de la línea recta entre $M$ y el centro óptico de la primera cámara, $C$, se proyectan en dicha cámara sobre el mismo punto $m$ y en la segunda cámara sobre una línea recta llamada la *línea epipolar* de $m$, $l'_m$, tal como se puede observar en la figura 2.2. Análogamente, todos los puntos de la línea recta entre $M$ y $C'$ se proyectan en la segunda cámara sobre el mismo punto $m'$ y en la primera cámara sobre una línea recta llamada la *línea epipolar* de $m'$, $l_{m'}$.

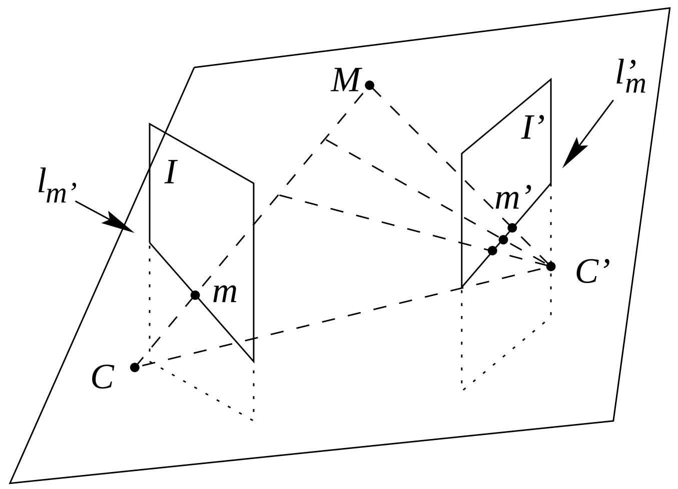



**Figura 2.2:** Geometría epipolar.

La relación entre un punto de la imagen y su línea epipolar es proyectiva lineal y puede escribirse como,

$$F\tilde{m} = l'_m \tag{2.4}$$
$$F^T\tilde{m}' = l_{m'} \tag{2.5}$$

donde $F$ es una matriz $3\times3$ llamada la *matriz fundamental* (para más detalles véase la sección A.2).

Conociendo los pares de puntos $m$, $m'$ que corresponden a un punto del mundo $M$ es posible hallar la localización de $M$ mediante una sencilla técnica de triangulación (figura 2.2). Sólo es posible recuperar la distancia del punto $M$ a la cámara (profundidad) si se tienen dos o más vistas del mismo. Así pues, para poder recuperar la estructura tridimensional de la escena, es crucial el proceso de identificación de los elementos homólogos en las diferentes imágenes.

## 2.2 El problema de la correspondencia

### 2.2.1 Introducción

Si un objeto es observado desde dos puntos de vista diferentes la posición de las proyecciones en cada imagen de un mismo punto de la escena 3D serán diferentes. Es posible definir el problema de la correspondencia como el de la búsqueda de una función $d : (u, v) \to (d_u, d_v)$ tal que dado el punto $m = (u, v)$ de la imagen de referencia, el punto $m' = m + d(m) = (u + d_u, v + d_v)$ representa la proyección del mismo punto del espacio 3D sobre la segunda imagen. La función $d$ se denomina *disparidad* y ha de ser conforme a la geometría epipolar.

En el caso particular de la geometría paralela (sección A.2.1), $d_v = 0$ y por tanto la función disparidad es un escalar de modo que dado el punto $m$, $m' = m + d(m) = (u + d, v)$. Esta es la formulación más común, ya que solamente requiere rectificar las imágenes, y permite obtener un mapa de disparidades cuya relación con la profundidad de los puntos de la imagen es sumamente sencilla: la función disparidad es inversamente proporcional a la profundidad del punto 3D proyectado,

$$d = \frac{bf}{z_c} \tag{2.6}$$

donde $f$ es la longitud focal, $b$ es la distancia entre las cámaras y $z_c$ es la profundidad del punto 3D en coordenadas de la cámara.

Así como el principio de la triangulación es muy simple, el problema de la correspondencia es de difícil solución y constituye el problema clave de la estereovisión. Este problema sería prácticamente irresoluble si no se dispusiera de información adicional que restrinja el espacio de soluciones posibles, ya que a priori cualquier elemento de una imagen podría ser el homólogo a cualquier elemento de la otra imagen. Sin embargo, es posible aplicar una serie de restricciones basándose en el modelo geométrico de las cámaras y en el modelo fotométrico de los objetos de la escena. La mayoría de los algoritmos de estereovisión se basan generalmente en las mismas restricciones expresadas mediante diferentes formalismos.

En general, los algoritmos de estereovisión utilizan como entrada las primitivas extraidas de la imagen y devuelven como resultado un mapa de disparidades asociado a estas primitivas, aunque a veces devuelven directamente su posición tridimensional. Se ha desarrollado un gran número de técnicas que se distinguen por la elección de las primitivas, por las restricciones que les son asociadas y por la estrategia de la correspondencia. Una vez identificadas las restricciones que permiten reducir el espacio de búsqueda, éstas se integran en un algoritmo de búsqueda que obtenga el conjunto de correspondencias final. Este algoritmo también puede interpretarse como un *método de optimización* que se compone de:

- un *criterio* que permita seleccionar y distinguir entre conjuntos de correspondencias correctos e incorrectos, que constituirá la función a optimizar, y

- un *algoritmo* que explore el espacio de búsqueda.

En esta sección examinaremos primero cuáles son estas restricciones y a continuación revisaremos las principales técnicas desarrolladas hasta el momento para resolver el problema de la correspondencia.

### 2.2.2 Restricciones de un sistema de visión estereoscópico

Jones [1997] clasifica las restricciones aplicables al problema de la correspondencia en dos tipos. Las *restricciones unarias* son las que se derivan de la geometría del sistema y de la similitud entre los elementos entre los que se intenta establecer la correspondencia. Estas restricciones contienen información específica a cada posible correspondencia. Las *restricciones binarias* son las que miden la compatibilidad entre pares de posibles correspondencias.

El conjunto de correspondencias potenciales es el producto cartesiano de los conjuntos de primitivas extraidas de cada imagen. Este conjunto puede ser enorme y por tanto, costoso de procesar. Las restricciones unarias permiten extraer un conjunto de posibles correspondencias más reducido, basado en la localización de los elementos y en su similitud. Las restricciones binarias sirven para proporcionar criterios de selección robustos en los casos en los que hay conflictos entre correspondencias.

#### Restricción epipolar

El punto homólogo a uno dado está restringido a pertenecer a una línea de la imagen determinada por la geometría del sistema estereoscópico [Ayache, 1991], que depende directamente del modelo geométrico de la cámara. Es la restricción geométrica más importante debido a que supone una reducción en el espacio de búsqueda del homólogo a un elemento dado. La búsqueda de una correspondencia pasa de ser bidimensional (un plano) a unidimensional (una recta) gracias a la relación geométrica que existe entre un punto del espacio 3D y sus dos proyecciones, tal como se detalla en la sección A.2 sobre la geometría epipolar.

#### Restricción de disparidad (o de profundidad)

La restricción sobre la disparidad surge de la restricción sobre las profundidades a las que se pueden encontrar los objetos de la escena. Generalmente se puede establecer un intervalo de disparidades $\mathcal{D} = [d_{min}, d_{max}]$ que depende de la configuración de las cámaras respecto a la escena y que está directamente relacionado con el intervalo de profundidades $[Z_{min}, Z_{max}]$ de los puntos de la escena. Por lo tanto, el punto homólogo a uno dado está restringido a pertenecer a un segmento de la recta epipolar correspondiente, lo que supone una mejora sobre la restricción epipolar (figura 2.3).

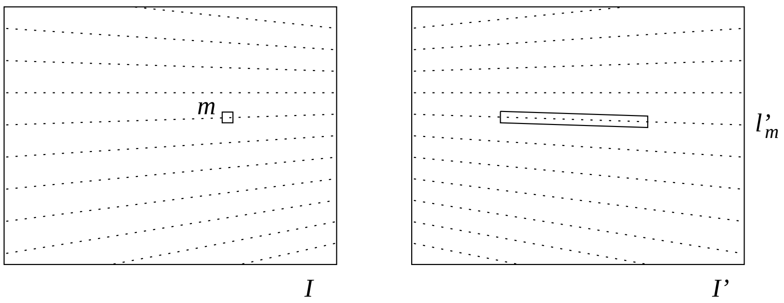



**Figura 2.3:** Restricción epipolar y restricción de disparidad: las líneas de puntos representan los pares de líneas epipolares y la zona encuadrada representa la zona de búsqueda del punto correspondiente al pixel $m$.

### Restricción de similitud

Es una restricción unaria que se basa en las propiedades de las primitivas empleadas en la correspondencia y que se combina con las dos restricciones geométricas anteriores en la búsqueda de los posibles homólogos de un elemento dado.

Generalmente, se basa en las propiedades fotométricas de los elementos que componen la escena, de modo que sólo aquellos candidatos que tengan intensidades similares a un elemento dado pueden ser elementos homólogos a él. El criterio de *similitud fotométrica* supone una hipótesis lambertiana sobre los objetos de la escena, es decir, que las propiedades fotométricas y de reflectancia de un objeto se conservan en su proyección, mostrándose iguales en dos vistas diferentes.

En el caso de primitivas complejas (segmentos curvos, esquinas, regiones, etc.) suele considerarse en la similitud las propiedades geométricas de las primitivas. Este es el caso del *límite en el gradiente de la disparidad* (o DGL), utilizado como restricción unaria[1] en el caso de primitivas con orientación, como son por ejemplo los segmentos lineales [Pollard *et al.*, 1985]. Generalmente, los atributos de una determinada primitiva se agrupan en un *vector de características* y a cada atributo se le asocia un peso que indica su importancia en el cálculo de la similitud entre dos elementos. De este modo se pueden considerar distintas propiedades para determinar cuáles de las correspondencias potenciales son posibles.

### Restricción de unicidad

Es la primera y más importante de las restricciones binarias: para cada punto de una imagen existe como máximo un punto homólogo en la otra imagen [Marr y Poggio, 1977] [Marr y Poggio, 1979]. Esta restricción se cumple siempre y cuando no haya objetos

[1] Originalmente se utilizó como una restricción de continuidad binaria [Burt y Julesz, 1980].

transparentes en la escena. Esta restricción obliga a seleccionar sólo una de las posibles correspondencias de cada elemento de la imagen. Si después de aplicar las restricciones anteriores todavía hay más de un candidato a la correspondencia, es necesario utilizar las restricciones binarias para realizar esta selección.

De la restricción de unicidad se deduce que dadas dos imágenes estereoscópicas, las correspondencias calculadas para los elementos de una imagen, y las correspondencias calculadas para los elementos de la otra imagen deben proporcionar los mismos pares de elementos homólogos. Esta formulación se denomina *restricción de consistencia izquierda-derecha* y se emplea en muchos métodos para validar el resultado de la correspondencia [Fua, 1991] [Fusiello *et al.*, 1997b].

### Restricción de continuidad

Basándose en las propiedades de continuidad y de cohesión de la materia, los puntos adyacentes en el espacio 3D permanecen adyacentes en cada proyección [Marr y Poggio, 1977, Grimson, 1981]. De ahí se deduce que la disparidad (o profundidad) de los elementos de la escena debe variar de manera suave a lo largo de la imagen. Aplicando esta restricción se puede establecer la compatibilidad o incompatibilidad entre dos posibles correspondencias dadas.

Existen numerosas formas de formular esta restricción en la práctica, dependiendo de las primitivas de la correspondencia. Por ejemplo, dos puntos adyacentes y sus correspondencias deberían satisfacer el *límite en el gradiente de la disparidad* [Burt y Julesz, 1980] [Pollard *et al.*, 1985] [Prazdny, 1985] excepto en el caso de que alguno de ellos sea un borde.

La *continuidad de la figura* (*figural continuity*) [Mayhew y Frisby, 1981] [Grimson, 1985] es otra formulación de esta restricción que permite evitar los problemas de discontinuidad en los límites de las superficies imponiendo la continuidad de la disparidad únicamente a lo largo de los contornos de las figuras de la imagen, y no a través de ellos. Por ejemplo, dos pares de puntos que pertenezcan al mismo contorno en ambas imágenes son compatibles. Otro ejemplo con primitivas lineales: dos segmentos de línea que están conectados en ambas imágenes están necesariamente conectados en el mundo 3D y por tanto son compatibles [Horaud y Skordas, 1989].

Los puntos donde se producen discontinuidades en la disparidad se denominan *contornos de oclusión* y pueden deberse a dos causas:

- La existencia de un cambio brusco de profundidad.
- La no existencia de correspondencia debido a una oclusión.

Esto significa que la función de reconstrucción, que asigna a cada punto de la imagen su profundidad, es continua a tramos.

### Restricción de conservación del orden

Dados dos elementos situados a lo largo de una línea epipolar, sus homólogos se encuentran a lo largo de la línea epipolar correspondiente, de modo que el orden de los elementos en la primera imagen coincide con el orden de los elementos correspondientes en la otra imagen. Esta restricción no se cumple siempre, pero geométricamente es poco probable que dos puntos se proyecten produciendo un ordenamiento diferente en las imágenes izquierda y derecha. Si la escena está formada por superficies de pequeña extensión con diferencias grandes en la profundidad respecto a otros objetos de la escena se puede violar esta condición (por ejemplo, la inversión de las patas de una silla, como se muestra en la figura 2.4). Sin embargo, esta restricción se utiliza en la mayor parte de las técnicas de correspondencia, siendo especialmente útil cuando en la escena pueden aparecer patrones repetitivos.

**Figura 2.4:** Un ejemplo de violación de la restricción de orden: inversión del orden de las patas de una silla.

### Restricciones topológicas

Además de la restricción de conservación del orden, existen otras restricciones topológicas que intentan servirse del hecho de que la estructura 3D vista en ambas imágenes es idéntica y que esto restringe su proyección desde distintos puntos de vista.

Algunas de estas restricciones tienen que ver con el paralelismo: líneas 3D que son casi paralelas en el espacio se proyectan en líneas 2D casi paralelas en cada cámara. En el caso de los segmentos lineales se traduce en *restricciones de colinearidad* para determinar la compatibilidad entre correspondencias.

Zhang y Faugeras [1992] formulan las *restricciones de rigidez* para el caso de la correspondencia de segmentos lineales, que consisten en cuatro restricciones acerca de la longitud, la distancia, y los ángulos de los segmentos.

Otras restricciones se fijan en que la posición relativa entre las primitivas sea similar de una imagen a otra. Esta restricción es muy efectiva, pero es menos cierta en el caso de vistas estereoscópicas muy separadas, donde las oclusiones pueden producir violaciones locales de la misma. Por esta razón, se tiende a implementar restricciones más débiles como es el caso de las relaciones entre segmentos lineales *left-of*, *right-of* del artículo de Horaud y Skordas [1989], que también incluye las relaciones *collinear-with* y *same-junction-as*. Estas relaciones proporcionan conjuntamente una descripción estructural de la escena que permite extraer relaciones de compatibilidad entre pares de correspondencias.

#### Restricciones jerárquicas

Los métodos de correspondencia jerárquicos proporcionan una alternativa eficiente a los métodos de búsqueda exhaustiva. En general se establece una jerarquía de primitivas y se obtiene la correspondencia en cada nivel de la jerarquía desde la primitiva más compleja (nivel más alto) a la más simple (nivel más bajo) consecutivamente. La principal ventaja es el ahorro computacional que se produce al reducir el espacio de búsqueda en las correspondencias de nivel más alto. La rigidez de la jerarquía hace que la correspondencia en los niveles más bajos sea dependiente de que la jerarquía se haya establecido correctamente. La desventaja radica en que las primitivas de mayor nivel son generalmente fáciles de definir (por ejemplo, *superficie* u *objeto*) pero suelen ser muy difíciles de extraer de manera robusta.

En general, cuanto más compleja es una primitiva, más descriptivos son sus atributos y hay menos ocurrencias de la misma en la imagen. Las primitivas complejas son más susceptibles de producir las correspondencias correctas mediante la comparación de sus atributos y, gracias a las relaciones jerárquicas, las correspondencias de alto nivel proporcionan restricciones considerables sobre las correspondencias posibles en las primitivas de niveles inferiores. Este aspecto es deseable para reducir el espacio de búsqueda considerablemente. Estas características de los sistemas jerárquicos han sido utilizadas en numerosos trabajos [Marapane y Trivedi, 1989] [Sander *et al.*, 1989] [Marapane y Trivedi, 1992] para obtener correspondencias más fiables.

### 2.2.3 Clasificación de las técnicas de correspondencia estéreo

El problema de la correspondencia no es un área exclusiva de la visión estereoscópica puesto que también aparece en otras áreas de la visión por computador como son el reconocimiento de objetos, el análisis de secuencias de imágenes y el análisis de imágenes tridimensionales. Un resumen del estado del arte del problema de la correspondencia en todas estas áreas se puede encontrar en Zhang [1993].

En las revisiones de las técnicas de visión estereoscópica [Barnard y Fishler, 1982] [Dhond y Aggarwal, 1989] [Faugeras *et al.*, 1992] [Lane y Thacker, 1994] los algoritmos de búsqueda de la correspondencia se clasifican en dos grandes familias:

- **Métodos basados en áreas**: también llamados métodos basados en superficies. Explotan la resolución radiométrica [Sonka *et al.*, 1993] de los pixeles por medio de áreas o ventanas sobre la imagen. Obtienen resultados muy buenos sobre las imágenes que presentan una textura importante. Permiten crear un mapa denso de disparidades siendo posible incluso obtener una precisión subpixel. Como desventaja está el hecho de que asumen una superficie continua y por tanto presentan problemas en presencia de discontinuidades de las superficies tridimensionales.

- **Métodos basados en características**: también llamados métodos basados en primitivas. Explotan las primitivas de alto nivel obtenidas de la imagen, que consisten en elementos con características distintivas como pueden ser los puntos de contorno, los segmentos lineales de los contornos, los segmentos curvos de los contornos, las regiones, etc. De estas primitivas se seleccionan las características distintivas que pueden ser atributos de posición, orientación, curvatura, etc. que sirvan para establecer la correspondencia. Proporcionan una información más dispersa aunque también más robusta cuanto más significativa es la primitiva.

  Se puede encontrar una revisión reciente de los métodos que emplean primitivas complejas en el artículo de Jones [1997].

Los métodos de correspondencia que se presentan en los capítulos 3 y 4 se clasifican a priori en la categoría de métodos basados en características, debido a que están basados en regiones resultantes de la segmentación de una o varias imágenes, que son primitivas de alto nivel. Sin embargo, estos métodos explotan la radiometría de los pixeles de la región de forma similar a los métodos basados en áreas, en lugar de considerar características típicas de las regiones como son el área, el perímetro, los momentos de inercia, la elongación, etc.

De esta manera se intentan aunar las dos ventajas principales de ambas familias de métodos: conseguir correspondencias robustas gracias al uso de primitivas de alto nivel de significado, evitando los problemas en presencia de discontinuidades, y conseguir un mapa denso de profundidades que proporcione un mapa tridimensional aproximado de la escena real que sea de utilidad para aplicaciones de robótica.

Además de estas dos grandes familias, las técnicas de correspondencia estereoscópica se pueden clasificar según la estrategia utilizada. Aunque no se pretende hacer un estudio exhaustivo del estado del arte, a continuación se estudian los aspectos fundamentales de las estrategias más conocidas, profundizando más en aquéllas que están relacionadas de una u otra manera con el trabajo que se expone en los siguientes capítulos.

### 2.2.4 Métodos basados en correlación

Estos métodos consideran cada imagen como una señal bidimensional en la que se busca para cada punto la traslación que minimiza un cierto criterio. El método más clásico de esta familia es la optimización de un criterio de correlación.

Las técnicas de estereovisión basadas en correlación son de los métodos más antiguos [Gennery, 1980] pero siguen en vigencia, teniendo numerosas aplicaciones en robótica [Faugeras *et al.*, 1993] [Zhang, 1996a]. Son métodos basados en áreas, donde para cada pixel de la imagen se calcula la correlación entre una ventana de tamaño fijo centrada en ese pixel y una ventana del mismo tamaño centrada en cada pixel de la otra imagen susceptible de corresponderle. Generalmente las imágenes se encuentran rectificadas para que el uso de ventanas de correlación sea más sencillo. Se elige como correspondiente aquél punto que maximice la medida de correlación y de él se deduce la disparidad. Para calcular la disparidad con precisión subpixel se puede aproximar los valores de correlación en el vecindario del máximo a una curva de segundo grado y calcular la disparidad óptima por interpolación.

Algunos algoritmos calculan puntos de interés en los que realizar la correlación para reducir el tiempo de computación. Los puntos de interés pueden ser pasos por cero de la segunda derivada de la intensidad [Moravec, 1977] [Castan y Shen, 1984], esto es, pixeles pertenecientes a los contornos (*edge pixels* o *edgels*), u otras primitivas de mayor nivel como esquinas (*corners*) [Zhang *et al.*, 1995]. En estos casos, es necesaria una etapa de interpolación entre las disparidades calculadas para obtener un mapa denso de disparidades.

Sin embargo, con las arquitecturas hardware modernas es posible realizar la correlación sobre todos los puntos de la imagen y retener sólo aquellas correspondencias que parezcan "válidas". El problema reside entonces en proporcionar una definición efectiva de la validez de una correspondencia y en añadir una etapa de postproceso donde se pueda asignar un valor de disparidad a los pixeles sin correspondencia. Por ejemplo, Fua [1991] utiliza como criterio de validez la restricción de consistencia izquierda-derecha y una interpolación entre las disparidades "válidas" para calcular las disparidades no encontradas.

La principal ventaja de estos métodos sobre los métodos basados en primitivas es la obtención de una representación densa de la escena observada. Otra ventaja de las técnicas de correlación es que resultan fáciles de paralelizar. Se pueden encontrar algunas implementaciones paralelas en los trabajos de Fua [1993] y Devernay [1996].

Sin embargo, para que los vecindarios de los puntos correspondientes tengan unas funciones de intensidad similares, es necesario, en teoría, que las superficies alrededor de los puntos cumplan las restricciones fotométricas y geométricas siguientes:

- restricción lambertiana: las superficies deben ser lambertianas en cada punto, es decir, que la intensidad de la proyección en cada imagen de un punto 3D debe ser independiente del punto de vista (no debe haber reflexiones especulares).

- restricción fronto-paralela: los métodos de correlación asumen implícitamente que la disparidad es constante localmente (dentro de la ventana de correlación). Esto es equivalente a decir que las superficies deben ser paralelas a los planos imagen de ambas cámaras.

- restricción de continuidad: la mayor parte de los métodos se basan en la hipótesis de que las superficies son al menos localmente continuas. Por ejemplo, estos métodos no son capaces de encontrar la correspondencia y reconstruir satisfactoriamente objetos de textura tridimensional compleja como son los matorrales o los árboles.

En la práctica, estas restricciones son laxas: la restricción lambertiana puede convertirse en “localmente lambertiana” si se normalizan localmente las intensidades, o si se filtran las imágenes previamente con un filtro laplaciana de la gaussiana [Nishihara, 1984a] de modo que se busca la correspondencia entre las variaciones locales de la intensidad en lugar de la propia intensidad. Una mejor solución sería incluir una función de reflectancia [Oren y Nayar, 1994] dentro de la función de correlación, pero este método tiene la desventaja de que se necesitaría conocer la posición de las fuentes luminosas.

Asímismo, la restricción fronto-paralela no es necesario que se verifique estrictamente, de modo que se pueden encontrar las correspondencias incluso en las superficies ligeramente inclinadas respecto a los planos imagen de las cámaras. Algunos métodos permiten evitar esta restricción teniendo en cuenta la deformación local de la superficie como la que presenta Devernay [Devernay y Faugeras, 1994] [Devernay, 1996] que consiste en hacer dos cálculos de la disparidad: una primera etapa mediante el método clásico de correlación y una segunda etapa denominada *correlación fina* que estudia la normal y la curvatura locales de la superficie, representadas respectivamente por las propiedades diferenciales de primer y segundo orden de la disparidad, para refinar y completar los resultados obtenidos en la primera etapa.

Un método que evita totalmente la restricción fronto-paralela es el de Super y Klarquist [1997] que, para cada punto de interés de la imagen, realiza una búsqueda en el espacio de superficies planas 3D. Cada superfice 3D hipotética se usa para definir un par de trozos de imagen con posiciones, formas, tamaños, orientaciones e intensidades consistentes con la hipótesis y entre sí. De este modo, se asegura que la comparación entre los dos trozos da un índice de correlación fiable de la similitud entre las intensidades. Dado que la deformación entre imágenes se tiene en cuenta explícitamente, no sufre la restricción fronto-paralela y permite la existencia de deformaciones grandes entre imágenes. Otra característica de este método es que obtiene directamente la posición tridimensional de los puntos de interés. Sin embargo, no se modelan las oclusiones ni la existencia de discontinuidades.

La tercera restricción hace que se produzcan problemas en presencia de discontinuidades de las superficies tridimensionales. Para el tratamiento de las discontinuidades en la profundidad se han desarrollado algoritmos que se basan en la selección adaptativa de la ventana de correlación, en la búsqueda de óptimos globales mediante técnicas de programación dinámica, en aproximaciones bayesianas y en el uso de más de dos cámaras.

Las investigaciones recientes se centran esencialmente sobre estos problemas que son el tratamiento de las oclusiones, de las discontinuidades y de las superficies no fronto-paralelas.

El uso de más de dos cámaras permite reducir el número de ambigüedades e incrementar la habilidad de la detección de oclusiones, al disponer de múltiples proyecciones de la escena. Muchos trabajos emplean sistemas trinoculares para hallar la correspondencia utilizando técnicas de predicción y verificación [Ayache y Lustman, 1987] [Kitamura y Yachida, 1990] [Robert y Faugeras, 1991], que generalmente se basan en primitivas cuyo nivel simbólico es mayor que el de un simple pixel (segmentos lineales, segmentos curvos).

Otros trabajos emplean más de tres cámaras, como Okutomi y Kanade [1993], y Sakamoto *et al.* [1997] que emplean un conjunto de cámaras en línea a distintas distancias para reducir el número de ambigüedades en la optimización de un criterio de correlación. Satoh y Ohta [1994] [1995] [1996] [Satoh *et al.*, 1996] emplean una matriz de $3 \times 3$ cámaras para aumentar la habilidad en la detección de oclusiones y en consecuencia, la precisión de la correspondencia.

No obstante, nosotros nos centraremos en los sistemas binoculares, y por tanto, estudiaremos las demás técnicas después de ver los principales criterios de correlación.

### Criterio de correlación

Para obtener buenos resultados es necesario elegir un criterio de correlación que permita obtener buenas correspondencias en cualquier situación. La mayor parte de los criterios de correlación derivan de dos medidas fundamentales:

- la distancia euclídea entre los dos vectores formados por las intensidades de las imágenes, conocida como SSD (*Sum of Squared Differences*), y
- el producto escalar de estos dos vectores, conocida como CC (*Cross-Correlation*).

Para comparar los distintos criterios, vamos a denotar la ventana de correlación centrada en un pixel cualquiera $m$ como $W(m)$, y los pixeles del vecindario de $m$ como $m_i$, $i = [1..N]$, siendo $N$ el número de pixeles de la ventana. Asímismo, para cada disparidad $d$, el pixel correspondiente a $m = (u, v)$ es $m' = (u + d, v)$ y los pixeles que pertenecen a la ventana alrededor de $m'$ los denotaremos como $m'_i \in W(m')$, siguiendo el mismo orden adoptado en $m_i \in W(m)$. Esta notación nos permitirá extender fácilmente la definición de los criterios de correlación a regiones arbitrarias como se verá en los capítulos 3 y 4 de esta tesis.

- *El criterio SSD* (*Sum of Squared Differences*) consiste en minimizar la suma de las diferencias de las intensidades sobre el conjunto de la ventana de correlación, o lo que es igual, maximizar la función

  $$C_{SSD}(m, m') = -\frac{1}{N} \sum_{\forall m_i \in W(m)} (I_1(m_i) - I_2(m'_i))^2 \qquad (2.7)$$

  donde $I_1$ e $I_2$ son las intensidades de cada una de las imágenes. Cuando la ventana es de tamaño fijo se puede prescindir de la división por $N$.

  Este criterio es muy sensible a las diferencias de iluminación entre las dos imágenes.

- *El criterio ZSSD* (*Zero-mean Sum of Squared Differences*) consiste en minimizar la suma de las diferencias de intensidad respecto a la media de intensidad calculada sobre la ventana de correlación, o de manera equivalente, maximizar la ecuación

$$C_{ZSSD}(m,m') = -\frac{1}{N} \sum_{\forall m_i \in W(m)} ((I_1(m_i) - \overline{I_1(m)}) - (I_2(m'_i) - \overline{I_2(m')}))^2 \tag{2.8}$$

donde $\overline{I_1(m)}$ e $\overline{I_2(m')}$ son las medias de intensidad dentro de las ventanas $W(m)$ y $W(m')$, respectivamente. Este criterio permite calcular la correspondencia en superficies donde la iluminación es ligeramente diferente de modo que hay diferencias pequeñas de desplazamiento en la función de transferencia de las cámaras.

En la práctica, se considera que las funciones $\overline{I_1}$ e $\overline{I_2}$ varían poco sobre la ventana de correlación y se utiliza el siguiente criterio

$$\begin{aligned} C_{ZSSD}(m,m') &= -\frac{1}{N} \sum_{\forall m_i \in W(m)} ((I_1(m_i) - \overline{I_1(m_i)}) - (I_2(m'_i) - \overline{I_2(m'_i)}))^2 \\ &= -\frac{1}{N} \sum_{\forall m_i \in W(m)} (I'_1(m_i) - I'_2(m'_i))^2 \end{aligned} \tag{2.9}$$

que es equivalente a filtrar las imágenes con un filtro laplaciana de la gaussiana [Nishihara, 1984a] para obtener las imágenes $I'_1$ e $I'_2$ y aplicar el criterio SSD a continuación.

- *El criterio ZNSSD* (*Zero-mean Normalized Sum of Squared Differences*) consiste en minimizar la suma de las diferencias de intensidad respecto a la media de intensidad calculada sobre la ventana de correlación, normalizada por la varianza local de las intensidades. Esto equivale a maximizar la ecuación

$$C_{ZNSSD}(m,m') = -\frac{\sum_{\forall m_i \in W(m)} ((I_1(m_i) - \overline{I_1(m)}) - (I_2(m'_i) - \overline{I_2(m')}))^2}{N\sigma_1(m)\sigma_2(m')} \tag{2.10}$$

donde $\sigma_1(m)$ y $\sigma_2(m')$ son las desviaciones típicas de la intensidad dentro de las ventanas $W(m)$ y $W(m')$, respectivamente. Este criterio permite calcular la correspondencia en superficies donde la iluminación es diferente tanto en ganancia como en desplazamiento y, por tanto, es insensible a las diferencias en la función de transferencia de las cámaras.

Se puede efectuar una simplificación similar a la anterior, dejando parte de los cálculos al filtrado de las imágenes. Además, en el caso de que la ventana de correlación sea fija se puede simplificar no sólo el número de pixeles $N$, sino también la variación de la intensidad en la ventana de referencia, $\sigma_1(m)$.

- *El criterio CC* (*Cross-Correlation*) consiste en maximizar el producto escalar de dos vectores de intensidad,

$$C_{CC}(m,m') = \frac{1}{N} \sum_{\forall m_i \in W(m)} I_1(m_i) I_2(m'_i) \tag{2.11}$$

En la práctica, este criterio es inutilizable porque favorece las zonas de intensidad elevada que no necesariamente son similares a la ventana de correlación de la imagen de referencia.

- *El criterio ZNCC* (*Zero-mean Normalized Cross-Correlation*) consiste en maximizar el coseno de dos vectores de intensidad,

$$C_{ZNCC}(m,m') = \frac{\sum_{\forall m_i \in W(m)}((I_1(m_i) - \overline{I_1(m)})(I_2(m'_i) - \overline{I_2(m')}))}{N\sigma_1(m)\sigma_2(m')} \tag{2.12}$$

de modo que si el resultado es 1 ambos vectores son idénticos mientras que si es -1 son totalmente opuestos.

Se puede aplicar la misma simplificación que en el criterio ZNSSD. En el caso de ventanas fijas, también se puede simplificar $N$ y $\sigma_1(m)$.

Como se ha anotado anteriormente, los criterios que dan mejores resultados son los que comparan las intensidades filtradas por la media, es decir, ZSSD, ZNSSD y ZNCC, ya que permiten relajar la restricción lambertiana implícita en las técnicas de correlación.

#### Selección adaptativa de la ventana de correlación

Nishihara [1984b] demostró que la probabilidad de obtener correspondencias falsas se reduce a medida que crecen el tamaño de la ventana de correlación y la cantidad de textura. Sin embargo, a medida que crece el tamaño de la ventana se pierde precisión hasta el punto de poder perder características importantes de la imagen. El tamaño de la ventana debe ser suficientemente grande para incluir suficiente variación de intensidad para la correspondencia y reducir la sensibilidad al ruido, pero suficientemente pequeña para evitar los efectos de la distorsión proyectiva.

De este razonamiento surgieron los algoritmos de correspondencia con *ventana adaptativa*. Kanade y Okutomi [1990] [1994] [Okutomi y Kanade, 1992] proponen seleccionar la ventana adaptativamente mediante la evaluación de la variación local de la intensidad y la disparidad. Emplean un modelo estadístico que representa la incertidumbre de la disparidad de los puntos de la ventana, que presumiblemente se incrementará a medida que se incrementa la distancia de los puntos al punto central. De este modo pueden calcular la incertidumbre de una estimación de la disparidad, teniendo en cuenta tanto la varianza de la intensidad como la de la disparidad. El algoritmo busca una ventana que produzca la estimación de la disparidad con la menor incertidumbre para cada pixel de la imagen, de modo que no sólo se controla el tamaño sino también la forma rectangular de la ventana.

Se han desarrollado numerosas versiones de la técnica de correspondencia con ventana adaptativa, como es la de Menard y Kropatsch [1997] que asocian una máscara gaussiana a la ventana de correlación, de modo que la influencia de los pixeles del vecindario son ponderados en función de su distancia al pixel central. Esto hace que la forma de la ventana sea circular en vez de rectangular, al mismo tiempo que permite controlar el radio de influencia mediante un parámetro que se puede adaptar a cada caso individual. Otro aspecto interesante de este método es que se obtiene la disparidad de máxima correlación a diferentes escalas, de modo que si en la escala de menor resolución no hay un único máximo, se asume que no hay punto correspondiente, y en caso contrario, se tiene en cuenta la disparidad anterior en sucesivas iteraciones hasta que se obtiene un máximo global a lo largo de la escala.

Otra versión de este tipo de algoritmos es la de Lotti [1996] que utiliza 4 ventanas adaptativas (arriba-izquierda, arriba-derecha, abajo-izquierda, abajo-derecha) en lugar de una para cubrir un máximo de formas geométricas alrededor del pixel central. El tamaño de cada ventana viene restringido por los contornos obtenidos mediante el operador de Canny-Deriche [Deriche, 1987] sobre las dos imágenes, y el resultado de cada una se valida por criterios de textura, de ruido y un umbral en el índice de correlación. La disparidad asociada al pixel central se calcula a partir del pico máximo de la curva de disparidades válidas. Finalmente, se realiza una segunda etapa de validaciones e interpolaciones sucesivas para completar el mapa denso de disparidades.

Otras versiones con múltiples ventanas intentan simplificar la técnica de la ventana adaptativa como son las que han desarrollado Geiger, Ladendorf y Yuille [1992] [1995], que utilizan dos ventanas fijas rectangulares a izquierda y derecha del pixel en estudio, una línea por debajo y por encima de la línea epipolar y de modo que ambas ventanas intersectan en una cierta área alrededor del pixel, y la de Intille y Bobick [1994a], que utilizan 9 ventanas diferentes, una centrada en el pixel central y el resto en las 8 direcciones alrededor del mismo. En ambos casos, para cada pixel y cada posible correspondencia se elige el mejor índice de correlación obtenido de las múltiples ventanas y se aplica una técnica de programación dinámica para encontrar las correspondencias. Más adelante profundizaremos también en este tipo de métodos que integran tanto técnicas basadas en áreas como basadas en primitivas (sección 2.2.5).

Fusiello *et al.* [1997b] [1997a] utilizan la misma técnica de múltiples ventanas que Intille y Bobick [1994a] pero utilizando una maximización del índice de correlación y posterior validación mediante la restricción de consistencia izquierda-derecha.

En resumen, los métodos basados en ventanas adaptativas intentan *evitar* los problemas generados por la presencia de oclusiones eliminando aquellas correspondencias cuya validez sea dudosa, determinando su incertidumbre mediante diversos métodos. Aunque los métodos basados en ventanas adaptativas consiguen obtener un conjunto de correspondencias válidas, no consiguen modelar las oclusiones explícitamente, y necesitan una etapa de postproceso en la que se completa el mapa de disparidades mediante alguna técnica de interpolación [Lotti, 1996].

Sin embargo, Geiger *et al.* [1992] muestran que las oclusiones pueden ayudar en el cálculo de la correspondencia. Según ellos, las oclusiones no son elementos a detectar y eliminar, porque así no se explota toda la información potencial que contienen. Es posible modelar las oclusiones y las discontinuidades de la profundidad explícitamente de modo que forman parte integrante del problema que se procede a resolver y por tanto, de la solución obtenida. De este modo las oclusiones y discontinuidades de la profundidad no son problemas a evitar, sino una fuente de información a tener en cuenta.

### 2.2.5 Métodos basados en programación dinámica

En las técnicas basadas en programación dinámica, la búsqueda de la solución se realiza mediante el algoritmo de Viterbi que consiste en partir el problema original en dos subproblemas, cada uno de los cuales se puede resolver de manera óptima y cuyos resultados se pueden procesar para alcanzar un óptimo global del problema original. La restricción de particionamiento que se aplica en el problema de la correspondencia estereoscópica es la restricción de orden y el criterio que selecciona el camino óptimo está basado en las diferencias fotométricas de los pixeles a emparejar. Estos métodos permiten obtener un conjunto de correspondencias para un conjunto de elementos, que es un óptimo global según una función de coste, en lugar de obtener un óptimo local para cada elemento considerado.

En general, en las técnicas basadas en programación dinámica, el espacio de búsqueda de la correspondencia se representa mediante un espacio bidimensional que incluye todas las posibles correspondencias para cada pixel de una línea epipolar. La solución al problema de la correspondencia se representa como un camino desde el primer punto de la línea epipolar al último, que minimiza una función de coste. En la función de coste se tienen en cuenta tanto los errores de similitud entre los puntos correspondientes como las oclusiones. De este modo se consigue obtener un conjunto denso de correspondencias que minimizan una función de coste global para cada línea epipolar.

Geiger *et al.* [1995] definen un espacio bidimensional llamado el *espacio de correspondencia* (*matching space*) cuyos ejes vienen dados por las líneas epipolares izquierda y derecha. Cada nodo del espacio de correspondencia contiene el error de similitud entre un elemento de la línea epipolar izquierda y uno de la derecha, calculado mediante una técnica de ventana múltiple que es una simplificación de la técnica de Kanade y Okutomi [1994]. La principal novedad consiste en que las oclusiones se modelan como saltos en el camino solución y se incorporan en la función de coste en función de la envergadura de este salto. Este modelo asume que una discontinuidad en la disparidad a lo largo de la línea epipolar siempre corresponde a una región oculta en la otra imagen. Así, se pueden incorporar las oclusiones en el proceso de correspondencia, imponiendo esta restricción sobre el conjunto de los posibles caminos en el espacio de correspondencias, llamada *restricción de oclusión*. En el espacio así definido se busca la solución óptima, utilizando además de la restricción de oclusión, la restricción de orden y la restricción de continuidad de la figura. Esta última supone una restricción débil de suavizado sobre el mapa de disparidades resultante. El coste computacional del algoritmo para cada línea epipolar es $O(N|\mathcal{D}|^2)$, donde $N$ es el tamaño de la línea epipolar y $|\mathcal{D}|$ el número de posibles disparidades.

Belhumeur [1993b] utiliza una aproximación bayesiana que relaciona las oclusiones y la correspondencia estéreo [Belhumeur y Mumford, 1992] [Belhumeur, 1993a] por medio de una función de densidad de probabilidad *a priori*, y estima las correspondencias mediante un algoritmo de máximo *a posteriori*. Para manejar las oclusiones y discontinuidades de la profundidad adecuadamente, Belhumeur afirma que debería mantenerse un mapa detallado de la geometría de la escena internamente, incluyendo no solamente la profundidad sino también la orientación de la superficie, las discontinuidades y los pliegues. Aunque el modelo es muy completo y proporciona una solución global al problema de la correspondencia, la búsqueda de la solución óptima requiere un esfuerzo computacional enorme, $O(N|\mathcal{D}|^2S^2)$, donde $S$ es el número de posibles valores de inclinación, para una dimensión.

Intille y Bobick [Intille y Bobick, 1994a] utilizan una representación del espacio de correspondencias similar a la de Geiger *et al.* , llamada *imagen del espacio de disparidades* (*disparity space image*, DSI), que consiste en un espacio bidimensional cuyos ejes vienen dados por una línea epipolar y todas las disparidades posibles. En el DSI, las discontinuidades de profundidad y las oclusiones se modelan de manera similar al espacio de correspondencia, considerando disparidades en lugar de posiciones absolutas en la línea epipolar correspondiente. La principal diferencia es que ellos proponen definir tres posibles estados para cada nodo, de modo que para calcular el estado de un nodo sólo se pueden considerar tres nodos vecinos y ciertas transiciones legales entre estados. Así se reduce el coste computacional del algoritmo a $O(3N|\mathcal{D}|)$.

Sin embargo, debido a que el coste de un nodo oclusión es una constante, el coste de una región oculta es linearmente proporcional a la anchura de la región y en consecuencia, pequeñas variaciones en el coste de cada oclusión pueden cambiar el camino mínimo global, particularmente en imágenes con ruido. Para intentar minimizar la sensibilidad del algoritmo al coste de oclusión, Intille y Bobick utilizan puntos de control verdaderos (*ground control points*) para guiar la correspondencia. Estos puntos de control pueden consistir en correspondencias conocidas de alta fiabilidad, generalmente obtenidas mediante alguna otra técnica de correspondencia. Introducir un punto de control en el DSI significa que todos los posibles caminos pasarán por él, de modo que el número de posibles caminos se reduce. Al mismo tiempo, se puede determinar una serie de nodos del DSI por los cuales no pasa ningún camino válido y por tanto se puede evitar el cálculo de esos nodos, con lo que el coste computacional también se reduce.

Otro modo de incorporar puntos de control en el DSI consiste en considerar no una sino varias correspondencias posibles para un punto dado. En esta línea, en el DSI se puede incorporar la información que proporcionan los bordes [Intille y Bobick, 1994b]. Trazando en el DSI la posición de un borde de la izquierda y la posición de su punto correspondiente en la derecha (que según la restricción de oclusión será también un borde) podemos observar que forman dos líneas que intersectan en la ordenada que corresponde a su disparidad. Por tanto, si trazamos las líneas correspondientes a todos los bordes de las líneas epipolares izquierda y derecha, todas las intersecciones que aparecen constituyen todas las correspondencias posibles entre bordes. Utilizar estos puntos de intersección como puntos de control permite asímismo reducir el espacio de búsqueda.

Algunos métodos que utilizan la programación dinámica como estrategia de correspondencia están basados en primitivas, como Bensrhair *et al.* [1996] que realizan la correspondencia entre *edgels* y después interpolan las disparidades entre los mismos, y Li [1994] que realiza la correspondencia de líneas en lugar de *edgels*.

En otros trabajos, los puntos de los bordes se usan para proporcionar información sobre la localización de las discontinuidades de la disparidad. Baker y Binford [1981] fueron los primeros en combinar la correspondencia basada en áreas y basada en primitivas en dos etapas. En primer lugar obtienen una correspondencia de puntos de bordes que aparecen en cada línea epipolar a diferentes resoluciones y a continuación obtienen los valores de disparidad en los intervalos obtenidos. Además, se intenta evitar el problema de las oclusiones considerando dos lados en cada borde (izquierdo y derecho) y emparejando bordes izquierdos con izquierdos y bordes derechos con derechos. Sin embargo, no se modelan explícitamente las oclusiones. Rojas *et al.* [1997] también proponen dos etapas, que se pueden aplicar conjunta o independientemente, e incorporan una función de visibilidad para manejar las oclusiones.

Ohta y Kanade [1985] utilizan intervalos limitados por contornos como primitivas para la correspondencia, utilizando la consistencia entre líneas adyacentes en la función de coste. Cox *et al.* [1992] [1996] intentan evitar el suavizado que producen las restricciones de continuidad aplicando una regularización basada en *restricciones de cohesión* (*cohesivity constraints*).

Recientemente, Birchfield y Tomasi [1998] proponen procesar cada línea independientemente y añadir un postproceso que consiste en propagar los valores de disparidad fiables en regiones de valores de disparidad no fiables. Este postproceso es efectivo en la propagación de las disparidades en las regiones que tienen poca variación en la intensidad. Sin embargo, aunque en el proceso se modelan las oclusiones, en el postproceso sólo se tiene en cuenta la fiabilidad de las disparidades para la propagación de la información.

### 2.2.6 Métodos basados en relajación

El problema de la correspondencia se puede traducir como la asignación de etiquetas únicas a un conjunto de primitivas de una imagen, dada una lista de posibles correspondencias de cada primitiva.

Este problema se puede resolver mediante un algoritmo iterativo y localmente paralelo donde, dadas las primitivas de la imagen izquierda y de la imagen derecha, se construye un conjunto inicial con todos los emparejamientos posibles. Este conjunto se organiza como una colección de nodos $\{a_i\}$ de modo que hay un nodo por cada primitiva de la primera imagen. A cada nodo $a_i$ se le asocia un vector de características geométricas de la primitiva de la primera imagen, y el conjunto de etiquetas, $e_i$, que representa las posibles correspondencias de las primitivas. Una etiqueta especial llamada el carácter nulo (*null label*) es incluido inicialmente en el conjunto $e_i$. Esta etiqueta significa considerar la posibilidad de que no haya correspondencia con la segunda imagen [Barnard y Thompson, 1980]. A cada etiqueta $e$ de $e_i$ se asocia un número $p_i(e)$ que se interpreta como una estimación de la probabilidad de que $e$ sea efectivamente la correspondencia de la primitiva, de modo que $p_i(e) \in [0,1]$ y $\sum_{e \in e_i} p_i(e) = 1$. Estas estimaciones de probabilidad se modifican teniendo en cuenta la coherencia entre los vecinos. Si hay un número de primitivas compatibles relativamente elevado entre el vecindario, entonces $p_i(e)$ aumentará, y en caso contrario disminuirá. Después de un número suficiente de iteraciones, un nodo es considerado como emparejable si hay una etiqueta con una probabilidad suficientemente elevada, y no emparejable en otro caso. Por tanto, lo esencial en este tipo de técnicas es la definición de la regla de actualización de las probabilidades.

En el área de correspondencia estereoscópica son famosos los algoritmos de Marr y Poggio [1976] [1979], Grimson [1981] [1985], y Pollard, Mayhew y Frisby [1985] que utilizan *edgels* como primitiva. En los trabajos más recientes, se tiende a utilizar la relajación probabilística con primitivas de mayor nivel simbólico como son los segmentos lineales [Christmas, 1995] [Christmas *et al.*, 1995] [Wilson y Hancock, 1997], y los segmentos curvos [Nasrabadi, 1992]. Algunos métodos utilizan redes neuronales como es el caso de Pajares *et al.* [1998], que aplican la relajación mediante la red neuronal de Hopfield.

### 2.2.7 Métodos basados en grafos

Tradicionalmente, el problema de la correspondencia basada en primitivas se ha formulado como la correspondencia entre dos estructuras relacionales. Dadas las primitivas de una imagen, éstas están conectadas por relaciones binarias e incluso ternarias como, por ejemplo, la distancia entre dos puntos, el ángulo entre dos rectas, un punto está sobre una recta, tres rectas intersectan en un punto, etc. Aunque las relaciones pueden ser $n$-arias, en la práctica se suele utilizar relaciones binarias ($n = 2$) y por tanto la estructura relacional se convierte en un grafo estándar. Dadas estas estructuras, la tarea del proceso de correspondencia consiste en identificar una subestructura "idéntica" dentro de las dos estructuras. Sin embargo, raramente se puede encontrar una subestructura idéntica porque las primitivas detectadas contienen ruido o están mal localizadas, o porque hay distorsiones en las imágenes, o porque un movimiento rígido en el espacio 3D generalmente induce a desplazamientos diferentes de las primitivas en la imagen. De ahí la introducción del concepto de *correspondencia inexacta* (*inexact matching*). Por tanto, el problema se traduce en la búsqueda del máximo par de subgrafos isomórficos tal que cada subgrafo sea una parte de los grafos relacionales iniciales.

La técnica habitual para resolver el problema de la correspondencia basándose en grafos consiste en construir un *grafo de asociación*. Cada nodo del grafo representa un posible emparejamiento entre una característica en la imagen izquierda y otra en la imagen derecha, y un arco entre dos nodos representa la compatibilidad entre los dos emparejamientos que enlaza.

Ayache y Faverjon [1987] usan segmentos lineales de la frontera como primitivas y representan todas las correspondencias posibles entre segmentos como nodos del grafo. Los arcos del grafo representan pares de segmentos compatibles, usando ciertas relaciones de vecindario. Exploran el grafo usando un algoritmo de predicción y propagación que usa algunas restricciones locales y globales (restricción epipolar, restricción de similitud geométrica, restricción de continuidad y restricción de unicidad). El método evita la búsqueda exhaustiva en el grafo, pero no hay garantía de que encuentre ni la correspondencia más numerosa ni la mejor.

Horaud y Skordas [1989] usan segmentos lineales de la frontera e introducen la búsqueda de *cliques máximos*[2] como estrategia de correspondencia estéreo. Nasrabadi y Liu[1989] aplican la misma estrategia a la correspondencia de segmentos curvos. Este método ha sido ampliamente utilizado en el área de reconocimiento de objetos en imágenes 2D [Bolles, 1979] [Bolles y Cain, 1982] y 3D [Yang *et al.*, 1989], donde la búsqueda del *clique máximo más grande*[3] obtiene el mejor conjunto de correspondencias entre regiones de la imagen y partes de un modelo del objeto. Sin embargo, el *clique* máximo más grande no es necesariamente la mejor correspondencia, debido a varias razones:

- la existencia de oclusiones, que hace que una característica correspondiente a una dada pueda aparecer distinta o incluso no aparecer;
- limitaciones en el proceso de extracción de características, que a menudo conlleva que dadas dos imágenes similares no proporcione los mismos resultados;
- alineación accidental, ruido, etc. que pueden producir diferencias en la extracción de características de una y otra imagen.

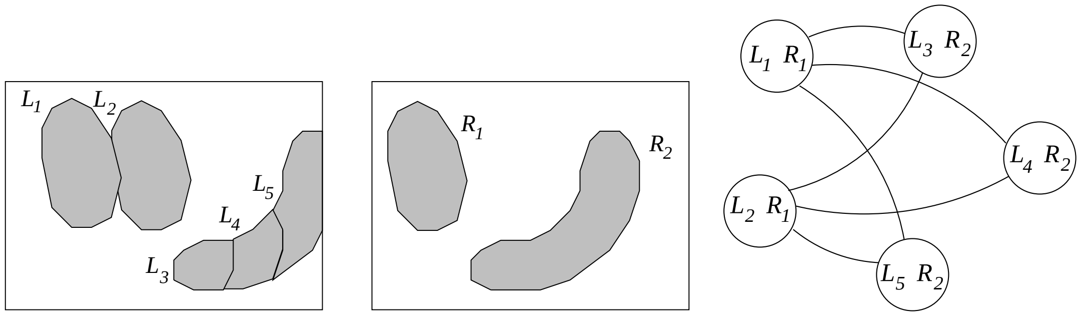
En la figura 2.5 se muestra un ejemplo sencillo de correspondencia basada en grafos, donde las primitivas son regiones. La figura ilustra un caso en el que hay seis *cliques* máximos del mismo tamaño (dos nodos cada uno) en el grafo de asociación resultante.

Para intentar obtener la mejor correspondencia estéreo, Horaud y Skordas asocian un peso a cada nodo del grafo que representa la similitud entre los dos segmentos de la frontera de ambas imágenes. Usan una función de beneficio para encontrar el *clique* máximo que maximiza la suma de pesos de sus nodos individuales, al que llamaremos a lo largo de este documento el *mejor clique máximo*.

Sin embargo, el problema de encontrar todos los *cliques* máximos de un grafo y el problema de encontrar sólo el *clique* máximo más grande, son problemas NP-completos [Balas y Yu, 1986] [Bolles y Haraud, 1986] [Davies, 1991]. Este es un problema muy

[2]Dado un grafo $G$, un *clique* de $G$ es cualquier subconjunto de nodos de $G$ en el que todos los nodos están conectados entre sí. Un *clique máximo* (*maximal clique*) es un *clique* al que si se le añade cualquier otro nodo del grafo deja de ser un *clique*.

[3]El *clique máximo más grande* (*largest maximal clique*) de un grafo $G$, es el *clique* máximo de $G$ con mayor número de nodos.




(a) Par estéreo de imágenes segmentado

(b) Grafo de asociación

**Figura 2.5:** Un ejemplo sencillo de errores en la segmentación.

importante que ha propiciado el desarrollo de diversas técnicas que intentan acelerar el proceso, como las implementaciones paralelas [El-Sonbaty y Ismail, 1998], o el uso de modelos conexionistas [Suganthan *et al.*, 1998]. Herault *et al.* [1990] derivan una función de coste asociada con la correspondencia de subgrafos relacionales, de modo que convierten el problema de la correspondencia en un algoritmo de *simulated annealing* usando el método de Metropolis.

Otro problema importante que aparece es que los pesos en los nodos pueden ser erróneos debido a que las fuentes de ruido influyen en la medición de la similitud entre las características.

Ranganath y Chipman [1992] usan una técnica de relajación difusa que incluye información estructural en los pesos de los nodos. En este método, los arcos también tienen un peso, que mide el ratio de compatibilidad entre dos nodos, construyendo así un *grafo de asociación mejorado* (*enhanced association graph*). Los pesos de los nodos se modifican mediante un proceso de relajación iterativo que depende del número y peso de los nodos conectados a cada nodo y la fortaleza de su relación. Si un nodo tiene un buen soporte de los nodos conectados a él su peso aumenta, y en caso contrario, disminuye. Después del proceso de relajación, los nodos sin soporte contextual (peso pequeño) se eliminan, reduciendo así el tamaño del grafo para la búsqueda del mejor *clique* máximo, de modo que se disminuye el primer problema.

### 2.2.8 Correspondencia de regiones

Los métodos basados en primitivas se pueden clasificar por el tipo de primitiva utilizada en el proceso de correspondencia. La selección de la primitiva es arbitraria y la única generalización real que se puede hacer es que la primitiva debe ser una parametrización útil de la imagen. En general, una primitiva útil debe tener las siguientes propiedades: unicidad, repetibilidad y significado físico [Lane y Thacker, 1994]. El objetivo es proporcionar correspondencias no ambiguas. La mayor parte de los trabajos basados en primitivas se han basado en la extracción de los contornos de la imagen (*edgels*, segmentos lineales, segmentos curvos) y en algunos casos intersecciones entre contornos (*junctions, corners*).

A medida que crece el contenido semántico de la primitiva, crece la dificultad de extracción de la misma, y crece la dispersión del mapa de disparidades. Pero al mismo tiempo, los algoritmos suelen ser más eficientes debido a que la cantidad de primitivas en la imagen disminuye, y las correspondencias obtenidas son más fiables debido a que se reduce el número de ambigüedades.

Las regiones son primitivas de alto nivel, con un elevado contenido semántico, lo que reduce la posibilidad de ambigüedades en la correspondencia. Marapane y Trivedi [1989] [1994] propusieron el uso de regiones como primitivas de la correspondencia debido a su estabilidad y su capacidad descriptiva, lo que las hace más tolerantes al ruido que las primitivas basadas en fronteras.

El uso de regiones hace que algunas de las restricciones de la correspondencia estéreo se asuman implícitamente o que sean más fáciles de introducir. Otra ventaja es el hecho de que en una imagen suele haber un número menor de regiones que de otras primitivas como son las basadas en fronteras. Por tanto, establecer las correspondencias será más eficiente y el número de correspondencias falsas o perdidas se reducirá.

Sin embargo, la correspondencia basada en regiones proporciona un mapa de disparidades muy disperso, por lo que Marapane y Trivedi concluyen que debería ser el primer paso de un sistema estereoscópico jerárquico donde el problema de la correspondencia debe resolverse primero usando regiones y después, fronteras y pixeles, de modo que finalmente se obtenga un mapa denso de disparidades. Una vez que se ha resuelto el problema de la correspondencia usando regiones, las fronteras y los pixeles pueden usarse para generar un mapa de disparidades de resolución fina.

De Agapito [1996] propone un sistema de dos niveles: regiones en el nivel alto de la jerarquía y esquinas en el segundo nivel. Obtiene la profundidad en las esquinas mediante triangulación y realiza una reconstrucción que consiste en la interpolación lineal entre las profundidades obtenidas. También hay métodos jerárquicos en los que en primer lugar se establece la correspondencia entre segmentos de las fronteras y después se calcula la disparidad entre los pixeles intermedios [Baker y Binford, 1981], por ejemplo, mediante adaptación de una función polinomial [Lacey *et al.*, 1998]. Otros métodos establecen la jerarquía entre diferentes conjuntos de regiones obtenidas con diferentes niveles de segmentación [Cohen *et al.*, 1989b] [Sander *et al.*, 1989], e incluso ha habido intentos de segmentar las imágenes al mismo tiempo que se establece la correspondencia en un algoritmo cooperativo [Cohen *et al.*, 1989a] [Randriamasy y Gagalowicz, 1991].

Muchos autores resaltan que el principal problema de los algoritmos basados en regiones es el manejo de las oclusiones. Sin embargo, veremos más adelante que en la técnica que se propone en el capítulo 4 no sólo se modelan las oclusiones, sino que además ayudan en el proceso de correspondencia.

Otra desventaja del uso de regiones radica en las diferencias entre los resultados de la segmentación de ambas imágenes. Pero en realidad ha habido pocos intentos de detectar o tener en cuenta estos efectos. Cuando se segmenta un par estéreo de imágenes, puede haber errores debido a, por ejemplo, el ruido, las condiciones de adquisición de las imágenes, las limitaciones de los procedimientos de segmentación, etc. Los errores que aparecen son los siguientes:

- *Regiones ocultas parcialmente*: regiones que aparecen en ambas imágenes pero con formas diferentes debido a oclusión parcial de una de ellas, situación en el borde de la imagen, etc.
- *Regiones ocultas*: regiones que aparecen sólo en una imagen.
- *Regiones fragmentadas*: algunas regiones adyacentes de una imagen pueden corresponder a una o más regiones en la otra imagen. Este problema también se conoce como sobre-segmentación o infra-segmentación[4] en el área de reconocimiento de objetos.

Las técnicas de correspondencia basadas en regiones tienen que ser capaces de manejar los errores de la segmentación. Las regiones ocultas parcialmente producen pesos incorrectos en los nodos, que pueden corregirse mediante técnicas de relajación que introducen información contextual en los pesos de los nodos. Las regiones que no aparecen en la otra imagen producen nodos, si los hay, con pocas compatibilidades con los demás nodos. Así, el proceso de relajación, que incluye información estructural en los pesos de los nodos, reduce el peso de estos nodos.

Sin embargo, el problema de las regiones fragmentadas es más difícil. Yang *et al.* [1989] corrigen la sobre-segmentación generando un *grafo de asociación aumentado*. El método considera todos los nodos que tratan de emparejar características diferentes de la imagen con la misma parte del modelo, y los une en un solo nodo. En su artículo prueban formalmente que la cardinalidad del clique máximo más grande no se incrementa cuando se unen dos nodos no adecuados y que se aumenta cuando se unen nodos adecuadamente. Por tanto, la mejor correspondencia es la que representa el clique máximo más grande del grafo de asociación aumentado, o uno de ellos si hay más de uno. Yang *et al.* realizan todas las uniones posibles en un proceso iterativo sin filtrar las uniones no adecuadas, lo que en casos reales podría hacer crecer demasiado el grafo de asociación.

Ranganath y Chipman [1992] corrigen la sobre-segmentación y la infra-segmentación teniendo en cuenta los pesos de los nodos. Su método considera el conjunto de todos

[4]En el área de reconocimiento de objetos, cuando se sobre-segmenta una imagen se obtienen más regiones que partes del modelo a reconocer (*oversegmentation*), y al contrario, la infra-segmentación (*undersegmentation*) ocurre cuando hay una región que se corresponde con varias partes del modelo.

los nodos que intentan emparejar la misma parte del modelo con regiones diferentes de la imagen, pero se unen sólo cuando todos los pesos de los nodos están por debajo de un determinado umbral. Por tanto, el número de nuevos nodos unión no es excesivamente grande. Sin embargo, en imágenes segmentadas reales una región fragmentada podría ser suficientemente grande para formar una buena correspondencia, mientras que algunas regiones pequeñas adyacentes a ella podrían quedarse sin correspondencia. Este método para la correspondencia de partes de un modelo con regiones de una imagen no funciona bien en el caso de correspondencia de regiones de dos imágenes debido a las numerosas diferencias que aparecen en la segmentación. Debido al criterio de similitud entre regiones, algunos emparejamientos se desechan desde el principio incluso antes de formar las uniones de regiones.

#### Criterios para la correspondencia de regiones

Los trabajos que usan regiones como primitiva para la correspondencia suelen utilizar los atributos de las regiones para calcular la similitud entre las mismas. Randriamasy y Gagalowicz [1991] usan el tamaño de la región, medias de características y posición del centro de gravedad. Marapane y Trivedi [1989] usan nivel de gris medio, área, perímetro, anchura, longitud, y razón de aspecto (*aspect ratio*). También proponen usar otras propiedades espectrales y espaciales tales como medidas de intensidad en canales simples o múltiples, compacidad, ejes mayor y menor, momentos, textura y descriptores topológicos. Cohen *et al.* [1989b] usan similitud en tamaño de las regiones, circularidad, posición del centro de gravedad, media de la intensidad, varianza de la intensidad, momentos espaciales, etc. Lee *et al.* [1993] usan algunos invariantes de momentos afines (*affine moment invariants*), que son los valores propios de una matriz que representa el movimiento aparente de la region entre dos imágenes.

Sin embargo, los errores de la segmentación pueden producir que algunas regiones en una imagen sean muy diferentes en tamaño y forma a las regiones correspondientes en la otra imagen.

### 2.2.9 Reconstrucción de la escena a partir de datos de rango

El objetivo final es, a partir de los datos de la profundidad, obtener una descripción lo más rica posible de la escena. Hay muchos trabajos sobre la segmentación de imágenes de rango entre los que destacan los basados en contornos y los basados en regiones [Besl y Jain, 1988].

Dado que es difícil la interpretación directa de una gran cantidad de datos de profundidad sin estructurar, Chen y Lin [1997] transforman estos datos en descripciones estructuradas mediante un método de segmentación de la información de rango que combina métodos basados en contornos y regiones de la imagen e incorpora técnicas de regresión para rechazar las falsas correspondencias (*outliers*) de los datos estéreo.

Wildes [1991] considera la siguiente pregunta: ¿Es la distancia a las superficies visibles el único resultado deseable de la visión estéreo? Como alternativa considera la posibilidad de interpretar la disparidad estéreo en términos de orientaciones de las superficies y discontinuidades, además de la distancia. Belhumeur [1993b] afirma que debería mantenerse un mapa detallado de la geometría de la escena internamente (profundidad, orientación, etc.) durante la correspondencia, de modo que todos estos elementos cooperen en la optimización de la misma.

Tarel *et al.* [1995] [1996] y Vezien *et al.* [1991] hacen la reconstrucción en base a primitivas geométricas de alto nivel basándose únicamente en las características geométricas globales de las proyecciones de las superficies sobre las imágenes. En estos métodos, se parte de un conjunto de pares de regiones correspondientes de donde se eliminan aquéllos pares que incluyen regiones mal segmentadas, haciendo un test de coherencia basado en invariantes, y sólo se reconstruyen pares de regiones válidos según este test. Un método donde se combinan la geometría y la fotometría permite además el tratamiento de las zonas ocultas [Tarel, 1996].

Otros autores prefieren considerar las superficies 3D de la escena durante la correspondencia en vez de durante la reconstrucción. Chabbi y Berger [1996] buscan superficies planas 3D entre las posibles correspondencias de caras 2D extraidas de las imágenes, basándose en la reconstrucción de los contornos de las caras 2D de modo que los segmentos de la superficie 3D sean coplanares. Sin embargo, necesitan del uso de un sistema trinocular para obtener tripletes de caras 2D que son validados mediante principios de geometría proyectiva.

En todos estos trabajos basados en regiones se asume generalmente una restricción importante sobre la forma de las superficies de la escena (superficies planas) que es utilizable en el caso de escenas de interiores y escenas urbanas.

## 2.3 Discusión

Como resumen de la revisión de las técnicas existentes de visión estereoscópica podemos extraer las siguientes conclusiones:

- Las **técnicas de correlación** clásicas tienen dos problemas principales que son el tratamiento de las oclusiones y las discontinuidades en la profundidad y el tratamiento de las superficies no fronto-paralelas.

  La **selección adaptativa de la ventana de correlación** y/o el uso de múltiples ventanas, permite evitar los problemas derivados de la presencia de oclusiones y discontinuidades en la profundidad, pero no saca provecho de la información que éstas proporcionan sobre la estructura tridimensional de la escena.

  Cada vez más, la selección adaptativa de la ventana intenta ceñirse a ciertas primitivas de más alto nivel como son los bordes obtenidos mediante un detector de bordes tipo Canny. Se intenta cubrir un máximo de formas geométricas alrededor del pixel central sin rebasar las discontinuidades en la intensidad que sean significativas.

- Las **técnicas basadas en programación dinámica** permiten obtener una solución óptima global para cada línea epipolar en lugar de un conjunto de óptimos locales, y además permiten modelar las oclusiones como parte del problema de la correspondencia.

  Sin embargo, en las investigaciones recientes, bien para aumentar la robustez de la solución, bien para reducir el espacio de búsqueda, o bien para ambas cosas, se tiende a incluir en el espacio de búsqueda información sobre los puntos de los bordes.

- Las dos conclusiones anteriores nos hacen pensar que las técnicas basadas en áreas tienden cada vez más a adaptarse a la información que proporcionan ciertas primitivas extraidas de las imágenes y que una generalización de esta tendencia sería considerar los bordes proporcionados por la segmentación de imágenes como único límite posible al crecimiento de la ventana adaptativa. Es decir, considerar las regiones resultantes de la segmentación, como primitivas para la correspondencia.

- En cuanto a los **métodos basados en primitivas**, según aumenta el nivel de la primitiva, aumenta la robustez de las correspondencias obtenidas, al mismo tiempo que se aumenta su dispersión. Debido a que es deseable obtener un mapa denso de disparidades, se han propuesto sistemas jerárquicos donde los resultados de la correspondencia de un tipo de primitivas sirven para inicializar la correspondencia de las primitivas del siguiente nivel en la jerarquía.

  Aunque algunos autores han propuesto el uso de **regiones como primitiva** (generalmente como nivel más alto del sistema jerárquico), las regiones han sido poco utilizadas debido a la dificultad que supone su extracción de la imagen, concretamente las diferencias en la segmentación de las regiones correspondientes, lo que dificulta los procesos de correspondencia y reconstrucción. Ha habido pocos intentos de resolver estos problemas producidos por los errores en la segmentación, y todavía no existe un método robusto que los solucione.

- Los **métodos basados en grafos** han sido abundantemente utilizados para resolver el problema de la correspondencia, sobre todo, correspondencia de segmentos lineales. Proporcionan una solución óptima, pero la búsqueda de los *cliques* máximos del grafo de asociación es un problema NP-completo.

- El proceso de **reconstrucción** de la escena tridimensional consiste generalmente en segmentar el mapa de profundidades obtenido a partir de las disparidades para obtener entidades de mayor nivel (superficies, objetos), que se puedan utilizar para las numerosas aplicaciones. En las técnicas en las que el mapa de profundidades es disperso además hace falta una etapa para completarlo, generalmente utilizando alguna técnica de interpolación. Aunque también se han propuesto métodos de reconstrucción basados en regiones, bajo la suposición de que representan superficies planas de la escena, éstos fallan cuando la segmentación de las regiones presenta los errores que ya se han mencionado.

Hemos elegido las regiones como primitiva de la correspondencia con el objetivo de, a partir de cada par de regiones correspondiente, obtener una reconstrucción de la superficie que representan. Para realizar la correspondencia entre regiones, tenemos que resolver los problemas derivados de las diferencias entre la segmentación de ambas imágenes. En esta línea se enmarca el trabajo que se presenta en el capítulo 3, donde se intentan resolver los problemas de segmentación mediante un método que une las regiones adecuadas para mejorar la correspondencia y la reconstrucción.

Sin embargo, la complejidad computacional de los métodos basados en grafos, nos ha llevado a buscar nuevas estrategias de correspondencia. La alternativa adoptada es considerar la correspondencia de regiones y la reconstrucción de las superficies que representan en un solo paso, de modo que se obtenga la escena 3D directamente. En el capítulo 4, se presenta un nuevo método que modela las superficies de la escena durante el proceso de correspondencia de modo que se maximice la correlación entre los pixeles no ocultos de las regiones. Aunque el problema ha de estar sujeto a ciertas restricciones comunes en otros trabajos (superficies planas), con este método se pretende obtener la reconstrucción de la escena teniendo en cuenta las oclusiones y las discontinuidades en la profundidad, en un algoritmo donde ambas cooperan en el proceso de correspondencia.

# Capítulo 3

# Correspondencia de regiones basada en grafos de asociación

Uno de los objetivos de este trabajo es establecer la correspondencia entre dos proyecciones de una escena para recuperar la información tridimensional de la misma. En este capítulo se realiza una serie de aportaciones a una familia clásica de técnicas, las basadas en grafos. Las regiones han sido poco utilizadas como primitiva en la literatura por la dificultad que supone su extracción de la imagen, y las dificultades que incorporan en la correspondencia los errores que se derivan de la segmentación. El trabajo que se presenta en este capítulo es el resultado de la investigación realizada en aras de la resolución de estos problemas en el ámbito de la correspondencia.

## 3.1 Introducción

En el área de aplicación de la robótica, las escenas de interés suelen ser escenas de interiores, donde la escena está formada principalmente por objetos fabricados por el hombre, con poca textura. En superficies poco texturadas, las técnicas basadas en áreas tienen dificultades, por lo que se suelen emplear técnicas basadas en características.

Las primitivas más utilizadas son las basadas en fronteras, como son los segmentos lineales, o los pixeles de la frontera. Aunque ha habido trabajos en los que se proponían primitivas de mayor contenido semántico que los segmentos lineales (segmentos curvos, intersecciones entre segmentos), son pocos los trabajos que proponen las regiones como primitiva de la correspondencia.

La correspondencia de regiones hace más fácil la introducción de algunas restricciones de la visión estereoscópica, tal como se comentó en el capítulo 2. El caso más claro lo constituye la restricción de continuidad. Debido a las propiedades de continuidad y cohesión de la materia, la disparidad (o la profundidad) debe variar de forma suave en los puntos adyacentes de la imagen, salvo en los límites de las superficies de la escena, donde pueden existir cambios bruscos de disparidad. Generalmente, los límites de las superficies suelen estar en los cambios bruscos de intensidad, mientras que las áreas de intensidad homogénea representan superficies donde la restricción de continuidad es aplicable. Las regiones representan áreas de intensidad homogénea, limitadas por discontinuidades en la intensidad y por tanto, se puede asumir que la función disparidad debe ser suave en el interior de las regiones, mientras que puede tener discontinuidades en las fronteras entre regiones.

Sin embargo, proporcionar un sólo valor de disparidad para cada región produce un mapa de disparidades muy disperso, por lo que Marapane y Trivedi [1989] proponen un método de correspondencia jerárquico que en primer lugar establezca la correspondencia entre regiones, para después establecer la correspondencia entre segmentos de las fronteras y finalmente entre pixeles.

Para establecer la correspondencia entre regiones utilizamos un método clásico de correspondencia basada en grafos [Ranganath y Chipman, 1992], dado que permite una representación de todas las posibles correspondencias en una estructura de datos donde se pueden incorporar fácilmente medidas de similitud entre regiones y de compatibilidad entre pares de regiones. Las técnicas basadas en grafos permiten a su vez la incorporación de técnicas de relajación que permiten añadir a cada posible emparejamiento información sobre el soporte que recibe del resto de emparejamientos. De este modo, el conjunto de pares de regiones resultante será más fiable. Además, el uso de regiones en el primer paso de la jerarquía supone normalmente un menor número de elementos en el grafo, dado que suele haber un número menor de regiones que de fronteras.

### 3.1.1 Trabajo previo

En la sección 2.2.7 se puede encontrar una revisión de los métodos de correspondencia basados en grafos, y en la sección 2.2.8 las ventajas y desventajas de la correspondencia basada en regiones, y las soluciones propuestas en la literatura a algunos de los problemas que conlleva.

Como método de partida, hemos elegido el método de correspondencia basado en el *grafo de asociación mejorado* desarrollado por Ranganath y Chipman [1992], debido a que este método incorpora numerosas mejoras sobre el método clásico: incorporación de pesos significativos en los nodos y en los arcos, una etapa de relajación que incorpora información contextual en los pesos de los nodos y un método para la resolución de los problemas derivados de los errores de la segmentación.

Aplicado a la visión estereoscópica, este método consiste en los siguientes pasos:

1. Obtener todas las regiones de las dos imágenes mediante algún método de segmentación.

2. Calcular los nodos iniciales del *grafo de asociación*. Cada nodo del grafo representa una posible correspondencia entre una región de la imagen izquierda, $L_i$, y una región de la imagen derecha, $R_j$. A cada nodo se le asigna un peso, $S(L_i, R_j)$, que representa la *similitud* entre las dos regiones que forman el nodo, basándose en los atributos de las regiones. Para que un par de regiones constituya una posible correspondencia (un nodo), la similitud entre las mismas ha de ser mayor que un determinado umbral, $T_s$.

3. Crear nuevas regiones a partir de las regiones iniciales. Para cada región de la izquierda, $L_i$, si aparece en más de un nodo y todos los pesos de estos nodos están por debajo de un umbral, que llamaremos $T'_s$, esto significa que ninguno de ellos representa una buena correspondencia. En este caso, las regiones emparejadas con $L_i$ se consideran candidatas a unirse. Para cada par de estas regiones, si sus atributos son similares, se crea una nueva región con la unión de ambas, se calcuan sus atributos, se crea un nodo que representa la correspondencia entre $L_i$ y la nueva región, y se incluye esta región en el conjunto de las candidatas. Para cada región de la derecha se procede análogamente.

   A partir de aquí y a lo largo de este capítulo, a cada región formada por la unión de dos o más regiones resultantes de la segmentación la llamaremos *región-unión*, y a cada una de las regiones componentes de una región-unión la llamaremos *región-fragmento*.

4. Calcular los arcos del *grafo de asociación*. A cada arco se le asigna un peso, $B(L_i, R_j; L_h, R_k)$, que representa el grado de *compatibilidad* entre los dos nodos que conecta. Si dos nodos son incompatibles, entonces no existe arco entre ellos. La introducción de regiones-unión conlleva la aparición de un nuevo caso de incompatibilidad: los nodos que contienen una región-unión son incompatibles con los nodos que contienen a cada una de las regiones-fragmento que la forman.

5. Realizar el *proceso de relajación*, para incluir información estructural en los pesos de los nodos, y al finalizar, borrar los nodos que obtengan pesos por debajo de un cierto umbral.

6. Encontrar el *mejor clique máximo* del grafo (el *clique* máximo cuya suma de los pesos de los nodos que lo componen sea máxima).

### Cálculo del peso inicial de los nodos y de los arcos

Típicamente, el peso de cada nodo del grafo se calcula a partir de una serie de atributos de las primitivas,

$$S(L_i, R_j) = \frac{n - E(L_i, R_j)}{n} \tag{3.1}$$

donde $n$ es el número de atributos y $E(L_i, R_j)$ es la suma ponderada del valor absoluto de las diferencias de los atributos de cada región. Cada atributo se pondera de modo que el peso asociado a él determina su importancia relativa en el conjunto de atributos,

$$E(L_i, R_j) = \sum_{k=1}^{n} |a_k(L_i) - a_k(R_j)| w_k \tag{3.2}$$

donde $w_k$ es el peso del $k$-ésimo atributo y $a_k$ es el valor del mismo.

Los pesos de los arcos se calculan de manera similar,

$$B(L_i, R_j; L_h, R_m) = \frac{n - E'(L_i, R_j; L_h, R_m)}{n} \tag{3.3}$$

pero basándose en las *relaciones* entre las dos regiones implicadas en cada par de nodos del grafo en lugar de los *atributos* de las regiones.

$$E'(L_i, R_j; L_h, R_k) = \sum_{k=1}^{n} |a'_k(L_i, L_h) - a'_k(R_j, R_m)| w'_k \tag{3.4}$$

Estas relaciones se pueden representar en binario (0/1) o mediante un valor real ($[0, 1]$). Por ejemplo, la adyacencia entre dos regiones de una imagen se puede representar como adyacente/no adyacente o basándose en un porcentaje de frontera común en pixeles.

### Proceso de relajación

El proceso de relajación propuesto por Ranganath y Chipman [1992] es un proceso iterativo que consiste en modificar el peso de cada nodo basándose en su peso inicial, en los pesos de sus arcos (que son constantes durante el proceso) y en los pesos de los nodos conectados a él. La regla de relajación permite integrar los pesos iniciales de los nodos y la consistencia relacional entre los nodos para conseguir una correspondencia mejor.

Los pesos de los nodos en el paso $t + 1$ se modifican dependiendo de los pesos de los nodos en el paso anterior, $t$, y el peso inicial de los nodos, como sigue:

$$S(L_i, R_j)^{t+1} = \alpha S(L_i, R_j)^0 + (1 - \alpha) \frac{1}{N} \sum_{h=1}^{N} \left( \max_{k=1}^{M} \left( S(L_h, R_k)^t B(L_i, R_j; L_h, R_k) \right) \right) \quad \forall i, j \tag{3.5}$$

donde $N, M$ es el número de regiones en la izquierda y en la derecha, respectivamente, $\alpha$ es un parámetro que determina la influencia del peso inicial en el resultado, y $B(L_i, R_j; L_h, R_k)$ indica la compatibilidad entre los nodos $(L_i, R_j)$ y $(L_h, R_k)$, o lo que es igual, el peso del arco que conecta ambos nodos, si lo hay, o 0 si no lo hay.

Este proceso de relajación iterativo termina cuando la diferencia entre los pesos de una iteración y la siguiente es suficientemente pequeña,

$$|S(L_i, R_j)^{t+1} - S(L_i, R_j)^t| < \varepsilon \quad \forall i, j \tag{3.6}$$

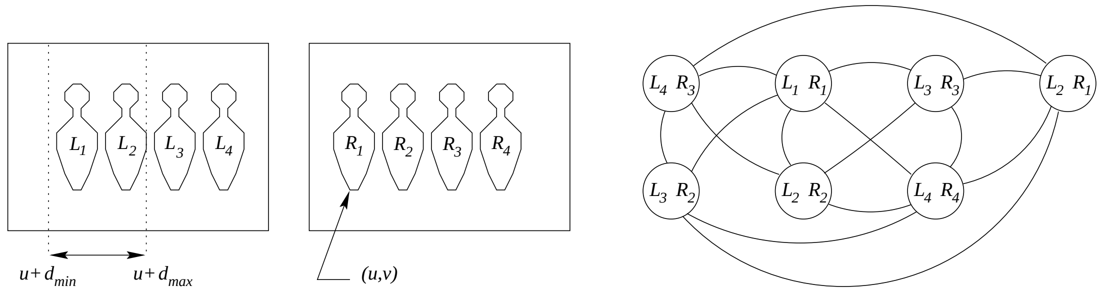
Los pesos de los nodos se normalizan en cada iteración, haciendo que la suma de los pesos permanezca constante,

$$S(L_i, R_j)^t \leftarrow \frac{S(L_i, R_j)^t \sum_{\forall h, \forall k} S(L_h, R_k)^0}{\sum_{\forall h, \forall k} S(L_h, R_k)^t} \quad \forall i, j \tag{3.7}$$

Sin embargo, cuando se incluyen uniones de regiones en el grafo de asociación, es necesario tenerlo en cuenta en el proceso de relajación. Esto es debido a que un nodo que contiene una región-unión, tiene menos sumandos en el sumatorio de la ecuación 3.5 que las regiones que la componen. Para compensar esta falta de términos, se incluye un factor de corrección $b_i$ en la regla de relajación. Supongamos que hay $N$ regiones en la imagen izquierda y $M$ en la derecha, incluyendo las nuevas regiones unión. Entonces la regla de relajación será:

$$\begin{aligned} S(L_i, R_j)^{t+1} &= \alpha S(L_i, R_j)^0 + (1 - \alpha)\, \frac{1}{N} \\ & \left[ b_i S(L_i, R_j)^t + \sum_{h=1}^{N} \left( \max_{k=1}^{M} \left( S(L_h, R_k)^t B(L_i, R_j; L_h, R_k) \right) \right) \right] \end{aligned} \tag{3.8}$$

Esta será la técnica de relajación que usaremos en todos los experimentos realizados.




(a) Par estéreo de imágenes segmentado (b) Grafo de asociación

**Figura 3.1:** Correspondencia basada en grafo de asociación, sin errores de segmentación.

### 3.1.2 Motivación

Como se vió en la sección 2.2.7 del capítulo anterior, se han usado técnicas basadas en grafos en muchos trabajos sobre correspondencia estéreo y reconocimiento de objetos. Estos trabajos obtienen resultados satisfactorios en la correspondencia cuando los errores de segmentación no son significativos.

En el área de la visión estereoscópica, el problema de las diferencias entre la segmentación de la imagen izquierda y la segmentación de la imagen derecha es mayor que en el área del reconocimiento de objetos, donde se dispone de un modelo del objeto u objetos que aparecen en la escena.

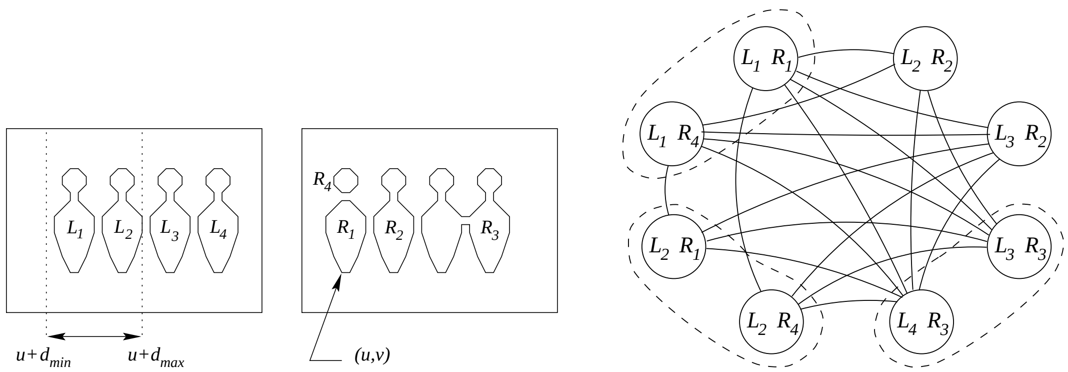
Incluso al segmentar las dos imágenes del par estéreo aplicando el mismo método de segmentación con los mismos parámetros, aparecen diferencias entre las imágenes segmentadas izquierda y derecha, como se puede observar en las figuras de la sección de resultados experimentales (por ejemplo, figura 3.10). Estas diferencias pueden provocar que algunas regiones se queden sin correspondencia y que algunas correspondencias del resultado sean incompletas o incluso erróneas.

La figura 3.1(b) muestra el grafo de asociación en un ejemplo libre de errores de segmentación (figura 3.1(a)). La restricción de disparidad permite reducir el número de nodos del grafo, por ejemplo, $L_2$ se podría corresponder con $R_1$ o con $R_2$, pero no con $R_3$ o con $R_4$. La restricción de orden permite establecer la incompatibilidad de algunos nodos, por ejemplo, si existiera el nodo $(L_1, R_2)$ sería incompatible con el nodo $(L_2, R_1)$ porque se invierte el orden de las correspondencias a lo largo de la línea epipolar. Dos nodos que contienen la misma región son incompatibles en base a la restricción de unicidad (por ejemplo, no puede existir un arco entre los nodos $(L_1, R_1)$ y $(L_2, R_1)$). Si no hay diferencias significativas entre los resultados de la segmentación de ambas imágenes, los atributos de las regiones proporcionan la información suficiente para calcular los pesos de los nodos y de los arcos, de modo que el mejor *clique* máximo representa la mejor correspondencia.

La figura 3.2(b) muestra el grafo de asociación en un ejemplo con errores de segmentación (figura 3.2(a)). Para subsanar estos errores, deberían crearse dos regiones-unión nuevas, $L_{(3,4)}$ y $R_{(1,4)}$. En la figura se han rodeado con línea discontinua los grupos de nodos que están relacionados con estas nuevas regiones.

En el grafo aparecen multitud de *cliques* máximos que podrían constituir la solución, y pequeños errores en los atributos de las regiones (debidos por ejemplo al ruido o las oclusiones) pueden hacer que el *clique* máximo de mayor peso no sea la mejor solución. El proceso de relajación es el que se encarga de incluir la información estructural en los pesos de los nodos, de tal forma que crecen los pesos de los nodos de compatibilidad elevada con nodos cuyos pesos son altos, mientras que decrecen los pesos de los nodos con pocos nodos compatibles o cuyos pesos de similitud y/o compatibilidad son bajos. De esta manera, se refuerzan los pesos de los nodos compatibles entre sí y se evitan las posibles incidencias en el cálculo de la similitud a partir de los atributos.

Según el método de Ranganath y Chipman [1992], para que se creen las regiones nuevas a partir de las regiones-fragmento tienen que ocurrir dos cosas: que inicialmente todas las regiones-fragmento formen un nodo con la región común (es decir, que la similitud de cada emparejamiento sea mayor que $T_s$), y que estos nodos sean considerados como candidatos para la unión de regiones (es decir, que la similitud sea menor que $T'_s$).




(a) Par estéreo de imágenes segmentado

(b) Grafo de asociación sin unión de regiones

**Figura 3.2:** Correspondencia basada en grafo de asociación, con errores de segmentación.

En la figura 3.2(b) se puede ver rodeados por línea discontinua los grupos de nodos que deberían dar lugar a una región-unión.

Vamos a explicar los problemas de esta aproximación con el ejemplo de la figura 3.2. Por un lado, el peso inicial del nodo $(L_1, R_4)$ es muy probable que sea bajo, ya que, como se puede observar, los atributos usuales de las regiones (área, perímetro, momentos de inercia, etc.) son muy diferentes. Si disminuimos $T_s$ lo suficiente como para que $(L_1, R_4)$ sea considerado como un emparejamiento posible podrían introducirse muchas falsas correspondencias en el grafo, lo que haría crecer el esfuerzo computacional. Análogamente sucedería con el nodo $(L_2, R_4)$. Por lo tanto, necesitamos de un criterio de similitud que reconozca la similitud entre dos regiones incluso en estas condiciones difíciles.

Por otro lado, si el peso inicial del nodo $(L_1, R_1)$ es alto, podría considerarse como un emparejamiento suficientemente bueno y por tanto no se crearía la nueva región-unión (y análogamente con el nodo $(L_2, R_1)$. Para facilitar la unión de $R_1$ y $R_4$ habría que aumentar excesivamente el umbral $T'_s$, lo que podría generar falsas regiones-unión que pasarían a engrosar el tamaño y la complejidad del grafo. Haría falta un método más elaborado que permitiera discriminar entre las uniones de regiones que pueden proporcionar mejores correspondencias y las que no.

También podría haber problemas si, además de las regiones-fragmento, existiera otra región en la imagen cuya similitud con la región común fuera alta. Aunque este caso no es habitual, se puede observar en el ejemplo que si el peso del nodo $(L_1, R_2)$ es mayor que el umbral $T'_s$, las regiones $R_1$ y $R_4$ no se unirían y el problema quedaría sin resolver.

### 3.1.3 Aportaciones

En este capítulo, se propone un nuevo método para manejar los errores derivados de la segmentación, que consiste en un paso de preproceso previo al cálculo clásico del grafo de asociación, realizando uniones de regiones que tengan probabilidad de poder encontrar su correspondiente de una forma más exacta.

Este método considera todos los casos posibles, pero rechaza las uniones que no proporcionen una buena correspondencia, comparando las correspondencias de regiones-fragmento a la correspondencia de la región-unión. Además, también se considera el caso de que la segmentación de áreas correspondientes en dos imágenes proporcione regiones muy diferentes. El método se basa en técnicas de grafos, construyendo un grafo previo (diferente al grafo de asociación) y calculando todos los conjuntos de regiones a unir.

También se propone un criterio de similitud adecuado para obtener una medida que reduzca el número de falsas correspondencias, y en consecuencia, el coste computacional. Los errores de la segmentación pueden producir que algunas regiones en una imagen sean muy diferentes en tamaño y forma a las regiones correspondientes en la otra imagen. Por tanto, proponemos el uso de técnicas basadas en correlación en lugar de atributos de las regiones para medir la similitud entre dos regiones. Como se verá más adelante, en el criterio de similitud se debe incluir además el tamaño del área de intersección entre las dos regiones con respecto a los tamaños de ambas regiones.

Por último, se propone un algoritmo subóptimo para resolver el problema de la búsqueda del mejor *clique* máximo. Este es el paso más costoso de todo el proceso de correspondencia, ya que se trata de un problema NP-completo. El algoritmo se basa en los pesos de los nodos para encontrar un *clique* máximo cercano a la solución. Aunque no se garantiza encontrar la solución óptima, los experimentos muestran que los resultados están muy próximos a la misma, consiguiendo reducir el coste computacional considerablemente.

El resto de este capítulo se organiza como sigue. En la sección 3.2 se presenta la etapa de preproceso para la unión de regiones basada en grafos. En la sección 3.3 se presenta el nuevo criterio de similitud basado en la correlación de las regiones. En la sección 3.4 se resume el algoritmo de preproceso desarrollado en las dos secciones anteriores y se detallan algunos aspectos de mejora en la eficiencia. En la sección 3.5 se detalla el algoritmo subóptimo para la búsqueda del mejor *clique* máximo de un grafo. Los resultados experimentales en presencia de diferentes niveles de ruido y los mapas de disparidades resultantes se presentan en la sección 3.6. Finalmente, en la sección 3.7 se discuten los resultados obtenidos y las conclusiones del trabajo presentado en este capítulo.

## 3.2 El problema de los errores en la segmentación

Proponemos sustituir el punto 3 del método de Ranganath y Chipman [1992] por un preproceso basado en grafos que una las regiones apropiadas. Hay algunos intentos de resolver el problema de los errores de segmentación uniendo algunos grupos de regiones [Yang *et al.*, 1989] [Ranganath y Chipman, 1992], como se detalla en la sección 2.2.8.

La idea principal es construir un grafo similar al grafo de asociación, donde los nodos representan todas las correspondencias posibles, con la diferencia de que los nodos que intentan hacer corresponder la misma región de una imagen están conectados con un arco. Este grafo, que denominaremos *grafo de asociación de incompatibilidades*, representa grupos de regiones que podrían unirse.

Aunque todas las uniones posibles son consideradas a priori, algunas uniones se rechazan mediante la realización de un test, basado en el *criterio de similitud* utilizado para calcular la similitud entre dos regiones. De cada grupo de regiones obtenemos una nueva región unión que es considerada posteriormente en el proceso de correspondencia. Finalmente, las nuevas regiones generan nuevos nodos que se añaden al conjunto inicial de nodos para construir el grafo de asociación del proceso de correspondencia.

### 3.2.1 El grafo de asociación de incompatibilidades

El grafo de asociación de incompatibilidades es un grafo donde los nodos representan todas las correspondencias posibles, y aquellos nodos que intentan hacer corresponder la misma región están conectados por un arco. Sin embargo, no todas estas incompatibilidades proporcionan uniones apropiadas, de modo que se debe aplicar algún filtro que rechace los arcos inadecuados. El grafo resultante representa todos los grupos de regiones que deberían ser unidas, de manera que las nuevas regiones resultantes se consideren para la correspondencia de regiones, que podrá realizarse, por ejemplo, con el método clásico basado en grafos [Ranganath y Chipman, 1992].

#### Nodos del grafo

Supongamos una imagen izquierda consistente en $N$ regiones $(L_1, L_2, ..., L_N)$ y una imagen derecha consistente en $M$ regiones $(R_1, R_2, ..., R_M)$. El primer paso para construir el grafo de asociación de incompatibilidades consiste en encontrar todas las posibles correspondencias entre regiones de ambas imágenes, usando la *restricción epipolar* y, si se conoce, la *restricción de disparidad* para reducir el espacio de búsqueda, además de un *criterio de similitud* para medir los pesos de los nodos y rechazar aquellos pares de regiones que no sean similares. Cada posible correspondencia es un nodo $(L_i, R_j)$ en el grafo de asociación de incompatibilidades, y el peso de dicho nodo, $S(L_i, R_j)$, es la medida de similitud entre ambas regiones.

### Incompatibilidades entre nodos

Dos nodos pueden ser incompatibles por varias razones, teniendo en cuenta la relación entre las regiones de una misma escena. Para unir las regiones apropiadas, interesan aquellas incompatibilidades que emergen del hecho de que dos nodos intenten establecer la correspondencia de una misma región (violación de la restricción de unicidad). Para crear los arcos del grafo de asociación de incompatibilidades, no nos interesan las incompatibilidades derivadas, por ejemplo, de la restricción de orden o la restricción de límite en el gradiente de la disparidad.

En la figura 3.3 se puede ver un grafo con arcos entre cada par de nodos que intentan establecer correspondencia con la misma región, correspondiente al ejemplo de la figura 3.2(a). Cada nodo que incluya una región-fragmento será incompatible con los nodos que intentan emparejar las otras regiones-fragmento con la misma región en la otra imagen. En consecuencia, habrá al menos una incompatibilidad entre cada par de regiones-fragmento.

Sin embargo, no todas estas incompatibilidades producen pares de regiones a unir. En el ejemplo (figura 3.2(a)) las regiones $L_1$ y $L_2$ podrían corresponderse con la región $R_1$ en la izquierda, pero no deberían unirse. Sin embargo, la región $R_1$ sí debería unirse a la región $R_4$ para corresponderse mejor con la región $L_1$.

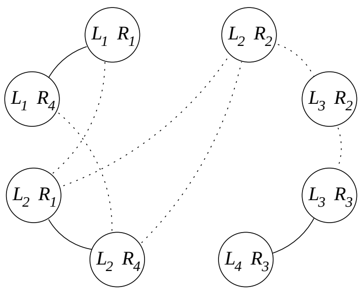



**Figura 3.3:** Grafo de asociación de incompatibilidades del ejemplo anterior: las líneas discontinuas corresponden a los arcos rechazados.

### Arcos del grafo

Hay que rechazar todas las incompatibilidades que no proporcionan regiones unión adecuadas y establecer arcos en el grafo para el resto de incompatibilidades. Se testean todos los pares de emparejamientos incompatibles para ver si generan una región-unión que se empareja con la región común mejor que cada región individual por separado. Si ambas regiones juntas no se emparejan con $L_k$ mejor que por separado, éstas no deberían unirse, y por tanto no se crea el arco.

Sean $(L_k, R_p)$ y $(L_k, R_q)$ dos nodos incompatibles y $R_{(p,q)}$ la región resultante de unir $R_p$ y $R_q$.

$$\begin{array}{ll} \textbf{si} & S(L_k, R_{(p,q)}) > S(L_k, R_p) \quad \textbf{y} \quad S(L_k, R_{(p,q)}) > S(L_k, R_q) \\ \textbf{entonces} & \text{establecer arco entre nodo } (L_k, R_p) \text{ y nodo } (L_k, R_q) \end{array} \tag{3.9}$$

donde $S()$ es el criterio de similitud entre dos regiones.

En el ejemplo de la figura 3.3, el arco entre los nodos $(L_1, R_1)$ y $(L_2, R_1)$ es rechazado porque el emparejamiento de la región-unión $L_{(1,2)}$ con $R_1$ no es mejor que el emparejamiento de cada región-fragmento con $R_1$ (véase figura 3.2(a)). Sin embargo, el arco entre los nodos $(L_3, R_3)$ y $(L_4, R_3)$ se crea porque la unión de las dos regiones, $L_{(3,4)}$, se corresponde con $R_3$ mejor que cada región-fragmento por separado.

#### Búsqueda de todos los *cliques* máximos

El último paso del preproceso consiste en encontrar todos los *cliques* máximos en el grafo de asociación de incompatibilidades. Cada *clique* máximo representa un grupo de regiones en una imagen que se emparejan con una región común en la otra imagen. Para cada *clique* máximo se construye una nueva región como la unión de todas las regiones-fragmento.

En el ejemplo, sólo hay tres *cliques* máximos formados por dos nodos cada uno, dos de los cuales dan lugar a la misma región-unión. Las regiones $R_1$ y $R_4$ unidas forman una nueva región, $R_{(1,4)}$, las regiones $L_3$ y $L_4$ forman la región $L_{(3,4)}$.

A continuación se calculan los emparejamientos que conciernen a las nuevas regiones y se añaden al conjunto inicial de nodos. Para calcular los arcos del grafo de asociación, todos los nodos concernientes a las regiones-unión se consideran incompatibles con todos los nodos concernientes a las regiones-fragmento que lo forman. En la figura 3.4 se muestra el grafo de asociación resultante.

Como el número de incompatibilidades entre nodos es normalmente bajo, el proceso de encontrar todos los *cliques* máximos es bastante rápido si usamos un algoritmo apropiado que saque ventaja de este hecho.

Las nuevas regiones proporcionarán nuevos posibles emparejamientos. Incluso pueden aparecer emparejamientos entre regiones-unión que se hayan añadido al grafo. Todos los nodos competirán en el proceso de correspondencia para encontrar la mejor solución.

## 3.3 El criterio de similitud

Sean dos imágenes $I_1$ e $I_2$ rectificadas, de modo que los puntos correspondientes están en la misma horizontal en la otra imagen, y sean $\mathcal{L}$ y $\mathcal{R}$ los conjuntos de regiones de

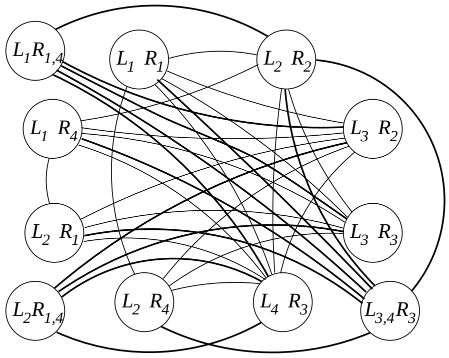



**Figura 3.4:** El grafo de asociación del ejemplo anterior después de unir las regiones. Los nodos y arcos nuevos están resaltados en línea más gruesa.

la primera y la segunda imagen, respectivamente. Queremos calcular cuál es la similitud entre pares de regiones, de modo que se pueda discriminar entre correspondencias correctas e incorrectas y se puedan comparar unas similitudes con otras.

### 3.3.1 Requisitos del criterio de similitud

El criterio de similitud no sólo debe proporcionar un índice que sirva para rechazar las correspondencias incorrectas, sino que además debe permitir evaluar la bondad de cada emparejamiento. Este criterio debería cumplir los siguientes requisitos:

1. Aceptar la correspondencia entre regiones parcialmente ocultas.
2. Aceptar la correspondencia entre una región fragmentada en una imagen con la región completa en la otra imagen.
3. Aceptar la correspondencia entre una región-unión y su región correspondiente en la otra imagen, con un peso mejor que cada región-fragmento por separado.
4. Rechazar pares de regiones no correspondientes.

El criterio de similitud utilizado en la mayoría de los trabajos consiste en comparar algunas propiedades de las regiones como son el área, los momentos de inercia, la media de intensidad, etc. En la literatura se ha probado que estas propiedades son significativas [Mundy y Zisserman, 1992], proporcionando buenas medidas en el primer y tercer caso. Sin embargo, cuando una región de una imagen aparece muy fragmentada en la otra imagen las propiedades de los fragmentos pueden ser muy diferentes respecto de la región completa.

Para conseguir todos los objetivos, proponemos un criterio basado en la correlación. Las medidas de correlación se utilizan principalmente en las técnicas de correspondencia basadas en áreas, donde el área correspondiente tiene tamaño y forma idénticos en la imagen izquierda y derecha, ya que viene determinado por una ventana que selecciona un determinado vecindario alrededor de un pixel central. En nuestro caso, el criterio debe adaptarse para manejar regiones de tamaños y formas diferentes.

### 3.3.2 Coeficiente de correlación

Consideremos $L_i$ una región en la imagen izquierda y $R_j$ una región en la imagen derecha. Ambas regiones raramente encajan con exactitud, de modo que tenemos que superponer y desplazar una región respecto a la otra para localizar el punto de máxima correlación.

La función de correlación sólo puede estar definida donde ambas regiones están definidas; sin embargo, las regiones pueden tener formas muy diferentes y, por tanto, es necesario adecuar la definición de la correlación a este caso particular. Para ello, se calcula la correlación considerando aquellos pixeles de $L_i$ cuyo pixel correspondiente en la imagen derecha pertenezca a $R_j$.

Denotaremos las áreas de intersección de las regiones $L_i$ y $R_j$ a una disparidad dada $d$, como $L_{i,j,d}$ y $R_{i,j,d}$ en la primera y segunda imagen, respectivamente, es decir:

$$L_{i,j,d} = \{(x, y) \in L_i / (x + d, y) \in R_j\} \tag{3.10}$$

$$R_{i,j,d} = \{(x + d, y) \in R_j / (x, y) \in L_i\} \tag{3.11}$$

Sean $\mu_{L_{i,j,d}}$ y $\mu_{R_{i,j,d}}$ las intensidades medias de las áreas de intersección y $\sigma_{L_{i,j,d}}$, $\sigma_{R_{i,j,d}}$ sus desviaciones típicas. Los criterios de correlación detallados en la sección 2.2.4 pueden ser expresados en función de estas áreas del siguiente modo:

- *Criterio SSD (Sum of Squared Differences)*,

  $$C_{SSD}(L_i, R_j, d) = -\frac{1}{N} \sum_{\forall (x,y) \in L_{i,j,d}} (I_1(x, y) - I_2(x + d, y))^2, \tag{3.12}$$

  donde $N$ es el número de pixeles del área de intersección de ambas regiones. Este criterio da unos resultados muy malos, debido a la gran diferencia existente en las formas y tamaños de las áreas de intersección a distintas disparidades. El coeficiente de correlación calculado según este método varía entre 0 y un valor mínimo negativo que denotaremos $-K_{SSD}$, es decir, está en el rango $[-K_{SSD}, 0]$.

- *Criterio ZSSD (Zero-Mean Sum of Squared Differences)*,

  $$C_{ZSSD}(L_i, R_j, d) = -\frac{1}{N} \sum_{\forall (x,y) \in L_{i,j,d}} ((I_1(x, y) - \mu_{L_{i,j,d}}) - (I_2(x + d, y) - \mu_{R_{i,j,d}}))^2, \tag{3.13}$$

  Este criterio no mejora los resultados obtenidos aplicando el criterio SSD, al mismo tiempo que resulta más costoso de evaluar, debido a que es necesario calcular cada vez la intensidad media del área de intersección. El coeficiente de correlación calculado según este método está en el rango $[-K_{ZSSD}, 0]$.

- *Criterio ZNSSD (Zero-Mean Normalized Sum of Squared Differences),*

$$C_{ZNSSD}(L_i, R_j, d) = -\frac{\sum_{\forall(x,y)\in L_{i,j,d}}((I_1(x,y) - \mu_{L_{i,j,d}}) - (I_2(x+d,y) - \mu_{R_{i,j,d}}))^2}{N\sigma_{L_{i,j,d}}\sigma_{R_{i,j,d}}}, \tag{3.14}$$

  El coeficiente de correlación calculado según este método está en el rango $[-1, 0]$. Debido a que la ventana es de tamaño variable no se puede prescindir de la división por $N$, ni de la división por la desviación típica, ya que ambas desviaciones varían de una disparidad a otra.

- *Criterio ZNCC (Zero-Mean Cross-Correlation),*

$$C_{ZNCC}(L_i, R_j, d) = \frac{\sum_{\forall(x,y)\in L_{i,j,d}}((I_1(x,y) - \mu_{L_{i,j,d}})(I_2(x+d,y) - \mu_{R_{i,j,d}}))^2}{N\sigma_{L_{i,j,d}}\sigma_{R_{i,j,d}}}, \tag{3.15}$$

  El coeficiente de correlación calculado según este método está en el rango $[-1, 1]$. Al igual que en el caso anterior, no se puede prescindir de la división por $N$, ni de la división por la desviación típica.

En los numerosos experimentos realizados para evaluar los criterios de correlación definidos, los criterios SSD y ZSSD han dado resultados muy pobres, mientras que los criterios ZNSSD y ZNCC han dado buenos resultados. La normalización aportada por la desviación típica de ambas áreas de intersección parece ser muy importante para conseguir un resultado de correlación significativo, dado que el número de puntos correlacionados es muy diferente en cada cálculo.

Dado que los dos criterios que dan mejores resultados son también los más costosos de calcular, para reducir los cálculos a realizar se llevaron a cabo experimentos utilizando criterios aproximados, es decir, sustituyendo la media y desviación típica de cada área de intersección por la media y desviación típica de la región completa. La motivación principal de esta aproximación se basa en que si el método de segmentación proporciona regiones donde la variación de la intensidad es suave, es probable que la media y desviación típica de la región en su totalidad sea parecida a las medias y desviaciones típicas de ciertos subconjuntos de pixeles de la misma.

Sin embargo, esta suposición no se cumple cuando una región ha sido infra-segmentada en una imagen y se compara con fragmentos de la misma en la otra imagen. La región completa no tiene por qué tener media de intensidad parecida a los fragmentos. Los resultados experimentales mostraron que esta suposición no era válida, especialmente en estos casos.

La medida de similitud entre dos regiones $L_i$ y $R_j$ podría definirse inicialmente como el *coeficiente de correlación* máximo dado un intervalo de disparidades $\mathcal{D} = [d_{min}..d_{max}]$:

$$S(L_i, R_j) = \max_{d\in[d_{min}..d_{max}]}(C_{ij}(d)) \tag{3.16}$$

donde $C_{ij}(d)$ es la correlación entre las áreas de intersección de $L_i$ y $R_j$ con disparidad $d$, calculada mediante cualquiera de los criterios anteriores.

### 3.3.3 Normalización del coeficiente de correlación

Para poder realizar el proceso de relajación normalizaremos los índices de correlación al rango $[0, 1]$, de modo que 0 significa que las dos regiones son totalmente diferentes y 1 significa que son idénticas:

- *Criterio SSD*
$$C_{ij}(d) = \frac{K_{SSD} + C_{SSD}(L_i, R_j, d)}{K_{SSD}} \tag{3.17}$$

- *Criterio ZSSD*
$$C_{ij}(d) = \frac{K_{ZSSD} + C_{ZSSD}(L_i, R_j, d)}{K_{ZSSD}} \tag{3.18}$$

- *Criterio ZNSSD*
$$C_{ij}(d) = 1 + C_{ZNSSD}(L_i, R_j, d) \tag{3.19}$$

- *Criterio ZNCC*
$$C_{ij}(d) = \frac{1 + C_{ZNCC}(L_i, R_j, d)}{2} \tag{3.20}$$

Sin embargo, independientemente del criterio de correlación elegido, la medida de similitud definida en la ecuación 3.16 no satisface el último objetivo de la lista de requisitos, es decir, rechazar pares de regiones no correspondientes. Como ejemplo extremo, supongamos dos regiones que a cierta disparidad intersectan en dos pixeles. Estas regiones tendrán una excelente similitud a esa disparidad si los dos pares de pixeles tienen intensidades similares, como se puede ver en la figura 3.5. Podemos deducir que la mejor disparidad no viene determinada por la disparidad con la que se obtiene la máxima correlación, sino que hay que tener en cuenta también el tamaño de la región intersectada.

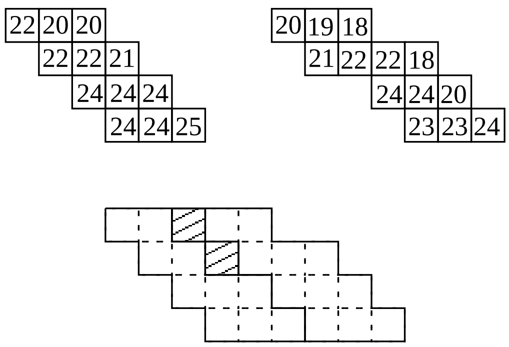



**Figura 3.5:** Necesidad de introducción de información sobre el área intersectada en el criterio de similitud basado en correlación: La disparidad de máxima correlación no es la mejor disparidad.

En la figura 3.6 se muestra otro ejemplo que nos induce a utilizar información sobre el área intersectada. En esta figura se muestran una región en la izquierda y dos regiones en la derecha que corresponden a la misma superficie de la escena real. Debido probablemente a que la superficie está inclinada respecto al plano imagen, las disparidades de correlación máxima entre la región de la izquierda y cada uno de los fragmentos varían ligeramente. Si en este ejemplo sólo consideramos el coeficiente de correlación, puede ocurrir que el valor de similitud de la región izquierda con cada uno de los fragmentos sea mayor que con la unión de ambos.

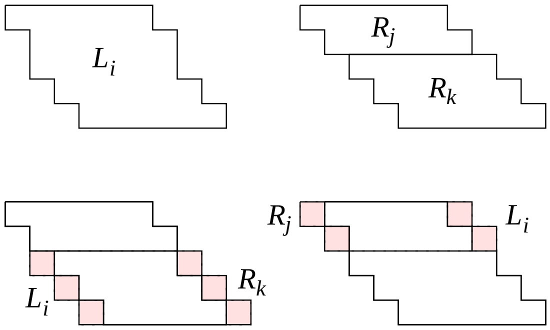



**Figura 3.6:** Necesidad de introducción de información sobre el área intersectada en el criterio de similitud basado en correlación: Un ejemplo de rechazo de la unión de dos regiones, por ser mayor la correlación de $L_i$ con $R_j$ que con $R_{j,k}$. Debería tenerse en cuenta que $R_k$ se puede quedar sin correspondencia.

El introducir información sobre el tamaño del área intersectada no debería producir otros errores, como puede ser que dos regiones cualesquiera se unan por intersectar una zona más amplia aunque el coeficiente de correlación sea bastante peor. Por ejemplo, en la figura 3.6 unir los dos fragmentos de la derecha sería un error si no se parecen lo suficiente a la región de la izquierda. Por tanto, el área de intersección debe influir, pero no introducir errores en el criterio de similitud.

### 3.3.4 Coeficiente de coincidencia

Para tener en cuenta el área de la zona intersectada entre las dos regiones, la medida de similitud puede definirse considerando junto al *coeficiente de correlación* un coeficiente que indique el porcentaje de área intersectada. Así, la ecuación 3.16 se transforma en

$$S(L_i, R_j) = \max_{d \in [d_{min}..d_{max}]} (N_{ij}(d) C_{ij}(d)) \tag{3.21}$$

donde $C_{ij}(d)$ es la correlación entre las áreas de intersección de $L_i$ y $R_j$ según la disparidad $d$, y $N_{ij}(d)$ es un coeficiente proporcional al área de la zona de intersección entre ambas regiones superpuestas con disparidad $d$.

La medida $N_{ij}(d)$, que a partir de ahora llamaremos *coeficiente de coincidencia*, debe estar normalizada, ya que el *coeficiente de correlación*, $C_{ij}(d)$, es una medida normalizada en el rango $[0, 1]$. De este modo, si alguno de estos dos coeficientes es un valor pequeño, el resultado será pequeño.

El coeficiente de coincidencia debería ser, pues, el número de pixeles correspondientes con respecto al tamaño de ambas regiones. Queda por determinar si la proporción ha de ser calculada respecto al área mínima o el área máxima de las dos regiones. Sin embargo, parece imposible cumplir todos los requisitos exigidos al criterio de similitud utilizando una sola expresión de similitud.

Definimos $N_{ij}(d)$ y $N'_{ij}(d)$ como los coeficientes de coincidencia con respecto al tamaño máximo y mínimo de ambas regiones, respectivamente,

$$N_{ij}(d) = \frac{A(L_{i,j,d})}{\max\left(A(L_i), A(R_j)\right)} \tag{3.22}$$

$$N'_{ij}(d) = \frac{A(L_{i,j,d})}{\min\left(A(L_i), A(R_j)\right)} \tag{3.23}$$

donde $A(L)$ indica el número de pixeles de la región $L$. El primer coeficiente resuelve los problemas planteados en las figuras 3.5 y 3.6, mientras que no funciona bien en el caso de las regiones parcialmente ocultas, que podrían ser rechazadas en el caso de que hubiera fragmentos pequeños. El segundo coeficiente es capaz de aceptar la correspondencia entre regiones parcialmente ocultas y rechazar los pares no correspondientes, pero puede dar malos resultados a la hora de crear las regiones-unión.

Si maximizamos el producto del primer coeficiente de coincidencia con el coeficiente de correlación, podemos calcular el valor de disparidad de mejor correspondencia entre ambas regiones. Llamemos *mejor disparidad* a este valor,

$$\hat{d}_{ij} = \operatorname{argmax}_{d \in [d_{min}..d_{max}]} \left(N_{ij}(d) C_{ij}(d)\right) \tag{3.24}$$

Finalmente, dada la mejor disparidad $\hat{d}_{ij}$ entre un par de nodos, proponemos una combinación de dos criterios,

$$S_1(L_i, R_j) = N_{ij}(\hat{d}_{ij}) C_{ij}(\hat{d}_{ij}) \tag{3.25}$$

$$S_2(L_i, R_j) = N'_{ij}(\hat{d}_{ij}) C_{ij}(\hat{d}_{ij}) \tag{3.26}$$

de usos muy diferentes:

1. Para la creación de los nodos del grafo, es necesario tener una medida de similitud que satisfaga los requisitos 1,2,4 de la lista. Para este propósito, el criterio apropiado sería $S_2$ (ecuación 3.26).

2. Para calcular $\hat{d}_{ij}$, comparar entre nodos y todas las operaciones en el proceso de correspondencia, es necesario cumplir los requisitos 1,3,4 de la lista. Para este propósito, el criterio apropiado sería $S_1$ (ecuación 3.25).

Nótese que el uso de dos medidas de similitud no causa un aumento importante en el coste computacional dado que la tarea más costosa es la búsqueda de la mejor disparidad (ecuación 3.24), que se realiza solamente una vez por cada emparejamiento.

### 3.3.5 Pesos de los arcos del grafo de asociación

Como consecuencia de la no utilización de los atributos de las regiones para el cálculo de los pesos de los nodos del grafo de asociación, tampoco se pueden utilizar las diferencias entre estos atributos para el cálculo de los pesos de los arcos [Ranganath y Chipman, 1992].

En este trabajo se propone calcular el peso de cada arco a partir de la diferencia entre las mejores disparidades de los nodos que conecta, $\hat{d}_{ij}$ y $\hat{d}_{hk}$, de modo que si la diferencia es cero, el peso será 1 y si la diferencia es la disparidad máxima, entonces el peso será 0.

$$B(L_i, R_j; L_h, R_k) = \frac{|\mathcal{D}| - |\hat{d}_{ij} - \hat{d}_{hk}|}{|\mathcal{D}|}. \tag{3.27}$$

De este modo, durante el proceso de relajación se refuerzan entre sí las correspondencias que tienen disparidades similares.

## 3.4 Algoritmo de la etapa de preproceso

Resumiendo lo expuesto en las secciones 3.2 y 3.3 podemos reescribir los pasos del preproceso previo al cálculo del grafo de asociación de la siguiente manera:

1. Para cada par de regiones $L_i$ y $R_j$, encontrar la mejor disparidad $\hat{d}_{ij}$ (ecuación 3.24), en un rango de posibles disparidades $\mathcal{D} = [d_{min}..d_{max}]$. Si el índice de similitud $S_2(L_i, R_j)$ (ecuación 3.26) es mayor que un cierto umbral $T_s$, se crea un nodo en el grafo de incompatibilidades para esta posible correspondencia y se asigna como peso del nodo el índice de similitud $S_1(L_i, R_j)$ (ecuación 3.25). En otro caso, el emparejamiento es rechazado.

2. Para cada par de nodos del grafo, $(L_i, R_j), (L_l, R_k)$, si alguna de las regiones coincide (por ejemplo, $i = l$), se considera la unión de las otras dos regiones (en el ejemplo, $R_m = R_j \cup R_k$), de modo que si el índice de similitud según el criterio $S_1$ de la región común con la región unión ($S_1(L_i, R_m)$) es mayor que cada uno de los pesos de ambos nodos, entonces se establece un arco entre ellos.

3. Se buscan todos los *cliques* máximos del grafo, y por cada *clique* máximo de dos o más nodos se crea una región-unión que incluya todas las regiones asociadas a la región común por estos nodos. Para cada región nueva, se calculan todos los emparejamientos posibles según el criterio establecido en el punto 1.

### 3.4.1 Mejoras en la eficiencia

La parte más costosa del algoritmo anterior es el cálculo de la mejor disparidad para un par de regiones dadas. Este cálculo se realiza en un bucle iterativo para todas las disparidades posibles $\mathcal{D} = [d_{min}..d_{max}]$ de modo que si en una disparidad dada, $d$, el producto de ambos coeficientes $N_{ij}(d)C_{ij}(d)$ es mayor que los anteriores, se guarda $d$, $A(L_{i,j,d})$ y $C_{ij}(d)$ para las siguientes iteraciones.

La parte más costosa computacionalmente es el cálculo de todas las correlaciones $C_{ij}(d)$ para todas las disparidades posibles. Sin embargo, el par de regiones sólo será aceptado como una posible correspondencia si cumple la condición,

$$N'_{ij}(d)C_{ij}(d) > T_s$$

Por tanto, podemos sacar partido de esta condición para intentar evitar calcular la correlación para algunas disparidades que de antemano podemos saber que no son la mejor disparidad. A continuación se detallan los aspectos que se pueden tener en cuenta para mejorar la eficiencia.

#### Rango efectivo de disparidades

El rango de disparidades posibles $\mathcal{D} = [d_{min}..d_{max}]$ además de depender de los límites de la disparidad, si los hay, se puede limitar para cada par de regiones teniendo en cuenta el rectángulo que inscribe a cada región, de modo que si a partir de una determinada disparidad los rectángulos de ambas regiones no intersectan, entonces el área de intersección es 0 y no es necesario considerar esos valores de disparidad.

Sea $V_i = (x_i^0, y_i^0), (x_i^1, y_i^1)$ el rectángulo que inscribe a la región $L_i$ y sea $V_j = (x_j^0, y_j^0), (x_j^1, y_j^1)$ el rectángulo que inscribe a la región $R_j$, donde $(x^0, y^0)$ indica el vértice superior izquierdo y $(x^1, y^1)$ indica el vértice inferior derecho de cada rectángulo. El *rango efectivo de disparidades* entre ambas regiones, que denotaremos $\mathcal{D}'_{ij} = [d'_{min}..d'_{max}]$ sería

$$d'_{min} = \max(d_{min}, x_i^0 - x_j^1) \quad (3.28)$$

$$d'_{max} = \min(d_{max}, x_i^1 - x_j^0) \quad (3.29)$$

#### Área mínima de intersección

Para que un par de regiones, $(L_i, R_j)$, se considere como una posible correspondencia, debe existir al menos un valor de disparidad $d$ dentro del rango efectivo de disparidades que cumpla que el producto $N'_{ij}(d)C_{ij}(d)$ sea mayor que un determinado umbral $T_s$.

Dado un valor de disparidad cualquiera $d_0$, si el coeficiente de coincidencia $N'_{ij}(d_0)$ es menor que $T_s$, entonces podemos descartar $d_0$ sin necesidad de calcular el coeficiente de correlación, ya que por definición $0 \leq C_{ij}(d) \leq 1, \forall d \in \mathcal{D}'_{ij}$, y en consecuencia,

$$N'_{ij}(d_0)C_{ij}(d_0) \leq N'_{ij}(d_0) \leq T_s \tag{3.30}$$

Por tanto, $d_0$ no cumple la condición necesaria para considerar el par $(L_i, R_j)$ como una posible correspondencia, y por tanto se puede descartar como mejor disparidad del mismo.

Dicho de otra manera, si se cumple que

$$A(L_{i,j,d_0}) \leq T_s \min(A(L_i), A(R_j)) \tag{3.31}$$

entonces el par $(L_i, R_j)$ con disparidad $d_0$ no es susceptible de formar un posible emparejamiento en el grafo. Por tanto, se puede establecer a priori un número mínimo de pixeles que debe contener la zona de intersección para cualquier disparidad $d$

$$A_{min} = T_s \min(A(L_i), A(R_j)) \tag{3.32}$$

de modo que si $A(L_{i,j,d_0}) < A_{min}$ entonces $d_0$ se descarta como mejor disparidad, y si $A(L_{i,j,d}) < A_{min}, \forall d \in \mathcal{D}'_{ij}$ el par $(L_i, R_j)$ se descarta como posible correspondencia.

Esta condición permite ahorrar un buen número de cálculos sobre las regiones:

- La intersección entre los dos rectángulos que inscriben a ambas regiones a una disparidad cualquiera $d$, es también un rectángulo, y el área de este rectángulo es el límite superior de $A(L_{i,j,d})$. Si el área de este rectángulo es menor que $A_{min}$, entonces $d$ se puede descartar sin necesidad de calcular siquiera el área de intersección.
- Si el rectángulo que inscribe el área de intersección es mayor que $A_{min}$, deberemos calcular el área de intersección $A(L_{i,j,d})$. Si se cumple $A(L_{i,j,d}) < A_{min}$ se puede descartar $d$ sin necesidad de calcular las medias y desviaciones típicas de las dos áreas de intersección, así como el coeficiente de correlación.

### Revisión del área mínima de intersección

Durante el cálculo de la mejor disparidad para un par de regiones, $L_i$ y $R_j$, dada una estimación de la mejor correlación calculada hasta el momento, $d_0$, se puede establecer un número mínimo de pixeles que debe contener la zona de intersección a otra disparidad $d_1$, de modo que si su área es menor que ese valor mínimo $d_1$ se puede desestimar sin necesidad de realizar el resto de los cálculos.

El razonamiento es análogo al anterior, esta vez especificando no el umbral de similitud mínima, sino la similitud máxima estimada hasta el momento.

En términos formales, dadas dos disparidades $d_0, d_1$

$$A(L_{i,j,d_1}) < A(L_{i,j,d_0})C_{ij}(d_0) \rightarrow d_0 \text{ es mejor que } d_1 \tag{3.33}$$

de modo que podemos descartar $d_1$ como *mejor disparidad.*

La demostración es sencilla: $d_1$ es mejor que $d_0$ si y sólo si su coeficiente de correlación es mayor que el de $d_0$, o lo que es igual, si y sólo si $N_{ij}(d_1)C_{ij}(d_1) > N_{ij}(d_0)C_{ij}(d_0)$. Se puede simplificar en ambos lados de la inecuación el máximo de las áreas de ambas regiones, quedando como sigue:

$$d_1 \text{ es mejor que } d_0 \leftrightarrow A(L_{i,j,d_1})C_{ij}(d_1) > A(L_{i,j,d_0})C_{ij}(d_0) \tag{3.34}$$

Sabiendo que $0 \leq C_{ij}(d) \leq 1, \forall d$ se cumple que

$$0 \leq A(L_{i,j,d})C_{ij}(d) \leq A(L_{i,j,d}), \forall d \tag{3.35}$$

Si se cumple que $A(L_{i,j,d_1}) < A(L_{i,j,d_0})C_{ij}(d_0)$ entonces

$$A(L_{i,j,d_1})C_{i,j(d_1)} < A(L_{i,j,d_0})C_{ij}(d_0) \tag{3.36}$$

y por tanto, como no se cumple la condición 3.34, $d_1$ no puede ser mejor que $d_0$.

Esta condición se aplica de la misma manera que la condición anterior. Podemos decir que el número de pixeles mínimo se inicializa con la ecuación 3.32, y cada vez que se encuentra una disparidad mejor que la estimada en el bucle de búsqueda de la mejor disparidad se actualiza el valor de esta variable

**si** $A(L_{i,j,d_1})C_{ij}(d_1) > A(L_{i,j,d_{ij}})C_{ij}(d_{ij})$ **entonces**
  $\hat{d}_{ij} \leftarrow d_1$
  $A_{min} \leftarrow \max\left(A_{min}, A(L_{i,j,d_1})C_{ij}(d_1)\right)$
**fin si**

En el caso de una región-unión se puede imponer como mínima similitud el máximo de las similitudes de las regiones hijas.

### Paralelización

Todos los cálculos relativos a regiones distintas se podrían paralelizar, aunque no se ha implementado un algoritmo paralelo en esta tesis. Una primera aproximación sería dividir la imagen izquierda en varias secciones a lo largo de las líneas epipolares, de modo que cada región de esta imagen esté asociada a una sola sección y cada sección pueda tener varias regiones asociadas. Las secciones deben solaparse para no dividir las regiones asociadas. La paralelización consistiría en dar a cada procesador las regiones involucradas en una sección, y las regiones involucradas en la misma sección en la imagen derecha (éstas pueden estar recortadas por los límites de la sección) para construir todos los nodos relativos a dicha sección.

## 3.5 Un algoritmo subóptimo para la búsqueda del mejor *clique* máximo

Aunque se ha dirigido un gran esfuerzo a la reducción del tamaño del grafo, el número de nodos puede ser todavía considerable cuando hay un número elevado de regiones por imagen. Buscar el mejor *clique* máximo es un problema NP-completo, de modo que el tiempo de cálculo crece exponencialmente con respecto al número de nodos del grafo.

Aquí proponemos un algoritmo subóptimo que tiene en cuenta los pesos de los nodos para reducir el tiempo computacional del método. Los nodos de mayor peso son los que tienen mayor probabilidad de formar parte de la solución final. Por tanto, podemos aprovechar la información proporcionada por el criterio de similitud (y mejorada por el proceso de relajación) para guiar la búsqueda del mejor *clique* máximo.

Dado el grafo de asociación $G = (N, A)$ donde $N$ es el conjunto de nodos y $A$ es el conjunto de arcos, proponemos el siguiente algoritmo subóptimo para obtener una solución al problema de hallar el mejor *clique* máximo:

1. Ordenar la lista $N$ en orden de peso decreciente.
2. Dividir la lista $N$ en dos listas: $N'$ con los $K$ primeros nodos de $N$ y $N''$ con el resto de nodos. $K$ se selecciona de modo que se cumpla alguna restricción de tiempo.
3. Obtener el subgrafo $G' = (N', A')$, donde $A' \subset A$ contiene solamente los arcos concernientes a los nodos de $N'$.
4. Buscar el mejor *clique* máximo en $G'$, que es un conjunto de nodos $S$.
5. Para cada nodo $n_i$ de $N''$ en orden de peso decreciente, si $n_i$ es compatible con todos los nodos de $S$ se añade $n_i$ a $S$, en otro caso se elimina $n_i$.

Este algoritmo no garantiza encontrar la mejor solución, pero, como se mostrará en los resultados experimentales, se obtiene una solución muy cercana a la óptima y en muchas ocasiones idéntica. El coste de este método es la suma de los costes de los cinco pasos, por lo que a continuación se estudian estos costes, siendo $n$ la longitud de la lista $N$, $K$ la longitud de la lista $N'$ y $n - K$ la longitud de la lista $N''$.

1. La ordenación de una lista de $n$ elementos tiene un coste medio de $2n \ln(n)$ y un coste cuadrático en el peor caso.
2. La división de $N$ en dos listas tiene un coste lineal en el peor caso, $O(n)$.
3. Obtener el subgrafo $G'$ tiene un coste lineal respecto al número de arcos de $G'$, que es $K^2/2$ en el peor caso.

4. Buscar el mejor *clique* máximo en $G'$ tiene un coste exponencial. Como la selección de $K$ se haría en función del tiempo máximo que se desee imponer al método, llamaremos $F(K)$ a este límite.

5. El cálculo de las compatibilidades de cada nodo de $N''$ con la solución estimada es lineal con respecto al tamaño de la solución, que inicialmente es $K$ en el peor caso. Por tanto, el cálculo de la compatibilidad de todos los nodos de $N''$, teniendo en cuenta que a la solución se le añaden los nodos compatibles en cada iteración, tiene coste cuadrático, $O(\frac{1}{2}(n-K+1)(n+K))$ en el peor caso.

En conclusión, si despreciamos los términos de coste inferior al cuadrático, el coste total es $O(K^2+F(K)+\frac{1}{2}(n^2-K^2))$ y dado que $K \leq n$, queda $O(\frac{1}{2}n^2+F(K))$. Aunque $F(K)$ es exponencial respecto de $K$, fijando un valor de $K$ que no sea muy grande, ponemos un límite a este término, de modo que el algoritmo es cuadrático respecto a $n$.

## 3.6 Resultados experimentales

Para validar el método en su totalidad, hemos de realizar pruebas con distintos tipos de segmentación y con distintos tipos de imágenes. En primer lugar, se mostrarán algunos de los experimentos realizados con dos métodos de segmentación diferentes. Se puede observar que nuestro método es capaz de realizar las uniones de regiones adecuadas para mejorar en lo posible la correspondencia de las regiones, independientemente del tipo de segmentación realizado.

Los experimentos que se muestran en este capítulo han sido realizados con el criterio ZNCC, que se utiliza comúnmente en las técnicas de estereovisión basadas en áreas [Papadimitriou y Dennis, 1995] [Devernay, 1996]). Sin embargo, los numerosos experimentos realizados muestran que el criterio ZNSSD proporciona resultados similares.

Un aspecto diferente, pero no menos importante de este problema, es la complejidad computacional, que depende del tamaño del grafo. Por tanto, en los experimentos se han evaluado todos los aspectos que influyen en el número de nodos del grafo, como son el umbral de similitud, $T_s$, y el número de nodos considerados en la búsqueda del mejor *clique* máximo, $K$.

Por último, para evaluar la robustez del método, se muestran los resultados obtenidos en una serie de imágenes sintéticas con distintos niveles de ruido.

### 3.6.1 Respuesta a diferentes métodos de segmentación

La figura 3.7 muestra un ejemplo de los pares de imágenes estereoscópicas utilizados en los experimentos. Hay un conjunto de pares sintéticos disponible

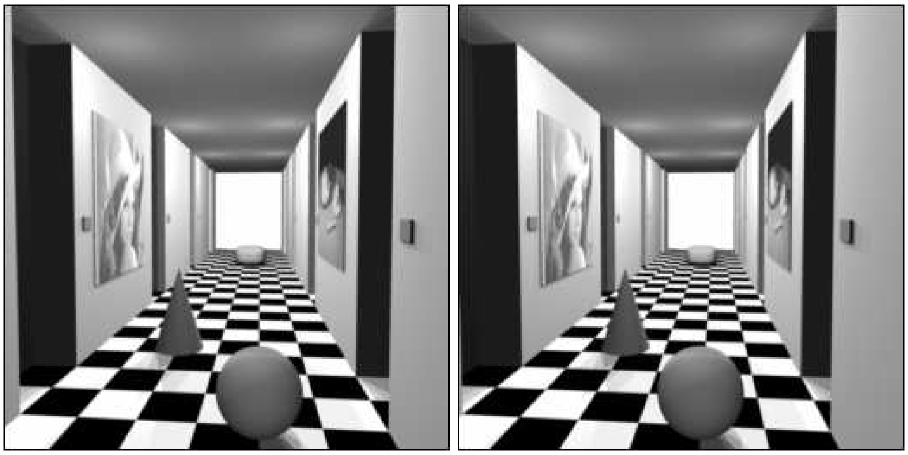


**Figura 3.7:** Par estéreo de imágenes sintéticas *corridor* (sin ruido).

en [Froehlinghaus, 1997] con diferentes niveles de ruido. En particular, la figura 3.7 muestra el par sin ruido.

Las figuras 3.8 y 3.10 muestran los resultados obtenidos al aplicar los dos métodos de segmentación empleados en los experimentos. Uno de los métodos utilizados para segmentar las imágenes es una técnica habitual de segmentación por unión de regiones [Rosenfeld y Kak, 1982] (figura 3.8). Otro método de segmentación utilizado en los experimentos es una técnica de agrupamiento desarrollada por Badenas *et al.* [1997] que agrupa pixeles cercanos en regiones dentro de una cierta varianza en el nivel de gris (figura 3.10). En todos los experimentos realizados, se ha aplicado el mismo método de segmentación a las dos imágenes del par estéreo, con los mismos parámetros de segmentación. En todos los casos, se aprecian diferencias en la segmentación de ambas imágenes, es decir, algunas regiones de la izquierda se corresponden con más de una región en la derecha y viceversa.

En el caso de segmentación mediante unión de regiones, filtramos las regiones de la imagen para evitar regiones muy pequeñas de pocos pixeles. En el caso de segmentación por agrupamiento, todas las regiones pequeñas se unen a la región más similar durante el proceso de agrupamiento [Badenas *et al.*, 1997]. Las áreas negras que se muestran en los mapas de disparidades de cada tipo de segmentación (figuras 3.9 y 3.11) corresponden a regiones filtradas (sólo en el primer caso) o regiones sin correspondencia (en ambos casos). Excepto para las áreas negras, las zonas más oscuras corresponden a puntos más lejanos en la escena.

Dado que asignamos disparidad constante dentro de cada región, los mapas de disparidades resultantes muestran una aproximación basada en planos frontoparalelos. Esta restricción se asume implícitamente también en el uso de la correlación. Esta aproximación del mapa de disparidades puede servir en una serie de aplicaciones de la robótica, donde el mapa que se necesita puede ser menos preciso, dado que este mapa da una idea general de la localización de los objetos en la escena. Para aplicaciones que necesi-

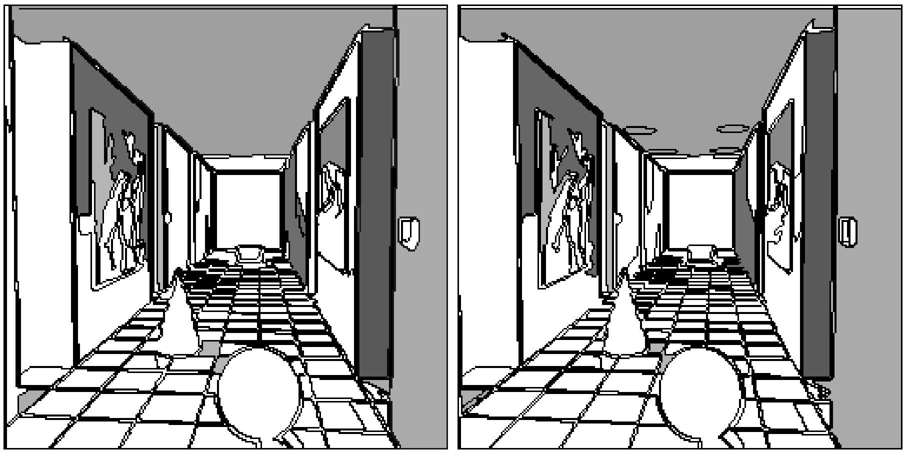


**Figura 3.8:** Segmentación por unión de regiones del ejemplo *corridor* y regiones unión resultantes del algoritmo.

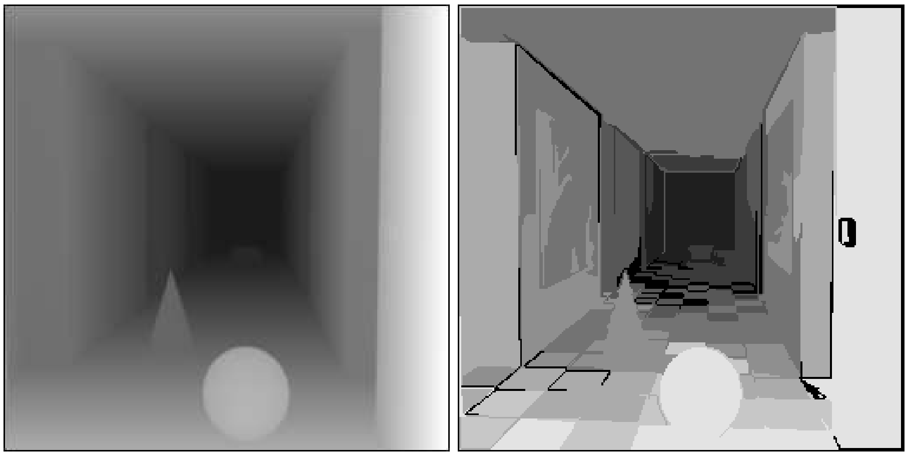


**Figura 3.9:** Mapa de disparidades real y mapa de disparidades obtenido a partir de la *mejor disparidad* de cada región.

ten precisión, este mapa puede ser una buena inicialización para la correspondencia de bordes y puntos en la escena.

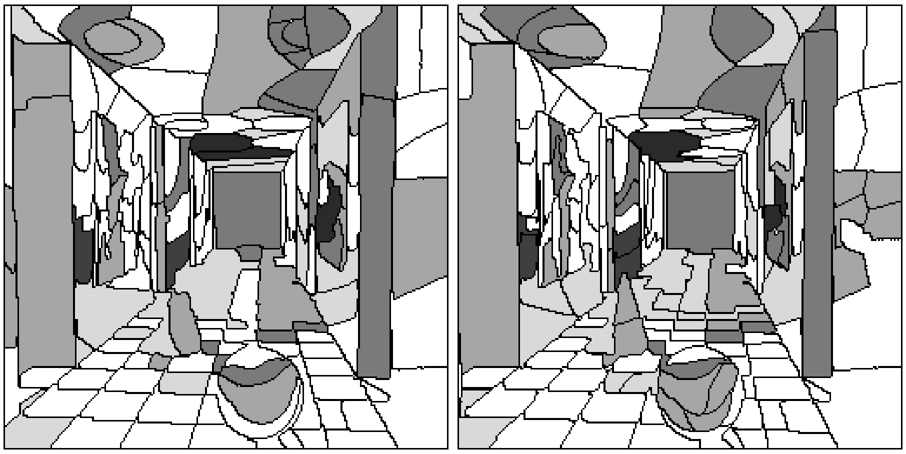


**Figura 3.10:** Segmentación por agrupamiento de regiones del ejemplo *corridor* y regiones unión resultantes del preproceso.

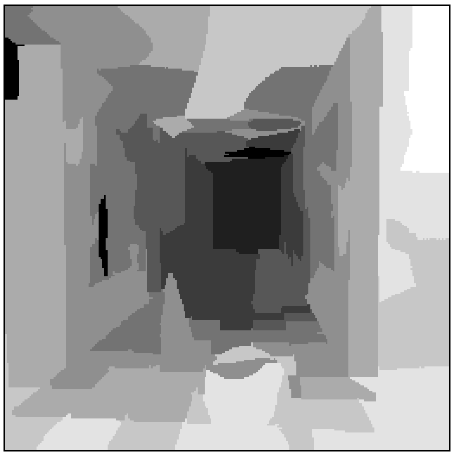
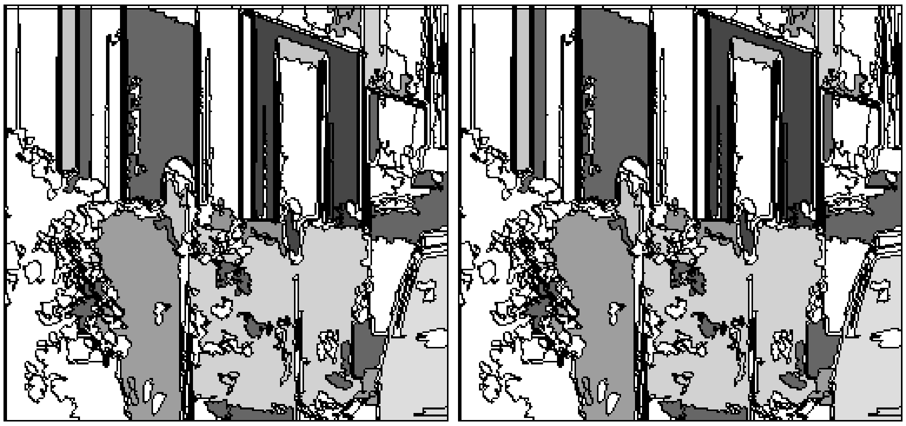


**Figura 3.11:** Mapa de disparidades obtenido a partir de la *mejor disparidad* de cada región.

La figura 3.8 muestra un ejemplo de las regiones que han sido unidas como resultado del algoritmo propuesto, y las regiones correspondientes en la otra imagen. Las regiones sombreadas con el mismo nivel de gris en la imagen izquierda son regiones unidas durante la etapa de preproceso y que después son emparejadas con su región o regiones correspondientes en la imagen derecha, también sombreadas con el mismo nivel de gris. Las correspondencias resultantes incluyen 8 regiones unión en la izquierda y 19 en la derecha, que contienen desde 2 hasta 5 regiones individuales.

Algunas de las uniones no son visibles porque algunas regiones consisten en contornos de las regiones o grupos de baldosas lejanas que se unen para emparejarse con grupos similares en la otra imagen. La figura 3.10 muestra los resultados con el segundo

(a) Par estéreo *parking meter*.



(b) Fronteras de las regiones (negro) y regiones-unión resultantes (grises).

**Figura 3.12:** Resultados de una escena real, el ejemplo *parking meter*, con segmentación mediante el método de unión de regiones.

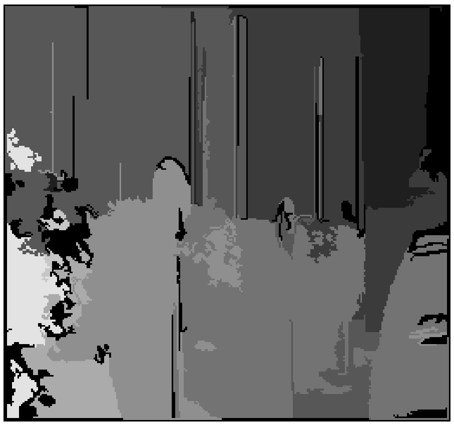


**Figura 3.13:** Mapa de disparidades resultante, ejemplo *parking meter*.

método de segmentación. Nótese que el algoritmo propuesto ha sido capaz de manejar el problema de las regiones fragmentadas de una imagen con respecto a la otra.

La figura 3.12 representa otro ejemplo, en este caso una escena real del conjunto de pruebas JISCT [Bolles *et al.*, 1993][1], del resultado del método propuesto para solucionar el problema de la fragmentación de regiones. Algunas regiones en la izquierda se han unido en el proceso para obtener correspondencias mejores con las regiones de la imagen derecha y así minimizar las diferencias en la correspondencia debidas a la fragmentación de las regiones. En la figura 3.13 se muestra el mapa de disparidades resultante.

Las figuras 3.14[2], 3.15 y 3.16 son otros ejemplos de escenas reales, donde se muestran algunas de las uniones de regiones obtenidas. No se muestran todas para preservar la claridad de los resultados. En los mapas de disparidades, que se han obtenido respecto a la imagen derecha, se puede observar que hay algunos errores en la correspondencia de algunas regiones del lado derecho de la imagen. Esto es debido a que sus regiones correspondientes se encuentran parcialmente fuera de la imagen.

También se pueden observar algunas imprecisiones en el cálculo de la disparidad, especialmente en aquellas regiones donde la varianza de la intensidad es muy pequeña (poca textura), como es, por ejemplo, la parte frontal de la mesa en la figura 3.16.

### 3.6.2 Complejidad computacional

Como ya se ha comentado, el cálculo de los *cliques* máximos es un problema NP-completo, de modo que cualquier algoritmo para hallar todos los *cliques* máximos en un grafo tiene un coste exponencial que depende del número de nodos en el grafo.

[1]Disponible en `http://www.ius.cs.cmu.edu/idb/html/jisct/index.html`

[2]`http://www.ius.cs.cmu.edu/idb/html/stereo/index.html`

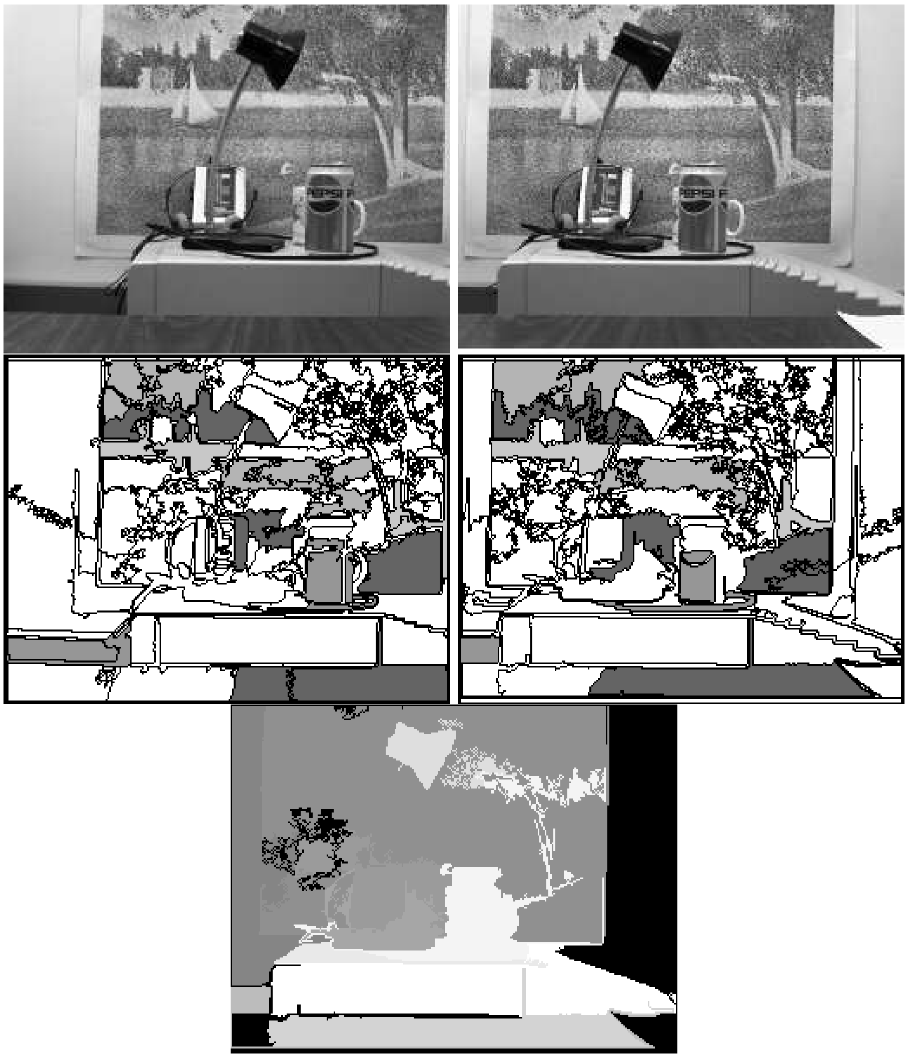


**Figura 3.14:** Resultados de una escena real, el ejemplo *pepsi*, con segmentación mediante el método de unión de regiones.

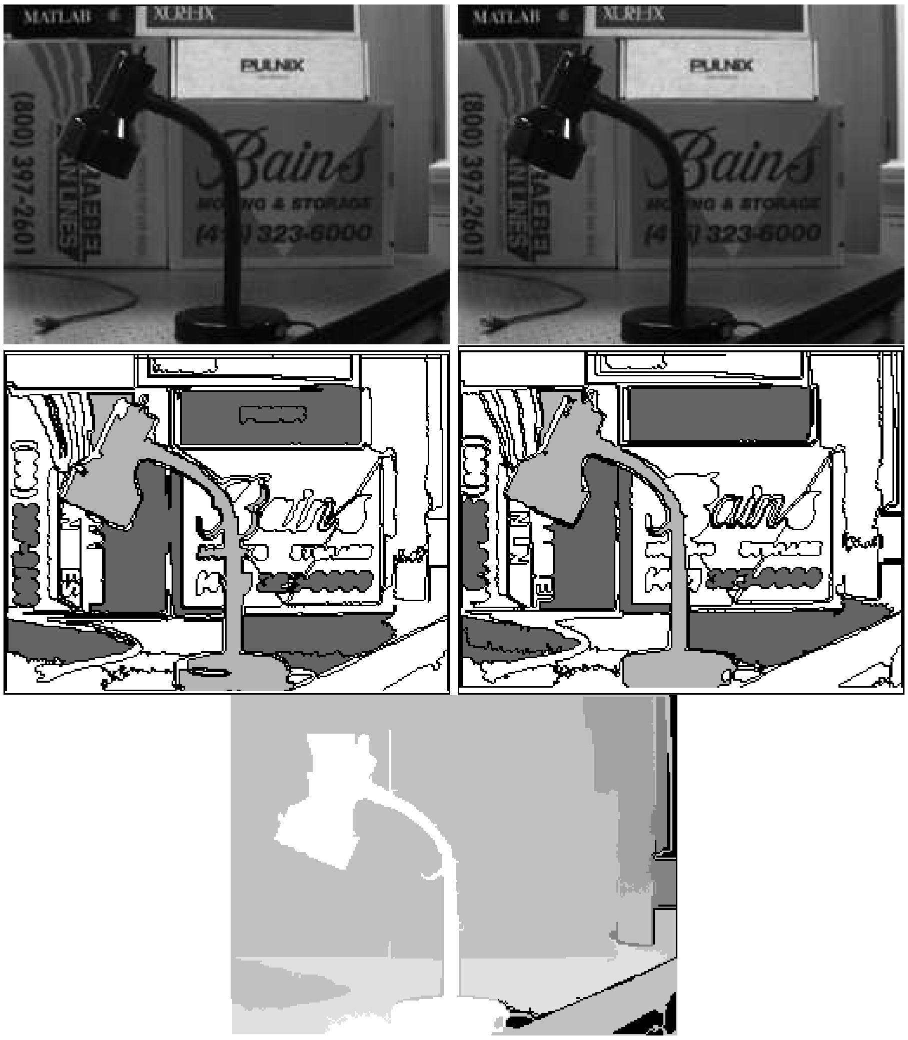



**Figura 3.15:** Resultados de una escena real, el ejemplo *textured lamp*, con segmentación mediante el método de unión de regiones.

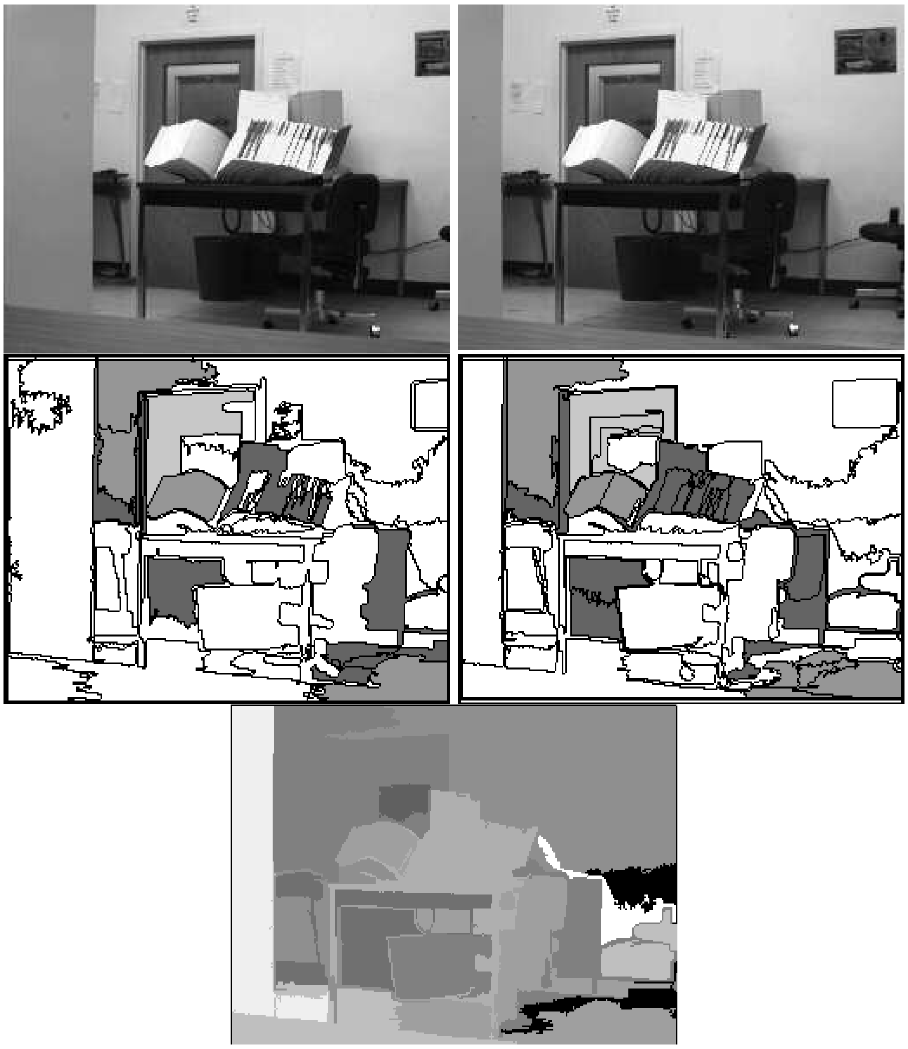


**Figura 3.16:** Resultados de una escena real, el ejemplo *lab*, con segmentación mediante el método de unión de regiones.

En el problema presentado de correspondencia de regiones, el número de nodos del grafo de asociación depende del umbral de similitud seleccionado para considerar los pares de correspondencias posibles. La tabla 3.1 muestra el coste computacional en segundos[3] del algoritmo utilizando umbrales diferentes. Para cada umbral de similitud se muestra también la proporción de correspondencias obtenidas, medido como el número de pixeles con correspondencia dividido por el total de pixeles de todas las regiones consideradas en la correspondencia. El número de regiones con correspondencia crece a medida que decrece el umbral porque se consideran más emparejamientos posibles en el proceso.

| umbral $T_s$ | 0.65 | 0.60 | 0.55 |
|---|---|---|---|
| # nodos | 70 | 85 | 100 |
| % correspondencia | 91.5 | 92.4 | 93.3 |
| tiempo (seg.) | 40.75 | 53.09 | 13969.49 |

**Tabla 3.1:** Porcentaje de correspondencia para el ejemplo *corridor* (sin ruido).

No obstante, para tratar de evitar la complejidad exponencial del cálculo de los *cliques* máximos, la tabla 3.2 muestra el resultado de aplicar el algoritmo de búsqueda descrito en la sección 3.5. Nótese que ahora los tiempos de cálculo han sido reducidos considerablemente con respecto a los tiempos mostrados en la tabla 3.1, y que de esta manera el coste computacional no se incrementa exponencialmente si el número de nodos se incrementa. Mas aún, el algoritmo subóptimo permite ejecuciones con umbrales más pequeños que proporcionan mejores correspondencias.

| umbral $T_s$ | 0.65 | 0.60 | 0.55 | 0.50 | 0.45 | 0.40 | 0.35 |
|---|---|---|---|---|---|---|---|
| # nodos | 70 | 85 | 100 | 118 | 141 | 192 | 248 |
| % correspcia. | 91.5 | 92.4 | 93.3 | 93.7 | 94.0 | 95.3 | 96.0 |
| tiempo (seg.) | 40.92 | 46.97 | 55.40 | 67.41 | 88.95 | 115.12 | 153.22 |

**Tabla 3.2:** Porcentaje de correspondencia para el ejemplo *corridor* (sin ruido), usando el algoritmo subóptimo con $K = 30$.

En este ejemplo, los resultados del algoritmo óptimo y del algoritmo subóptimo son exactamente los mismos. Cuando el número de nodos del grafo crece considerablemente, sólo se puede aplicar el algoritmo subóptimo. Aunque el algoritmo propuesto no garantiza encontrar la mejor solución, los resultados son idénticos en la mayoría de los experimentos realizados. En el resto de experimentos, los resultados son muy similares (alrededor del 1% de diferencia en el porcentaje de correspondencia). La figura 3.17

[3]Todos los tiempos de cálculo que se muestran en las tablas y gráficas de este capítulo se han obtenido con un HP-725 75 MHz.

muestra los resultados del ejemplo anterior con un umbral de 0.55 y valores diferentes para el número de nodos considerados en la búsqueda del mejor *clique* máximo, $K$. En la figura se puede observar cómo el tiempo de computación crece exponencialmente respecto a $K$.

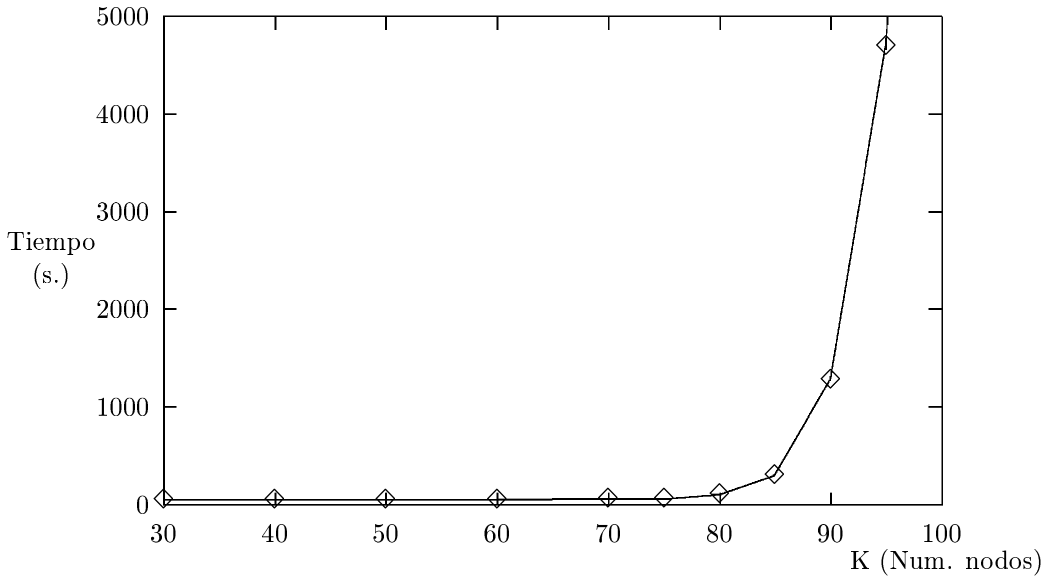



**Figura 3.17:** Crecimiento del tiempo de ejecución respecto al número de nodos considerados para encontrar el mejor *clique* máximo.

La selección de $K$ se puede realizar en función de esta gráfica, que muestra que a partir de $K = 75$ el tiempo computacional se dispara. Sin embargo, para evaluar la aproximación a la solución que se obtiene con el algoritmo subóptimo, se han utilizado distintos valores de $K$. En particular, los resultados que se muestran en la tabla 3.2 y en las distintas figuras han sido obtenidos con $K = 30$. La figura 3.18 muestra los mapas de disparidades resultantes de aplicar ambos algoritmos para encontrar la solución, el mejor *clique* máximo y el algoritmo subóptimo. Nótese que son bastante similares, por tanto, el cálculo del mejor *clique* máximo puede sustituirse por la solución aproximada propuesta.

### 3.6.3 Evaluación de la influencia del ruido

Para poder evaluar la influencia del ruido en el algoritmo propuesto hemos utilizado pares estéreo con ruido blanco de diferente varianza. Por ejemplo, en la figura 3.19 se muestran los resultados para un par de imágenes estéreo sin ruido, y otro con ruido de varianza 100. Los mapas de disparidades resultantes son similares, en el sentido de que proporcionan una aproximación a la solución, que se muestra en la figura 3.9. Aunque los resultados no son precisos, ambos mapas dan una idea de la posición de los objetos en la escena.

(a) Óptimo (86.7%). (b) Subóptimo (86.5%)

**Figura 3.18:** Mapas de disparidades y porcentajes de correspondencia del ejemplo *corridor* usando los algoritmos óptimo y subóptimo ($T_s = 0.80$).

(a) Sin ruido (97.2%) (b) Ruido de varianza 100 (96.3%)

**Figura 3.19:** Mapas de disparidades y porcentajes de correspondencia del ejemplo *corridor* con distintos niveles de ruido y el mismo umbral, $T_s = 0.40$ (algoritmo subóptimo).

La figura 3.20 muestra el porcentaje de correspondencia con respecto al umbral de similitud utilizado en las imágenes con distinto nivel de ruido. Nótese que cuanto mayor es el nivel de ruido en la imagen, el umbral de similitud necesario para obtener resultados similares en el porcentaje de correspondencia es menor. Las medidas de similitud entre regiones correspondientes descienden debido al ruido, de modo que el número de nodos del grafo decrece.

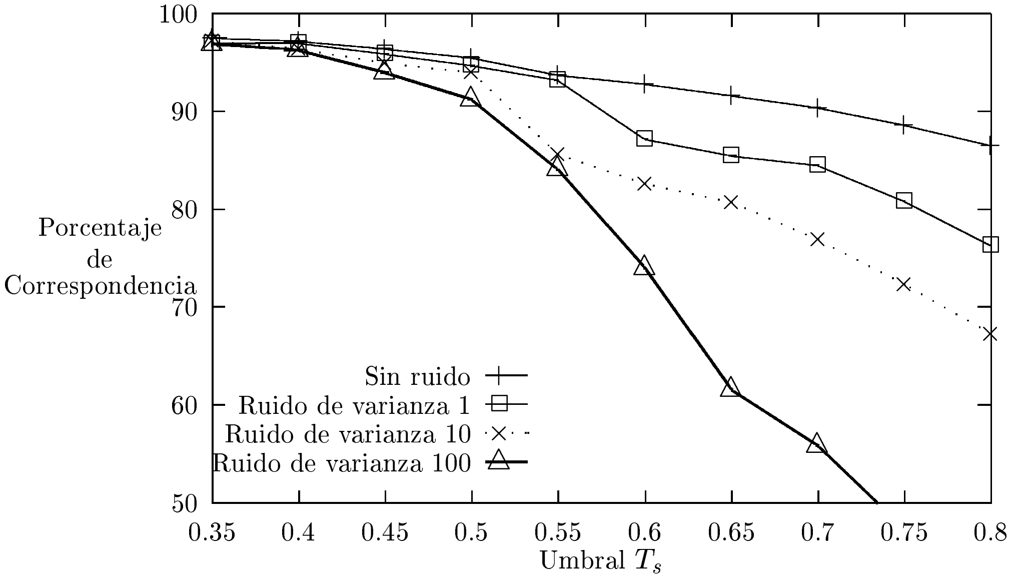



**Figura 3.20:** Efecto del ruido en el porcentaje de correspondencia.

La figura 3.21 muestra la relación entre el número de nodos en el grafo y el porcentaje de correspondencia obtenido para los ejemplos con ruido y sin ruido. Se puede observar que en el ejemplo con ruido se necesitan más nodos en el grafo para obtener el mismo porcentaje de correspondencias que en el ejemplo sin ruido. Esto significa que se incluyen más nodos en el grafo para resolver los problemas de segmentación que aparecen debidos al efecto del ruido.

En los ejemplos con ruido, es posible obtener porcentajes de correspondencia similares al ejemplo sin ruido decrementando el umbral de similitud. Es decir, el método propuesto puede conseguir resultados en presencia de ruido similares a los obtenidos con imágenes libres de ruido, lo que significa que el método es robusto en presencia del ruido.

## 3.7 Discusión y conclusiones

En este capítulo se ha presentado un método basado en grafos para la búsqueda de las correspondencias de regiones, que permite manejar los errores de la segmentación. La técnica consiste en una etapa de preproceso en el método clásico basado en grafos, para

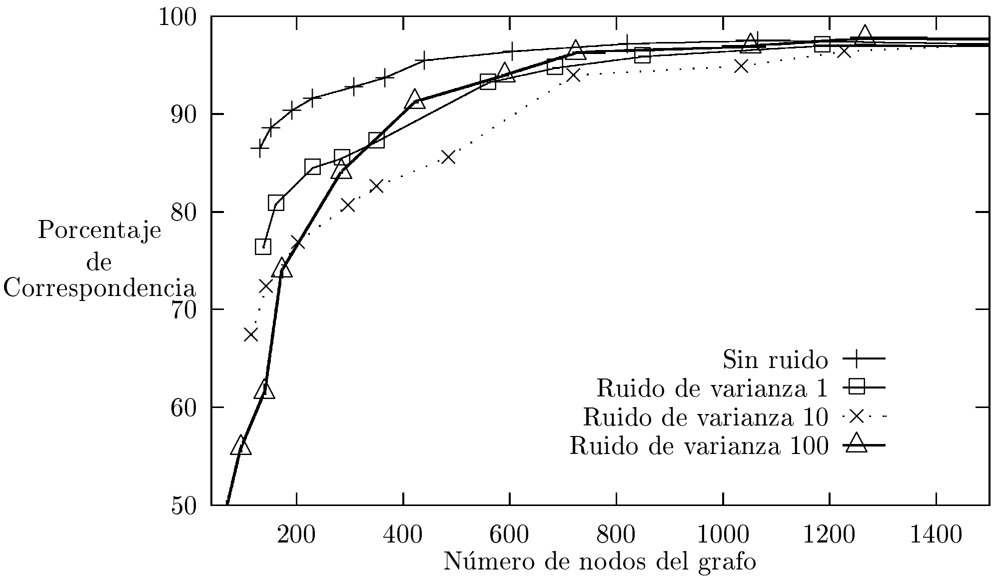



**Figura 3.21:** Relación entre el tamaño del grafo y el porcentaje de correspondencia con distintos niveles de ruido.

calcular las regiones unión apropiadas a considerar en el proceso de correspondencia. Este método permite evitar los errores que se deben a la segmentación considerando tanto las regiones individuales como las regiones unión obtenidas en el preproceso.

El único coste adicional con respecto al método que no considera los errores en la segmentación, es el cálculo de todos los *cliques* máximos en el grafo de asociación de incompatibilidades. Dado que el número de arcos es normalmente bajo, el proceso de búsqueda de todos los *cliques* máximos es bastante rápido, utilizando un algoritmo que aproveche este hecho.

El uso de dos criterios de similitud basados en técnicas de correlación, uno para crear nodos del grafo y otro para el resto del proceso de correspondencia, proporciona un menor número de ambigüedades, reduciendo así el tiempo de cálculo y produciendo correspondencias de regiones más fiables.

También se ha presentado un algoritmo subóptimo de búsqueda del mejor *clique* máximo de un grafo, que permite obtener una solución aproximada con un coste cuadrático respecto al tamaño del grafo.

Los experimentos llevados a cabo muestran que los resultados son satisfactorios con respecto al problema de la segmentación. La construcción de un grafo con las incompatibilidades entre correspondencias es una solución robusta para unir regiones y, en consecuencia, obtener mejores correspondencias.

Como desventaja principal de la correspondencia de regiones está el hecho de que el mapa de disparidades obtenido es un mapa disperso, que proporciona un solo valor de disparidad por cada región. Por lo tanto, las regiones que representan objetos de la escena con formas geométricas curvas o planos inclinados (por ejemplo, el plano del suelo, o una esfera) han sido asignados con una disparidad constante a lo largo de la región para obtener un mapa de disparidades denso. Estos mapas de disparidades podrían servir como inicialización a un proceso de correspondencia jerárquico que proporcione más resolución. Esta sería una línea de trabajo futuro, enfocada a integrar el método propuesto como parte de un algoritmo de correspondencia jerárquico.

# Capítulo 4

# Correspondencia de regiones basada en la minimización de una función de energía

Generalmente el proceso de reconstrucción se aborda después del proceso de correspondencia. La correspondencia proporciona un *mapa de disparidades*, lo más denso posible, que son fácilmente transformadas en profundidades mediante una simple triangulación. La reconstrucción procede a segmentar el *mapa de profundidades* (también denominado *imagen de rango*) para obtener una representación de mayor contenido semántico, como son superficies y objetos, que sean utilizables en múltiples aplicaciones. En este capítulo se presenta un método que realiza ambas etapas en un solo proceso, donde la correspondencia y la reconstrucción cooperan en la recuperación de la escena tridimensional.

## 4.1 Introducción

En el área de aplicación que nos incumbe (escenas de interiores) los objetos de la escena son principalmente objetos fabricados por el hombre, con poca textura. Por tanto, las técnicas basadas en áreas tienen dificultades en las zonas de intensidad homogénea, mientras que las técnicas basadas en primitivas proporcionan un mapa de disparidades disperso, de modo que, a medida que crece el nivel significativo de la primitiva, crece su dificultad de extracción, y crece la dispersión en los resultados. La combinación de métodos basados en características y métodos basados en áreas no es nueva en la visión estereoscópica, pero generalmente se había realizado como dos pasos diferenciados en el proceso de la correspondencia.

En el capítulo anterior, se proponía una combinación de características de alto nivel, como son las regiones, con una técnica de correlación en un mismo proceso para la búsqueda de la correspondencia. Sin embargo, a la hora de aplicar la correlación, las zonas de las regiones que no intersectan entre sí a una disparidad dada no se tenían en consideración, debido a los problemas generados por las diferencias que aparecen entre las segmentaciones de ambas imágenes.

Para evitar las dificultades generadas por las diferencias en la segmentación de las regiones, en este capítulo se propone un nuevo método que, basándose en la segmentación de una sola de las imágenes que se toma como referencia, busca las regiones correspondientes en la otra imagen sin segmentar. El objetivo es *encontrar para cada región de la imagen de referencia, la profundidad que proporcione una región correspondiente lo más parecida posible a la región de referencia.* La profundidad dentro de una región es una función que a priori podría variar bastante a lo largo y ancho de la región. Por tanto, será necesario imponer algún tipo de restricción a la forma de la función profundidad para poder resolver el problema planteado, como se verá más adelante.

La función profundidad debe tener en cuenta dos cuestiones importantes relativas a la restricción de continuidad (ya comentadas en la sección 2.2.2):

- debido a las restricciones de continuidad y de cohesión de la materia, la profundidad debe cambiar suavemente en las zonas homogéneas, y
- las discontinuidades en la profundidad pueden aparecer en donde existen discontinuidades en la intensidad.

Teniendo en cuenta la restricción de continuidad y la existencia de discontinuidades, restringiremos la forma de la función profundidad en el interior de cada región (zonas homogéneas), mientras que en las fronteras (discontinuidades en la intensidad) no se realiza ninguna restricción para permitir la aparición de discontinuidades en la profundidad.

### 4.1.1 Planteamiento del problema

Sean $I_1$ e $I_2$ un par de imágenes estereoscópicas, donde $I_1$ es la imagen que tomamos como referencia, y supongamos que conocemos los parámetros de calibración de ambas proyecciones. Si se conociese la profundidad de todos los pixeles de $I_1$, es decir, dispusiésemos del mapa de profundidades, $\mathcal{Z}$, podríamos obtener una réplica de $I_1$ a partir del mapa $\mathcal{Z}$ y los valores de intensidad de $I_2$. La réplica, que llamaremos $T$, es una imagen que contendría para cada pixel $m$ de $I_1$ la intensidad de su pixel correspondiente $m'$ en $I_2$.

$$T(m) = I_2(m'), \quad \forall m \in I_1 \tag{4.1}$$

$$m' = f_{12}(m, \mathcal{Z}(m)) \tag{4.2}$$

donde $m'$ es el pixel que corresponde a $m$ dada la profundidad de $m$, $\mathcal{Z}(m)$. Para una revisión detallada de cómo obtener el punto correspondiente a un pixel dado a partir de su profundidad, véase el apéndice A.4 (ecuación A.33).

Si $\mathcal{Z}$ es el mapa de profundidades verdadero (*ground truth depth map*), en condiciones ideales, $T$ sería igual a $I_1$, excepto en aquellos pixeles de $I_1$ cuyo pixel correspondiente esté oculto en $I_2$, para los que no existe correspondencia. Por tanto, $T$ es la réplica más similar a $I_1$ que se puede obtener a partir de $I_2$ y $\mathcal{Z}$. Cuando $\mathcal{Z}$ no coincide exactamente con el mapa de profundidades verdadero, pero no está lejos del mismo, se podría calcular mediante algún método variacional los incrementos o decrementos de $\mathcal{Z}(m_i)$ necesarios, de modo que $\mathcal{Z}$ se acerque hacia la solución. Por tanto, el problema de la correspondencia se puede abordar como la búsqueda de $\mathcal{Z}$ que minimice las diferencias entre $I_1$ y $T$. Sin embargo, éste es un problema *mal planteado*[1] (*ill-posed problem*) de modo que la existencia, unicidad y estabilidad de la solución no puede garantizarse en ausencia de restricciones adicionales [Terzopoulos, 1986].

Se puede encontrar un método variacional para la reconstrucción estereoscópica en [Robert y Deriche, 1996], que consiste en minimizar una función de energía basada en diferencias entre propiedades de los pixeles correspondientes. La función de energía incluye un término de regularización necesario para alcanzar la convergencia, suavizando la función de la profundidad ($\mathcal{Z}(u, v)$) en las áreas donde la intensidad es homogénea, pero al mismo tiempo permitiendo que aparezcan discontinuidades en la profundidades donde existan discontinuidades en la intensidad. En su caso, la forma de la función de profundidad se restringe mediante este término de regularización que debe cumplir los dos requisitos: suavizar la profundidad en las áreas de intensidad homogénea, no suavizar la profundidad a través de los cambios bruscos de intensidad, y sí a lo largo de ellos. Este método, aunque novedoso e interesante, no obtiene buenos resultados con escenas reales, y en el caso de escenas sintéticas necesita de al menos tres imágenes de la escena para obtener las discontinuidades en la función profundidad.

En esta tesis se propone usar regiones como primitiva de la correspondencia en lugar de pixeles de la imagen para poder aplicar el método de minimización en escenas donde los objetos estén poco texturados, y detectar las oclusiones y discontinuidades en la profundidad. Sin embargo, como ya se ha mencionado, este problema requiere imponer alguna restricción a la solución. En concreto, se propone imponer una restricción sobre la profundidad de los pixeles que pertenecen a una misma región. Por ejemplo, asumir una escena hecha de superficies planas es aplicable a entornos de robots donde las escenas están generalmente compuestas por objetos fabricados por el hombre. Aplicar esta restricción sobre la distribución de la profundidad de cada región hace que no necesitemos a priori ningún término de regularización adicional sobre la función profundidad.

En el ejemplo de la figura 4.1 se muestra la proyección en dos planos imagen de una escena formada por dos superficies, $S_a$ y $S_b$. Supongamos que segmentamos la imagen $I_1$, obteniendo un conjunto de regiones, $\{R_a, R_b\}$, y que buscamos su correspondencia en $I_2$. En la figura se muestra cuál sería la región correspondiente de $R_b$ en función de su profundidad.

[1]Un problema está *bien planteado* (*well-posed*) según Hadamard si hay: existencia de la solución, unicidad de la solución y estabilidad o continuidad de la solución. La estabilidad o continuidad de la solución se refiere a que hayan pequeñas variaciones de las medidas respecto al ruido de las observaciones.

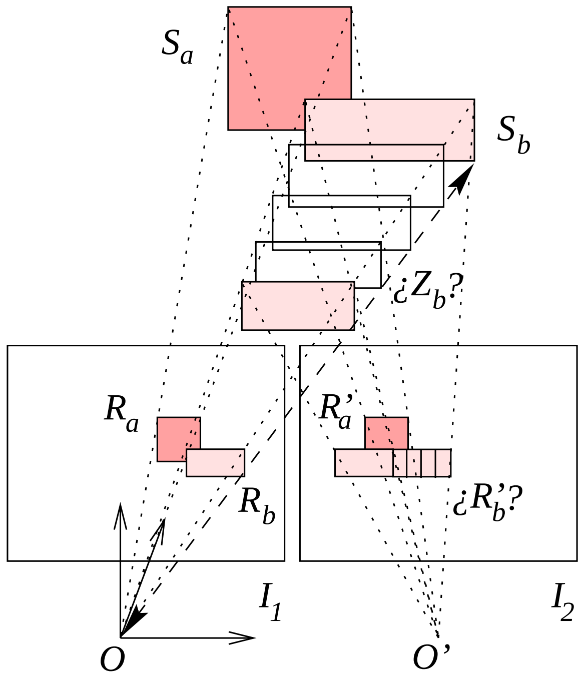



**Figura 4.1:** Planteamiento del problema de la correspondencia.

Las técnicas clásicas consisten en segmentar la imagen $I_2$ y calcular la correspondencia entre las regiones de la izquierda y de la derecha basándose en sus características geométricas o en sus características fotométricas [Marapane y Trivedi, 1989] [Marapane y Trivedi, 1994] [Cohen *et al.*, 1989b] [Sander *et al.*, 1989] [de Agapito, 1996]. Otros trabajos realizan la segmentación de $I_1$ e $I_2$ a la vez que la correspondencia mediante un algoritmo cooperativo que intenta evitar los errores de segmentación [Cohen *et al.*, 1989a] [Randriamasy y Gagalowicz, 1991].

Sin embargo, nuestro planteamiento es muy diferente: proponemos empezar con una estimación de la profundidad de cada región, $\{Z_a, Z_b\}$, de modo que podamos calcular las regiones correspondientes $\{R'_a, R'_b\}$, para después modificar las profundidades hacia la solución. Para ilustrar esta idea, en la figura 4.1 se muestra cómo varía $R'_b$ en función de $Z_b$. En este ejemplo, si el error de similitud entre las regiones $R_b$ y $R'_b$ se puede disminuir, se modifica $Z_b$ (y por tanto $R'_b$) en esa dirección, hasta que se alcance un mínimo.

La función de energía que se propone puede tener múltiples mínimos locales. Para que el algoritmo iterativo que busca el mínimo global sea capaz de converger a la solución, es necesario que el valor inicial de la profundidad no se encuentre lejos de la misma. Una solución muy común en los sistemas de visión consiste en utilizar un esquema multiescala, que procese un conjunto de representaciones de la imagen de diferente resolución, desde la más burda hasta la más fina (*coarse-to-fine*), de modo que los resultados de un nivel de resolución se emplean para inicializar el nivel siguiente de la escala.

Al igual que la imagen de referencia se segmenta en un conjunto de regiones

$$I_1 = \bigcup_{i=1..N} R_i$$

el mapa de profundidades $\mathcal{Z}$ estará formado por un conjunto de funciones $Z(R_i)$, cada una de las cuales está definida sobre una región $R_i$ de la imagen.

$$\mathcal{Z} = \bigcup_{\forall R \subset I_1} Z(R)$$

La idea básica del método que se propone consiste en partir de una estimación inicial de la profundidad de cada región, $Z(R)$, que se modifica iterativamente mediante la minimización de una función de energía basada en diferencias de intensidad de los pixeles de la región $R$ con su región correspondiente. Para determinar la profundidad $Z(R)$ de cada región, será necesario realizar alguna restricción sobre la forma de esta función, lo que significa realizar alguna suposición sobre la forma de las superficies de la escena.

La restricción de asumir una escena formada por superficies planas ha sido utilizada habitualmente en las técnicas de reconstrucción basadas en la correspondencia de regiones [Vézien y Gagalowicz, 1991] [Tarel y Vézien, 1995] [Tarel, 1996] [Chabbi y Berger, 1996]. Sin embargo, estos trabajos realizan la reconstrucción en un paso posterior a la correspondencia, mientras que nosotros pretendemos realizar ambos procesos en un solo paso.

Un médodo que realiza la correspondencia y la reconstrucción en un solo paso es [Super y Klarquist, 1997] que, para cada punto de interés de la imagen, realiza una búsqueda en el espacio de superficies planas 3D. Cada superfice 3D hipotética se usa para definir un par de trozos de imagen con posiciones, formas, tamaños, orientaciones e intensidades consistentes con la hipótesis y entre sí. De este modo, se asegura que la comparación entre los dos trozos da un índice de correlación fiable de la similitud entre las intensidades. Dado que la deformación entre imágenes se tiene en cuenta explícitamente, permite la existencia de deformaciones grandes entre imágenes. Sin embargo, no se modelan las oclusiones ni la existencia de discontinuidades.

La reconstrucción basada en regiones se ha abordado en pocos trabajos con escaso éxito debido a los errores de segmentación. Chabbi *et al.* [1996] no consideran las propias regiones en la reconstrucción, sino sus fronteras, asumiendo que son coplanares en el espacio 3D. Vezien *et al.* [1991] y Tarel *et al.* [1995] [1996] obtienen la posición tridimensional de las superficies asumiendo que las superficies de la escena son planas y basándose en las características geométricas globales de las proyecciones de las superficies sobre las imágenes. Sin embargo, los pares de regiones que están mal segmentados no son válidos para estos métodos y se eliminan mediante un test de coherencia. Las regiones eliminadas suponen una pérdida importante en la reconstrucción de la escena tridimensional.

Un problema importante que hay que considerar durante la correspondencia es la detección de oclusiones, y su influencia en el establecimiento de la correspondencia. En otros estudios basados en regiones este aspecto se ha dejado aparte o se ha tratado durante la etapa de reconstrucción. En este trabajo se tratan las oclusiones durante la correspondencia, como se verá en el capítulo 5, en un proceso cooperativo que encuentra las profundidades y las oclusiones al mismo tiempo, y el resultado es la reconstrucción tridimensional de la escena.

### 4.1.2 Aportaciones

El método que se propone consiste en calcular la profundidad mediante la minimización de una función de energía basada en la correlación entre una región de la imagen de referencia y su región correspondiente en la otra imagen. A pesar de que se han establecido similitudes en ciertos aspectos con otros métodos, que han sido detalladas en la sección anterior, el método que se propone introduce aspectos novedosos calculando las profundidades a partir de las regiones producto de la segmentación de una imagen de referencia, introduciendo las restricciones de continuidad de la figura a la vez que permitiendo discontinuidades de la profundidad en los bordes.

Correspondencia y reconstrucción se realizan en un solo paso. La profundidad se calcula para cada región de modo que la profundidad de cada pixel se puede obtener siguiendo la restricción que se imponga. Las dos restricciones que se han investigado hasta el momento son:

1. La profundidad es constante para todos los pixeles de la región. Esta restricción supone que la escena está formada por planos paralelos al plano imagen de la imagen de referencia (*restricción de planos fronto-paralelos*).
2. La profundidad varía linealmente para todos los pixeles de la región. Esta restricción supone que la escena está formada por planos de cualquier tipo (*restricción de planos de cualquier orientación*).

Por tanto, se propone **mantener una representación de la escena tridimensional**, de modo que en todo momento se dispone de un mapa denso de profundidades, cuya precisión con respecto al mapa de profundidades verdadero dependerá de las oclusiones y la validez de la restricción asumida.

Junto con la búsqueda de la correspondencia, también se trata la **detección de oclusiones**, y su influencia en el establecimiento de la correspondencia. Como se verá más adelante, las oclusiones tienen un papel importante en la búsqueda de la correspondencia, y su correcta detección aumenta la precisión de los resultados.

En este capítulo se presenta el estudio, desarrollo y resultados experimentales de la primera suposición (restricción de planos fronto-paralelos) y un estudio de la posibilidad de generalización a la segunda suposición (restricción de planos de cualquier orientación).

En el siguiente capítulo se estudia cómo puede la detección de oclusiones mejorar las correspondencias obtenidas mediante este método.

En la sección 4.2 se detalla cómo se aplican las restricciones de los sistemas estereoscópicos al método de correspondencia que se presenta en este capítulo. En la sección 4.3, se explica la función de energía elegida, y las consideraciones básicas de la estrategia de minimización. En la sección 4.4, se trata el caso de escenas formadas por planos paralelos al plano imagen. A continuación, se muestran los resultados obtenidos en los numerosos experimentos realizados con pares de imágenes sintéticas y reales en la sección 4.5. En la sección 4.6 se estudia la generalización del método a la suposición de escenas formadas por superficies planas cualesquiera. Finalmente, las conclusiones sobre el trabajo mostrado en este capítulo, así como una discusión sobre los resultados experimentales obtenidos, se resumen en la sección 4.7.

## 4.2 Restricciones del proceso de correspondencia

### 4.2.1 Restricción epipolar

Como ya se ha comentado al principio de este capítulo, dado un pixel $m$ en la imagen de referencia, se puede calcular el pixel correspondiente en la segunda imagen, $m'$, a partir de los parámetros de calibración y la profundidad estimada de ese pixel,

$$m' = f_{12}(m, \mathcal{Z}(m)) \tag{4.3}$$

donde la función $f_{12}$ depende de las matrices de proyección de las imágenes $I_1$ e $I_2$, que contienen los parámetros intrínsecos y extrínsecos de calibración (apéndice A.4, ecuación A.33).

Dada una región $R$ en la imagen de referencia, se puede calcular su región correspondiente en la segunda imagen, $R'$, a partir de los parámetros de calibración y la profundidad estimada de la región,

$$R' = g_{12}(R, Z(R)) \tag{4.4}$$

donde la función $g_{12}$ no es más que una generalización de la función $f_{12}$, de modo que cada pixel $m_i \in R$ corresponde a cada punto $m'_i \in R'$ a una profundidad dada $\mathcal{Z}(m_i)$, y $\mathcal{Z}(m_i)$ se calcula a partir de la ecuación que determina la profundidad de la región, $Z(R)$, es decir, depende de la profundidad de la región según la restricción impuesta.

$$R' = g_{12}(R, Z(R)) = \{m'_i = f_{12}(m_i, \mathcal{Z}(m_i)), \forall m_i \in R\} \tag{4.5}$$

De esta manera se restringe la búsqueda de la correspondencia al área que determina la restricción epipolar. Queda por definir cuál es la relación entre las profundidades de los pixeles que pertenecen a una misma región, es decir, la forma de la función $Z(R)$.

### 4.2.2 Restricción de profundidad (o de disparidad)

En este método es fundamental proporcionar un rango de profundidades válido de la escena real. En el caso de que a una determinada escala no se tenga suficiente información sobre una región para obtener una aproximación de su profundidad (esto puede ocurrir sobre todo en los niveles de resolución más burda), el algoritmo podría diverger hacia la máxima o mínima profundidad. Este resultado se propaga en los siguientes niveles de la escala, donde al disponer de mayor información es posible obtener una mejor aproximación. El hecho de limitar el rango de profundidades no supone una pérdida importante de generalidad, debido a que, por lo general, se dispone de esta información sobre la escena, sobre todo cuando se trata de interiores.

Es una restricción muy utilizada, generalmente para acelerar el proceso de la correspondencia, ya que de este modo, la búsqueda de la correspondencia se restringe a un segmento de la línea epipolar, aquél que determine el rango de profundidades dado, $[Z_{min}, Z_{max}]$.

### 4.2.3 Restricción de continuidad de la figura

Las discontinuidades en la profundidad normalmente aparecen en las discontinuidades de la intensidad, mientras que las regiones de intensidad homogénea normalmente representan superficies suaves. El uso de regiones hace que introducir estas restricciones sea más sencillo. Por un lado, se permiten discontinuidades en la profundidad en los contornos de las regiones y, por otro lado, la continuidad de la profundidad a lo largo de la figura se modela explícitamente en la restricción sobre la forma de la profundidad dentro de la región.

### 4.2.4 Restricción de orden

La restricción de orden se asume al inicializar el primer mapa de profundidades a un valor constante y propagar los resultados de un nivel a otro. Aunque posteriormente no se impone que se siga cumpliendo la restricción de orden, es muy improbable que se viole esta condición si la segmentación de las imágenes de los distintos niveles de escala y el proceso de correspondencia funcionan con normalidad.

## 4.3 La función de energía

La técnica de minimización aborda el problema de la correspondencia como la minimización de una función de energía, y nosotros proponemos expresar esta energía en términos de correspondencia de regiones, basándonos en las regiones obtenidas de la segmentación de la imagen de referencia. Por tanto, la función de energía a minimizar se puede escribir como

$$E(\mathcal{Z}) = \int_{R\subset I_1} F(R, Z(R))dR \tag{4.6}$$

donde la función $F$ denota el error en similitud entre cada región $R$ de la imagen de referencia y su correspondiente $R'$ calculada a partir de la profundidad estimada de la región, $Z(R)$ (ecuación 4.5).

Una vez se ha obtenido $R'$ (la región correspondiente a $R$ según $Z(R)$), podemos calcular el error en la similitud entre ambas regiones utilizando alguna medida de correlación. Por ejemplo, el método de correlación ZNCC (*zero-mean normalized cross correlation*) es una medida a maximizar, de modo que podemos minimizar la siguiente función de error:

$$F(R, Z(R)) = -C_{ZNCC}(R, R') = -\frac{1}{N}\sum_{\forall m_i\in R}\mathcal{I}_1(m_i)\mathcal{I}_2(m_i') \tag{4.7}$$

donde $m_i' = f_{12}(m_i, \mathcal{Z}(m_i))$, $N$ es el tamaño de $R$ y $\mathcal{I}_1(m_i)$, $\mathcal{I}_2(m_i')$ son las intensidades normalizadas de media cero de los pixeles $m_i$ y $m_i'$ con respecto a su región, respectivamente:

$$\mathcal{I}_k(m_i) = \frac{I_k(m_i) - \mu_k(R)}{\sigma_k(R)}, \quad m_i \in R \subset I_k \tag{4.8}$$

donde $\mu_k(R)$ y $\sigma_k(R)$ son la media y la desviación típica de la intensidad de la región $R$.

Sustituyendo la ecuación 4.7 en la función de energía general (ecuación 4.6) resulta

$$E(\mathcal{Z}) = \int_{R\subset I_1} -\frac{1}{N}\sum_{\forall m_i\in R}\mathcal{I}_1(m_i)\mathcal{I}_2(m_i')\ dR \tag{4.9}$$

donde $m_i' = f_{12}(m_i, \mathcal{Z}(m_i))$, y $\mathcal{Z}$ es tal que, debido a la restricción de continuidad, los pixeles correspondientes $m_i', m_j'$ a dos pixeles $m_i, m_j$ de la misma región están relacionados entre sí.

Esta función se podría reescribir en función de las dos variables independientes de la imagen, $m = (u, v)$, y el mapa de profundidades $\mathcal{Z}$, que es el conjunto de todas las profundidades y es lo que se quiere obtener,

$$E(\mathcal{Z}) = \int_u\int_v -\frac{1}{N}\mathcal{I}_1(u,v)\mathcal{I}_2(f_{12}((u,v), \mathcal{Z}(u,v)))\ du\ dv \tag{4.10}$$

con la condición de que las profundidades de todos los pixeles que pertenecen a la misma región cumplan la restricción impuesta. De modo que, la función se puede reescribir como

$$E(\mathcal{Z}) = \int_u\int_v F(u, v, z, \frac{\partial z}{\partial u}, \frac{\partial z}{\partial v})\ du\ dv \tag{4.11}$$

donde $z = \mathcal{Z}(u, v)$ y $(\frac{\partial z}{\partial u}, \frac{\partial z}{\partial v})$ representa la variación de la profundidad respecto a $u$ y $v$, que ha de ser pequeña en las zonas de intensidad homogénea.

De acuerdo con la ecuación de Euler [Elsgoltz, 1977], la $z$ que hace mínima $E(\mathcal{Z})$ es solución de la ecuación diferencial de segundo orden,

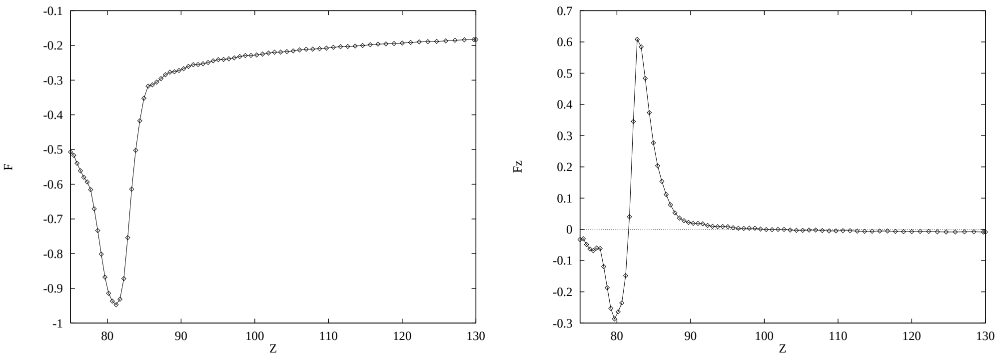
$$F_z - \frac{\partial}{\partial u}\{F_p\} - \frac{\partial}{\partial v}\{F_q\} = 0 \tag{4.12}$$

donde $p = \frac{\partial z}{\partial u}$, $q = \frac{\partial z}{\partial v}$ y $F_z, F_p, F_q$ son las derivadas parciales de $F$ con respecto a $z, p$ y $q$, respectivamente.

Sin embargo, bajo la suposición de que las superficies de la imagen son planas, $p$ y $q$ son constantes en el interior de las regiones (0 en el caso de planos fronto-paralelos), mientras que en las fronteras de las regiones no se aplica ninguna restricción, para permitir que aparezcan discontinuidades en la función $\mathcal{Z}$. Por tanto, bajo la consideración de superficies planas, $F$ no depende de $p$ y $q$, de modo que $F_p = 0$ y $F_q = 0$, y por tanto, $z = Z(R)$ es solución de

$$F_z = 0 \tag{4.13}$$

En la figura 4.2(a) se muestra un ejemplo de cómo varía el error de correlación $F$ entre una región y su región correspondiente en función de su profundidad, $Z$, suponiendo que ésta sea constante en toda la región. En la figura 4.2(b) se muestra la derivada de esta función, $F_z$. Se puede observar que el punto donde el error de correlación alcanza un mínimo, la función $F_z$ pasa por cero, de negativo a positivo. El mínimo global representa la profundidad a la que se encuentra la superficie representada por la región ejemplo y su región correspondiente.




(a) $F$ en función de $Z$ (b) $F_z$ en función de $Z$

**Figura 4.2:** Error de similitud entre una región y su región correspondiente en función de la profundidad.

En la figura se puede observar que para cada paso por cero en $F_z$ correspondiente a un mínimo de la función $F$, existe una zona de $F_z$ negativa a la izquierda y otra positiva a la derecha. Esto nos permite, dada una $Z$ inicial estimada, decidir si $Z$ se debe incrementar o decrementar para alcanzar el paso por cero.

Para cada valor de $Z$ que queramos calcular, se puede diseñar un algoritmo iterativo que incremente o decremente $Z$ en función de $F_z$:

$$Z^{t+1} = Z^t - \Delta t F_z \tag{4.14}$$

Por lo tanto, el problema se podría resolver mediante un algoritmo por descenso de gradiente. Sin embargo, puede haber más de un paso por cero debido a la existencia de más de un mínimo local. Para obtener el mínimo global, la $Z$ inicial a incrementar/decrementar, debe estar suficientemente cerca del paso por cero correspondiente al mismo. Por esta razón, la profundidad se debe estimar a distintas resoluciones, desde la más burda a la más fina, para dirigir la convergencia del algoritmo al mínimo global.

### 4.3.1 Diferencia con las técnicas clásicas basadas en áreas

Esta técnica se podría asimilar a una técnica de ventana adaptativa, donde el tamaño y forma de la ventana depende de los resultados de la segmentación. Sin embargo, si llamamos $W_1$ a la ventana de la imagen de referencia, y localizamos en $I_2$ los puntos correspondientes a cada pixel perteneciente a $W_1$, obtenemos la ventana correspondiente, $W_2$, que puede ser de *distinto tamaño y forma* que $W_1$.

Un problema que aparece entonces es el de calcular la correlación entre dos ventanas de tamaños y formas diferentes, $W_1$ y $W_2$. En las figuras 4.3 y 4.4 se muestra una región de una imagen en la izquierda, que es la proyección de una determinada superficie tridimensional, $S$. Cada pixel de la región tiene un tamaño de $k_u \times k_v$ medido en las unidades espaciales correspondientes. En la parte derecha, donde se muestra la proyección de la superficie $S$ en la otra imagen, se puede observar que el área que corresponde a cada pixel podría tener medidas diferentes de $k_u$ y $k_v$. Existen dos posibles soluciones a este problema:

- Calcular la correlación mediante una integral de la superficie, considerando que una imagen es una función de dos variables constante sobre la superficie de cada pixel (constante a tramos) (figura 4.3).
- Calcular la intensidad de los puntos de la ventana $W_2$ (que corresponden a las coordenadas enteras de la ventana $W_1$) por interpolación lineal entre los valores de intensidad de los puntos de la segunda imagen y calcular la correlación entre dos ventanas iguales como en las técnicas clásicas de correlación (figura 4.4).

Nosotros hemos escogido el segundo método por ser el más simple y el más rápido de calcular. De hecho, la técnica habitual va a consistir en obtener una imagen transformada $T$ a partir de la segunda imagen, $I_2$, según la ecuación 4.1. Tras esta transformación, las dos regiones correspondientes $R$ y $R'$ se encuentran en las mismas posiciones de las imágenes $I_1$ y $T$, respectivamente. Esta transformación es aplicable a todos los pixeles de $I_1$ exceptuando aquellos cuyos pixeles correspondientes no aparezcan en $I_2$, por ejemplo, por estar ocultos.

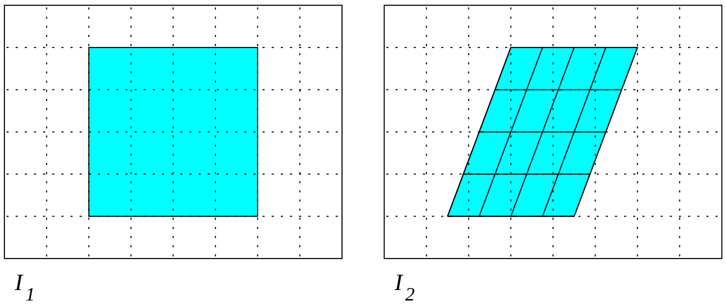



**Figura 4.3:** Cálculo de la correlación: integración sobre la superficie.

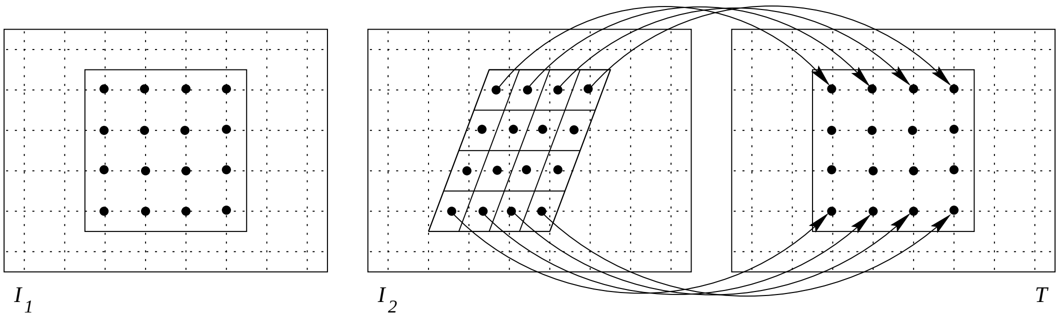



**Figura 4.4:** Cálculo de la correlación: interpolación entre los valores de los pixeles.

## 4.4 Escenas con superficies planas fronto-paralelas

### 4.4.1 Restricción de profundidad constante dentro de una región

Bajo esta restricción, la profundidad de cada región, $Z(R)$, se puede representar mediante un escalar. Por tanto, la región correspondiente a una región cualquiera $R$ es:

$$R' = g_{12}(R, Z(R)) = \{m' = f_{12}(m, z), \forall m \in R, z = Z(R) \in \mathbb{R}\} \tag{4.15}$$

Por ejemplo, si se asume un sistema de geometría paralela donde solamente se ha desplazado una cámara respecto de la otra en el eje $X_c$, el contorno de $R'$ es exactamente de la misma forma y tamaño que el contorno de $R$, y por tanto para hallar $R'$ solo sería necesario calcular el desplazamiento respecto de $R$, es decir, la disparidad.

En definitiva, el mapa de profundidades se puede representar en un vector donde cada valor determina la profundidad de cada región de $I_1$.

### 4.4.2 Cálculo de la derivada de la función de energía

De las ecuaciones 4.7 y 4.13 obtenemos que para cada región $R$ hemos de encontrar $Z(R)$ tal que

$$F_z(R, Z(R)) = \frac{\partial}{\partial z}\left\{-\frac{1}{N}\sum_{\forall m_i \in R} \mathcal{I}_1(m_i)\mathcal{I}_2(m_i')\right\} = 0 \tag{4.16}$$

Debido a que $R$ es independiente de la variación de $z = Z(R)$, la intensidad de cada pixel de $R$, $m_i$, la intensidad media de $R$, su desviación típica, y por consiguiente, $\mathcal{I}_1(m_i)$, son constantes respecto a $z$, y por tanto

$$F_z(R, Z(R)) = -\frac{1}{N}\sum_{\forall m_i \in R} \mathcal{I}_1(m_i)\frac{\partial}{\partial z}\left\{\mathcal{I}_2(m_i')\right\} = 0 \tag{4.17}$$

La región correspondiente, $R'$, sí que varía respecto a $z$, y en concreto, todos los pixeles de $R'$, $m_i'$. En consecuencia, varía también la intensidad de los $m_i'$, su media y su desviación típica. El cálculo desarrollado paso a paso de las derivadas parciales de estas funciones respecto a $z$ se detalla en el apéndice 4.A con el fin de no romper la continuidad de la exposición del método.

Finalmente, se obtiene (ecuación 4.63)

$$F_z(R, Z(R)) = \frac{1}{N}\sum_{m_i \in R} \mathcal{I}_1\left(m_i\right)\left(\mathcal{I}_2\left(m_i'\right)\mathcal{H}\left(R'\right) - I_2'(m_i')\frac{\partial m_i'}{\partial z}\right) = 0 \tag{4.18}$$

donde a la función $\mathcal{H}(R')$ la llamaremos el *gradiente medio ponderado* de la región $R'$ (ecuación 4.61) y es

$$\mathcal{H}(R') = \frac{1}{N}\sum_{m_i' \in R'} \mathcal{I}_2(m_i')I_2'(m_i')\frac{\partial m_i'}{\partial z} \tag{4.19}$$

En ambas ecuaciones aparece la expresión $I_2'(m_i')\frac{\partial m_i'}{\partial z}$, que indica la primera derivada de la intensidad (gradiente de intensidad) en la dirección de la línea epipolar.

### 4.4.3 Análisis de la derivada de la función de energía

Para analizar el significado de la ecuación 4.18, vamos a desarrollar la ecuación como sigue,

$$\frac{1}{N}\sum_{m_i \in R} \mathcal{I}_1\left(m_i\right)\mathcal{I}_2\left(m_i'\right)\mathcal{H}\left(R'\right) - \frac{1}{N}\sum_{m_i \in R} \mathcal{I}_1\left(m_i\right)I_2'(m_i')\frac{\partial m_i'}{\partial z} = 0 \tag{4.20}$$

El primer término de la ecuación es proporcional al índice de correlación (ecuación 4.7), mientras que el segundo término es muy similar a la definición de $\mathcal{H}$.

Vamos a definir una función similar a $\mathcal{H}$, que relacione intensidades en la imagen de referencia con gradientes direccionales en la otra imagen, es decir,

$$\mathcal{H}_{12}(R, R') = \frac{1}{N} \sum_{m_i=1} \mathcal{I}_1(m_i) I'_2(m'_i) \frac{\partial m'_i}{\partial z} \tag{4.21}$$

Podemos reescribir la ecuación 4.18 como

$$C\left(R, R'\right) \mathcal{H}\left(R'\right) - \mathcal{H}_{12}\left(R, R'\right) = 0 \tag{4.22}$$

Tanto $\mathcal{H}$ como $\mathcal{H}_{12}$ dependen del gradiente de intensidad de todos los pixeles de la región en la dirección de la línea epipolar, y pueden considerarse medidas similares. Ambas realizan una comparación entre la intensidad normalizada de media cero y el gradiente direccional de intensidad, para todos los pixeles de la región. La diferencia estriba en que $\mathcal{H}$ compara intensidad y gradiente en $R'$, mientras que $\mathcal{H}_{12}$ compara intensidad en $R$ con gradiente en $R'$.

La correlación entre las intensidades de ambas regiones, $C$, varía en el rango $[-1, 1]$. Si $R'$ es la region correspondiente a $R$, en condiciones ideales, $C = 1$ y $\mathcal{H} = \mathcal{H}_{12}$, de modo que $C\mathcal{H} - \mathcal{H}_{12} = 0$ y la profundidad no debe incrementarse ni decrementarse: el algoritmo iterativo se detiene cuando se alcanza la profundidad solución. En otro caso, $C\mathcal{H} - \mathcal{H}_{12} \neq 0$ nos proporciona una medida de cómo debe modificarse la profundidad para acercarse hacia la solución. Si $C\mathcal{H} - \mathcal{H}_{12} > 0$ hay que decrementar $Z$, mientras que si $C\mathcal{H} - \mathcal{H}_{12} < 0$ hay que incrementarla. Es decir, la solución está en el *paso por cero* de $C\mathcal{H} - \mathcal{H}_{12}$.

### Caso particular de geometría paralela

En el caso de geometría paralela y bajo la suposición de superficies fronto-paralelas, la profundidad es constante en toda la región y por tanto podemos extraer $\frac{\partial m'_i}{\partial z}$ como factor común, de modo que podríamos reescribir la ecuación 4.18 como

$$F_z = \frac{-\alpha'_u b}{z^2} \left(C\mathcal{H}' - \mathcal{H}'_{12}\right) = 0 \tag{4.23}$$

donde $\mathcal{H}'$ y $\mathcal{H}'_{12}$ son similares a $\mathcal{H}$ y $\mathcal{H}_{12}$ con la correspondiente modificación. El gradiente de intensidad se calcula en este caso en la dirección $u$ (que es la dirección de las líneas epipolares).

Si el incremento de la profundidad se calcula proporcional respecto a $F_z$,

$$Z^{t+1} = Z^t - \Delta t F_z \tag{4.24}$$

se puede observar que cuanto mayor es $z = Z(R)$, menor es el incremento del gradiente, mientras que cuanto menor es $z$, mayor es el incremento. Este comportamiento no es deseable, debido al hecho de que el desplazamiento de un objeto de una imagen a otra (disparidad) es mayor cuanto más cerca está, y viceversa. Por tanto, lo más razonable sería realizar incrementos más pequeños de $Z$ cuanto más cerca se encuentre el objeto.

Por otro lado, hay que evitar que $\Delta Z$ produzca desplazamientos de $m'_i$ mayores que 1 pixel, para evitar que la región correspondiente, $R'$, se desplace mucho de una iteración a otra, ya que esto podría hacer que se sobrepase el paso por cero y se alcance un paso por cero diferente. Dada la profundidad mínima de los objetos de la escena, podríamos calcular un factor $\Delta t$ de modo que el mayor valor de $|F_z|$ posible produzca un incremento de 1 pixel como máximo. Sin embargo, esto prodría producir desplazamientos muy pequeños cuando $Z$ es grande, de modo que la convergencia en las regiones correspondientes a objetos lejanos de la escena sería muy lenta.

Debido a que la información de que partimos son las imágenes, proponemos calcular los incrementos de profundidad a partir del incremento de disparidad deseado. Es decir, obtener la magnitud de $\Delta Z$ a partir de un incremento de disparidad predeterminado, y la dirección del incremento (el signo de $\Delta Z$) a partir de $F_z$. Por tanto, el factor $\frac{\partial m'_i}{\partial z}$ se puede simplificar ya que la magnitud de $F_z$ no determina la magnitud de $\Delta Z$.

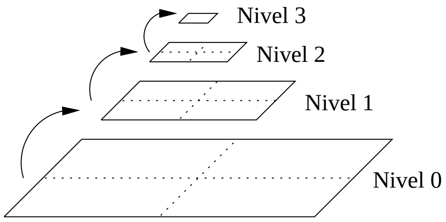
Aunque en el caso de geometría no paralela no se puede sacar $\frac{\partial m'_i}{\partial z}$ como factor común, ni tampoco simplificarlo, sí que se puede extender el resto del razonamiento en cuanto a la conveniencia de fijar el incremento de profundidad en función del incremento de disparidad deseado.

### 4.4.4 El algoritmo de correspondencia

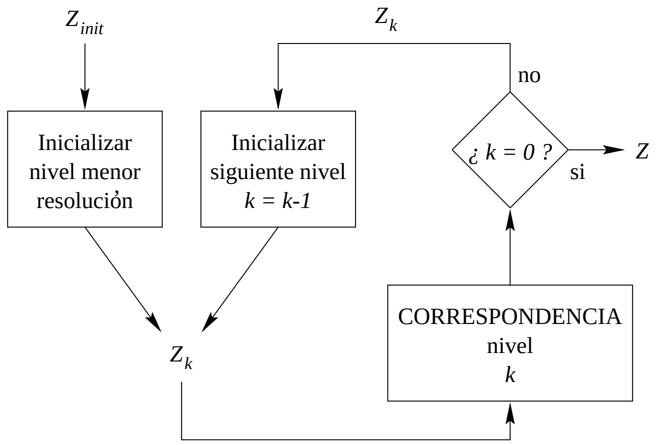
El objetivo es calcular un incremento o decremento pequeño en la profundidad en cada paso del algoritmo iterativo para acercar $Z$ suavemente hacia la solución. Para evitar obtener un mínimo local en vez de la solución global se aplica el algoritmo en un esquema multiescala. El conjunto de los niveles de la escala para cada imagen forma una estructura piramidal de modo que el nivel de menor resolución está formado por un número muy pequeño de pixeles (por ejemplo, $2 \times 2$, $3 \times 5$, etc, dependiendo del tamaño inicial de la imagen). El algoritmo converge a una solución en cada nivel, que se utiliza para inicializar el siguiente nivel. De este modo, la solución se guía desde los niveles de menor resolución hacia un mínimo global, que se va refinando a medida que aumenta la resolución.

El algoritmo general multiescala se muestra en la figura 4.5, donde se puede observar que hay tres operaciones principales:

- Inicializar las profundidades del nivel de menor resolución, que consiste en asignar un mismo valor de profundidad a todas las regiones de ese nivel.




(a) Estructura piramidal




(b) Algoritmo multinivel

**Figura 4.5:** Esquema multiescala: algoritmo general.

- Inicializar las profundidades de los demás niveles de resolución, a partir de las profundidades del nivel anterior resultantes de la correspondencia. Por ejemplo, se puede crear el mapa de disparidades del nivel $k$, escalarlo al siguiente nivel $k-1$, y calcular la profundidad de cada región haciendo una media de las profundidades de cada pixel de la región.

- Establecer la correspondencia en cada nivel $k$, dada una inicialización de $Z$ y dar como resultado la $Z$ que minimice la función de energía. Este proceso, que consiste en incrementar/decrementar la profundidad iterativamente hasta alcanzar un paso por cero en $F_z$, se detalla en el algoritmo 4.1.

Como ya se ha comentado, una cuestión importante del algoritmo es la cantidad del incremento/decremento de profundidad, $\Delta Z$. Por un lado, hay que evitar que $\Delta Z$ produzca desplazamientos de $m'_i$ mayores que 1 pixel, para evitar que la región correspondiente, $R'$, se desplace mucho de una iteración a otra ya que esto podría hacer que se sobrepase el paso por cero y se alcance un paso por cero diferente. Como cada $m'_i = (u', v')$ se calcula a partir de $m_i = (u, v)$ y la profundidad actual estimada, $Z^t$, calculamos la nueva profundidad, $Z^{t+1}$, a partir del incremento $\Delta u'$ o $\Delta v'$ deseado. En los algoritmos utilizaremos $\Delta d$ para representar este incremento de manera general.

Cuando $C\mathcal{H} - \mathcal{H}_{12} > 0$ la profundidad debe decrementarse y viceversa. Si $Z$ se ha decrementado iterativamente para hacer decrecer $C\mathcal{H} - \mathcal{H}_{12}$ hacia 0 y obtenemos un valor negativo, el último decremento debe deshacerse y repetirlo con un $\Delta d$ más pequeño, para obtener mayor precisión en la solución. La idea es utilizar incrementos fijos de $\Delta d$ hasta acercarse al paso por cero, y a continuación utilizar decrementos más pequeños hasta que se obtiene la solución con la precisión requerida. Por lo tanto, el algoritmo tiene dos etapas, que a partir de ahora llamaremos de *acercamiento*, y de *refinamiento*.

En el algoritmo 4.1, los nuevos incrementos de profundidad son calculados mediante la función $IncZ$ a partir del $\Delta d$ actual, que también controla que no se sobrepasen los límites de la profundidad, $Z_{min}$ y $Z_{max}$. El algoritmo se detiene cuando ocurre alguna de estas condiciones:

- $F_z = 0$, que es la condición principal.
- $Z$ converge a un valor, es decir, después de incrementar $Z$, su valor no ha cambiado. Esto quiere decir que se está intentando incrementar $Z$ más allá del valor máximo, o por debajo del valor mínimo. Este es un caso que puede ocurrir, sobre todo en los niveles altos de la estructura multiescala.
- Durante la etapa de refinamiento, $|\Delta d|$ se ha reducido sucesivamente, pero no se obtiene $F_z = 0$. Esto quiere decir que se está muy cerca del paso por cero, pero es muy difícil obtenerlo con total precisión. En lugar de ponerle un límite en la precisión a $F_z$, se lo ponemos a $\Delta d$, y lo llamaremos $\Delta d_\epsilon$, de modo que cuando $|\Delta d| < \Delta d_\epsilon$ se calcula una aproximación a la solución mediante interpolación entre las profundidades a ambos lados del paso por cero.

### Incremento de profundidad

En el apéndice 4.B se desarrolla la relación entre los incrementos de disparidad y los incrementos de profundidad. Dado un incremento de profundidad, $\Delta Z$, en el caso general, el incremento de $u'$ que se produce (para $v'$ es análogo) es:

$$\Delta u' = \alpha'_u \frac{(at'_z - bt'_x)\,\Delta Z}{(bZ + t'_z)\,(a(Z + \Delta Z) + t'_z)} \tag{4.25}$$

donde $a = k_{11}u + k_{12}v + k_{13}$, $b = k_{31}u + k_{32}v + k_{33}$ y $Z$ es el valor de profundidad antes de realizar el incremento.

Si fijamos $\Delta u'$, entonces podemos calcular el $\Delta Z$ adecuado para cada caso,

$$\Delta Z = \frac{\Delta u'\,(bZ + t'_z)^2}{\alpha'_u\,(at'_z - bt'_x) - \Delta u' b\,(bZ + t'_z)} \tag{4.26}$$

Análogamente, se podría fijar $\Delta v'$, o seleccionar aquél de los dos cuyo incremento sea mayor ante una geometría dada.

La desventaja de calcular $\Delta Z$ basándonos en un $\Delta u'$ prefijado es que $a$ y $b$ dependen del pixel que se esté procesando, $(u, v)$, y por tanto, habría que determinar un pixel $(u, v)$ de la región sobre el que realizar el cálculo. Este pixel podría ser aquél que proporcione el incremento mayor para asegurarnos de que no se sobrepasa el $\Delta u'$ fijado, o algún punto significativo, como por ejemplo, el centroide.

---

**Algoritmo 4.1 CORRESPONDENCIA1**, Cálculo de $Z$ del nivel $k$, sin detección de oclusiones.

---

**Entradas:** $Z_{min}, Z_{max}, \Delta d_{ini}, \Delta d_{\epsilon}$, nivel $k$ de $I_1$ e $I_2$, y $Z$ inicial del nivel $k$.
**Salidas:** $Z$ final del nivel $k$.

Sea $\mathcal{R}$ el conjunto de regiones resultante de la segmentación de $I_1$.
**para cada** región $R \in \mathcal{R}$ **hacer**
  $Z^0 \leftarrow Z(R)$
  $F_z^0 \leftarrow C(R, R')\,\mathcal{H}(R') - \mathcal{H}_{12}(R, R')$, donde $R' = g_{12}(R, Z^0)$
  $\Delta d \leftarrow \text{sign}(F_z^0)\Delta d_{ini}$
  $t \leftarrow 0$
  **mientras** $(F_z^t \neq 0)$ y ($Z$ no converge) **hacer**
    $Z^{t+1} \leftarrow IncZ(Z^t, \Delta d,$ parámetros de calibración, $Z_{min}, Z_{max})$
    **si** $Z^{t+1} = Z^t$ **entonces**
      $Z$ converge
    **sino**
      $F_z^{t+1} \leftarrow C(R, R')\,\mathcal{H}(R') - \mathcal{H}_{12}(R, R')$ donde $R' = g_{12}(R, Z^{t+1})$
      **si** sign$(F_z^{t+1})$ =sign$(F_z^0)$ **entonces** { *signos iguales* $\Rightarrow$ *continuar* }
        $t \leftarrow t + 1$
      **sino** { *signos diferentes* $\Rightarrow$ *recalcular* $Z^{t+1}$ }
        **si** $|\Delta d| \geq \Delta d_{\epsilon}$ **entonces** { *refinamiento* $\Rightarrow$ *reducir incremento* }
          $\Delta d \leftarrow \Delta d/2$
        **sino** { *ultimo refinamiento* $\Rightarrow$ *interpolar* }
          $Z^t \leftarrow \text{Interpolar}(Z^t, F_z^t, Z^{t+1}, F_z^{t+1})$
          $Z$ converge
        **fin si**
      **fin si**
    **fin si**
  **fin mientras**
  $Z(R) = Z^t$
**fin para**

---

En el caso particular de la geometría paralela, se simplifica enormemente esta operación,

$$\Delta Z = -\frac{\Delta u' Z^2}{\alpha'_u t'_x + \Delta u' Z} \tag{4.27}$$

de modo que el incremento de $Z$ sólo depende de la profundidad actual, del $\Delta u'$ que se desea fijar, y de los parámetros de calibración. $\Delta u'$ coincide en este caso con el incremento de disparidad.

Para evaluar el método, realizaremos los experimentos sobre pares de imágenes rectificadas, de modo que el cálculo de los incrementos de profundidad sea lo más simple posible. Que las imágenes estén rectificadas no supone una pérdida de generalidad, como se puede deducir del apéndice A sobre geometría epipolar. Es más, en muchos de los métodos basados en correlación, la rectificación de las imágenes es un requisito imprescindible, mientras que son pocos los métodos que son aplicables directamente a cualquier tipo de geometría [Super y Klarquist, 1997] [Robert y Deriche, 1996].

### 4.4.5 Análisis de coste

El algoritmo 4.1 contiene dos bucles anidados:

- El bucle exterior, que se ejecuta para cada región del conjunto de regiones de la imagen de referencia, es decir, $|\mathcal{R}|$ veces.
- El bucle interior, que se ejecuta hasta que se alcanza la convergencia de la profundidad de la región en curso, y que en el peor de los casos se realiza un número de iteraciones $n$ que calcularemos más adelante.

La operación más costosa de todas las que se realizan es el cálculo de $C\mathcal{H} - \mathcal{H}_{12}$, que requiere el cálculo de la región correspondiente $R'$, su media y desviación típica, el gradiente de los pixeles de $R'$ en la dirección de la línea epipolar, la correlación $C$ y los gradientes medios ponderados $\mathcal{H}$ y $\mathcal{H}_{12}$. Todos estos cálculos son de coste lineal respecto al número de pixeles de la región, $N$, y por tanto, el coste de la operación es también lineal, $O(N)$. Esta operación se realiza una vez en cada iteración del bucle interior más una vez en cada iteración del bucle exterior, esto es, $|\mathcal{R}|(n+1)$ veces.

Por tanto, sólo nos queda averiguar cuál es el valor de $n$ para obtener el coste computacional del algoritmo. Para ello, tenemos que considerar las dos etapas del algoritmo: acercamiento y refinamiento.

El caso peor de la etapa de acercamiento, consiste en empezar con $Z = Z_{max}$ y decrementar $Z$ sucesivamente hasta obtener $Z = Z_{min}$, o viceversa. Esto es equivalente a decir que empezamos con disparidad $d = d_{min}$ y la incrementamos sucesivamente hasta obtener $d = d_{max}$, o viceversa. Como los incrementos de disparidad en esta etapa son constantes, $d$ alcanza el valor $d_{max}$ cuando:

$$d = d_{min} + n_1 \Delta d_{ini} \geq d_{max}$$

El número de iteraciones $n_1$ de la etapa de acercamiento sería entonces:

$$n_1 = \frac{d_{max} - d_{min}}{\Delta d_{ini}} = \frac{|\mathcal{D}|}{\Delta d_{ini}}$$

El caso peor de la etapa de refinamiento, consiste en reducir $\Delta d$ sucesivas veces hasta obtener un valor menor que $\Delta d_\epsilon$. Suponiendo que la reducción sea una división por dos, el número de divisiones, $i$, que habría que realizar sería tal que:

$$\frac{\Delta d_{ini}}{2^i} < \Delta d_\epsilon$$

Por tanto,

$$i = \log_2 \frac{\Delta d_{ini}}{\Delta d_\epsilon} + 1$$

Hay que contar con el hecho de que cada vez que se realiza la reducción, sea ésta de la naturaleza que sea (división por 2, interpolación, etc.), es porque se ha sobrepasado el paso por cero y la reducción supone un reintento de la iteración previa. Si el reintento resulta ser un avance válido hacia la solución, se pospone la siguiente reducción hasta la iteración siguiente. Asumiendo que la reducción sea una división por dos, en el peor caso se realizarían reducciones y avances sucesivos, esto es, dos iteraciones por reducción posible. Por tanto, el número total de iteraciones de la etapa de refinamiento en el peor caso sería $n_2 = 2i$. Es decir:

$$n_2 = 2\log_2 \frac{\Delta d_{ini}}{\Delta d_\epsilon} + 2$$

La suma de las dos etapas proporciona un número de iteraciones en el peor caso de

$$n = n_1 + n_2 = \frac{|\mathcal{D}|}{\Delta d_{ini}} + 2\log_2 \frac{\Delta d_{ini}}{\Delta d_\epsilon} + 2$$

Por último queda señalar que, sabiendo que la suma de todos los pixeles de todas las regiones es el total de pixeles de la imagen, la suma para todas las regiones del coste de una operación lineal respecto al número de pixeles de cada región, es lineal respecto al número de pixeles total de la imagen. Por tanto, podemos decir que, para una imagen de referencia de $N_u \times N_v$ pixeles, el coste total del algoritmo es

$$O\left(N_u N_v \left(\frac{|\mathcal{D}|}{\Delta d_{ini}} + \log_2 \frac{\Delta d_{ini}}{\Delta d_\epsilon}\right)\right)$$

De los dos sumandos, el término que domina teniendo en cuenta los valores de $|\mathcal{D}|$, $\Delta d_{ini}$ y $\Delta d_\epsilon$ que se suelen dar en la práctica es el primero, de modo que el coste en el caso peor es

$$O\left(\frac{N_u N_v |\mathcal{D}|}{\Delta d_{ini}}\right)$$

es decir, que en el caso peor, para cada pixel de la imagen, se explora todo el rango de disparidades a intervalos de $\Delta d_{ini}$ pixeles.

Este es el coste computacional del nivel de mayor resolución de la escala ($k = 0$). Para calcular el coste computacional de cualquier nivel $k > 0$ de la escala hay que tener en cuenta que el número de pixeles de la imagen se divide por dos tanto en el eje $u$ como en el eje $v$, y que el intervalo de disparidades a explorar también se reduce. Por tanto, para un nivel cualquiera $k$ de la escala el coste es

$$O\left(\frac{1}{\Delta d_{ini}} \frac{N_u}{2^k} \frac{N_v}{2^k} \frac{|\mathcal{D}|}{2^k}\right)$$

Teniendo en cuenta que

$$\sum_{k=0}^{N} \frac{1}{8^k} < \sum_{k=0}^{\infty} \frac{1}{8^k} = \frac{8}{7}$$

si sumamos los costes de todos los niveles obtenemos

$$O\left(\frac{8}{7} \frac{N_u N_v |\mathcal{D}|}{\Delta d_{ini}}\right)$$

Este coste es comparable con el de los métodos basados en áreas, que se puede denotar $O(N_u N_v |\mathcal{D}| p)$, donde $p$ es el número de pixeles de la ventana sobre la que se aplica la correlación. Tanto $p$ como $\frac{1}{\Delta d_{ini}}$ no suelen ser números muy grandes. Por ejemplo

$$\begin{array}{rcl} \text{ventana de } 5 \times 5 & \Rightarrow & p = 25 \\ \Delta d_{ini} = 0.1 & \Rightarrow & \dfrac{1}{\Delta d_{ini}} = 10 \end{array}$$

Cabe resaltar que el coste calculado para el caso peor está muy por encima del coste real en los experimentos realizados, gracias al acercamiento paulatino que se realiza en los distintos niveles de la escala hacia la solución. Sin embargo, es difícil, si no imposible, definir un caso medio del comportamiento del algoritmo, y por lo tanto, sólo podemos comprobar este funcionamiento a partir del número de iteraciones que ejecuta el algoritmo en los experimentos realizados.

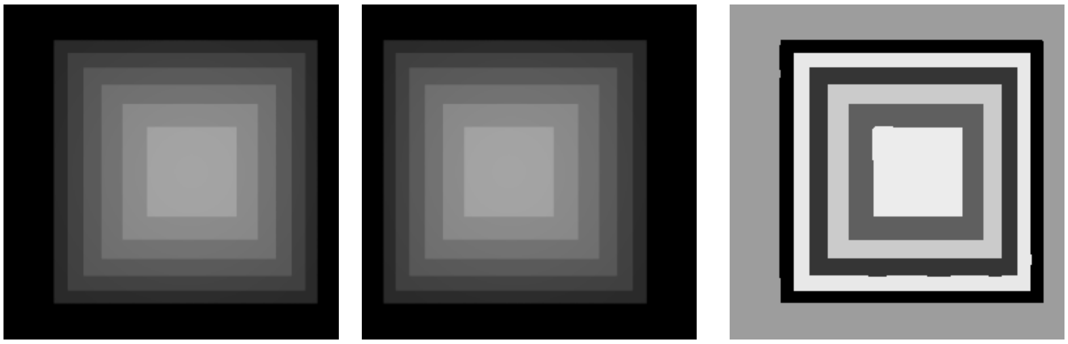
## 4.5 Resultados experimentales

Se han realizado experimentos con imágenes sintéticas y reales y con distintos tipos de segmentación. Los algoritmos de segmentación utilizados en los experimentos son los mismos que los descritos en la sección de resultados experimentales del capítulo 3, es decir, segmentación por unión de regiones [Rosenfeld y Kak, 1982] y segmentación por agrupamiento [Badenas *et al.*, 1997].

En los primeros experimentos, el principal objetivo es evaluar el funcionamiento del método, es decir, comprobar si los cálculos que se realizan para cada región son significativos o no. Para ello, se han realizado diversas pruebas acerca de la evolución de $F_z$ en función de $Z$ en los distintos niveles del esquema multiescala.

### 4.5.1 Experimentos con imágenes sintéticas

En la figura 4.6, se muestra un par sintético de imágenes del que conocemos su profundidad verdadera (de 82 a 117 cm.) y la segmentación de la imagen izquierda mediante la técnica de agrupamiento. El mapa de profundidades verdadero y el mapa resultante del algoritmo propuesto con un rango de profundidades $Z_{min} = 75$ y $Z_{max} = 130$, se muestran en la figura 4.7. Las áreas más claras corresponden a puntos más lejanos, mientras que las más oscuras corresponden a puntos más cercanos.



(a) Imagen izquierda (b) Imagen derecha (c) Segmentación de la imagen izquierda

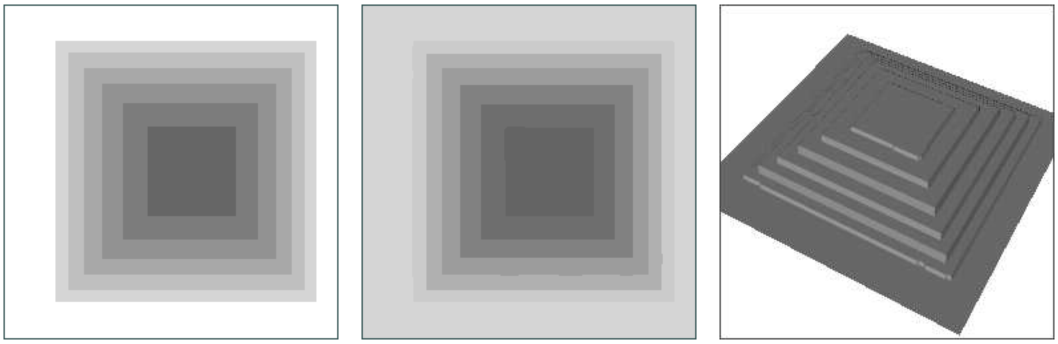
**Figura 4.6:** Par estéreo de imágenes sintéticas *pirámide*.

Los resultados muestran que se han encontrado todas las profundidades, excepto para la región del fondo, que debido a estar limitada por los bordes de la imagen y tener un nivel de gris casi constante, adquiere una profundidad similar a la de la única región adyacente a ella. El mapa de profundidades que se obtiene 4.7(b) es bastante similar al esperado 4.7(a) excepto en la región del fondo.

Sean $R_0, R_1, ...R_5$ las regiones que representan los planos del ejemplo *pirámide* desde el más cercano al más lejano. En la tabla 4.1 se detalla la profundidad verdadera y la obtenida para cada una de estas regiones, así como el número de pixels de la región. Exceptuando la región del fondo, el error relativo medio resultante es del 3.78%.

| Región | $R_0$ | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | fondo |
|---|---|---|---|---|---|---|---|
| Z verdadera | 82 | 89 | 96 | 103 | 110 | 117 | - |
| Z obtenida | 81.5722 | 84.7316 | 90.5292 | 98.9442 | 105.386 | 113.892 | 116.994 |
| # pixels | 4662 | 6977 | 6050 | 7634 | 7627 | 7622 | 24964 |

**Tabla 4.1:** Resultados del ejemplo sintético *pirámide*, $\Delta d_{ini} = 0.5$, $\Delta d_{\epsilon} = 0.00001$.



(a) Profundidad verdadera (b) Profundidad obtenida (c) Reconstrucción 3D

**Figura 4.7:** Mapas de profundidades del ejemplo *pirámide*: Las áreas más claras corresponden a los puntos más lejanos.

### 4.5.2 Selección de $\Delta d_{ini}$

La selección del incremento de disparidad que determina cuál es el incremento de profundidad en cada paso, tiene que ver con la suposición que se ha realizado sobre la escena. Hasta este momento, debido a que la profundidad es constante dentro de las regiones y no se tienen en cuenta las oclusiones, la selección de $\Delta d_{ini}$ no es muy crítica, siempre que al menos cumpla:

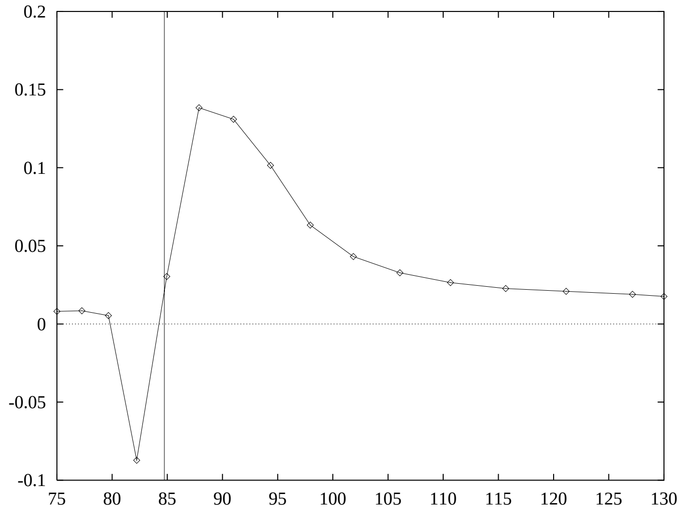
$$0 < \Delta d_{ini} \leq 1$$

Esto es debido a que la convergencia del algoritmo para cada región de la imagen es independiente del resto de regiones.

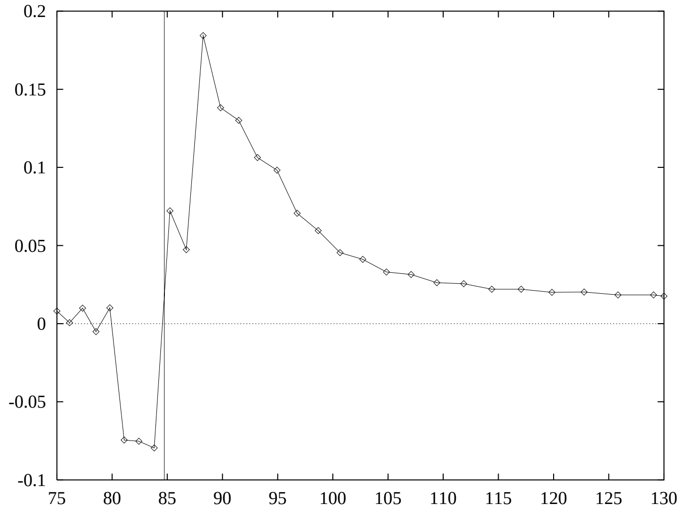
En la figura 4.8 se muestra cómo varía $F_z$ (eje de ordenadas) dependiendo de $Z$ (eje de abcisas) para la región $R_1$ del ejemplo. La línea vertical indica dónde se encuentra el paso por cero. Cada gráfica de esta figura se ha obtenido con un $\Delta d$ diferente, de modo que puede observarse que cuanto mayor es $\Delta d$, la aproximación al paso por cero es más burda.

Sin embargo, el algoritmo 4.1 propuesto permite alcanzar el paso por cero con la precisión que se desee, ya que al detectar que ha sido sobrepasado, se reduce el $\Delta d$ inicial hasta encontrar el paso por cero con la precisión requerida. Otra manera de realizar esta aproximación fina es interpolar entre los dos valores de $Z$ a ambos lados del paso por cero, tantas veces como sea necesario hasta alcanzar la precisión requerida, en lugar de dividir $\Delta d$ por un valor prefijado.

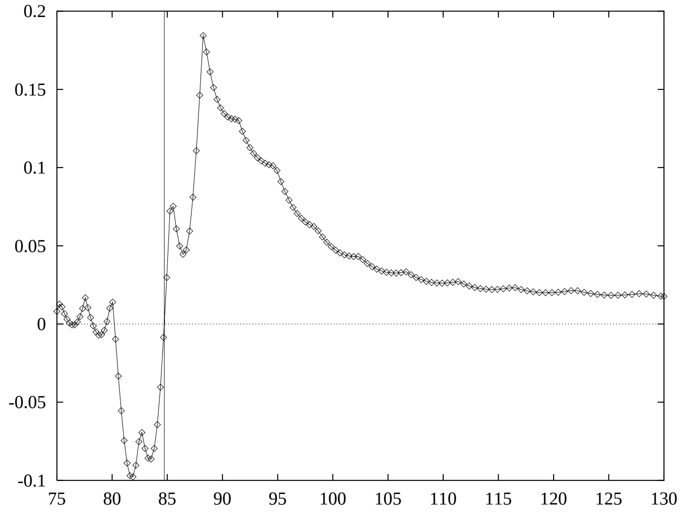
La selección de un $\Delta d_{ini}$ mayor tiene la ventaja de que el algoritmo converge más rápidamente, ya que realiza menos iteraciones para llegar a los alrededores del paso por cero. El nivel de precisión requerido (representado por $\Delta d_\epsilon$) también influye en el número total de iteraciones (a mayor precisión, más iteraciones).




(a) Nivel 0 de la escala, $\Delta d = 1$.




(b) Nivel 0 de la escala, $\Delta d = 0.5$.




(c) Nivel 0 de la escala, $\Delta d = 0.1$.

**Figura 4.8:** Evolución de $F_z, \forall Z$, de la región $R_1$ del ejemplo *pirámide*. La línea vertical indica el paso por cero.

Los resultados de la tabla 4.1 han sido obtenidos con $\Delta d_{ini} = 0.5$ y $\Delta d_\epsilon = 0.00001$. El número de iteraciones en cada nivel y para cada región con estos parámetros varía entre 1 y 30, como se puede observar en la tabla 4.2 (un guión significa que esa región no existe en ese nivel). El número de iteraciones medio (sin tener en cuenta los niveles 7 y 6, dado que el número de iteraciones es casi nulo) es de 25.

| Número de iteraciones (niveles 7 a 0, respectivamente) | | |
|---|---|---|
| Región | $\Delta d_{ini} = 0.5$ $\Delta d_\epsilon = 0.00001$ | $\Delta d_{ini} = 0.5$ $\Delta d_\epsilon = 0.001$ |
| $R_0$ | 0,0,25,26,28,25,22,25 | 0,0,15,17,16,14,14,16 |
| $R_1$ | -,0, 1,26,22,26,24,24 | -,0,1,16,12,16,16,14 |
| $R_2$ | -,0,25,26,27,25,30,23 | -,0,17,17,16,16,18,18 |
| $R_3$ | -,0, 2,24,24,23,25,22 | -,0,2,15,15,15,14,14 |
| $R_4$ | -,-, 0,24,22,25,25,29 | -,-,0,13,12,15,16,18 |
| $R_5$ | -,-,27,29,21,25,21,25 | -,-,17,14,14,14,12,14 |
| fondo | 0,1,25, 3,24,25,28,24 | 0,1,15,2,15,15,16,15 |

**Tabla 4.2:** Número de iteraciones en el ejemplo sintético *pirámide*, con $\Delta d_{ini} = 0.5$, $\Delta d_\epsilon = \{0.00001, 0.001\}$.

Cabe destacar que, en este ejemplo, la mayor parte de las iteraciones son debidas a la etapa de refinamiento. Generalmente no es necesaria tanta precisión, por lo que se puede aumentar $\Delta d_\epsilon$, y en consecuencia el número de iteraciones se reducirá. Por ejemplo, con un $\Delta d_\epsilon = 0.001$, los resultados obtenidos para este ejemplo son idénticos, mientras que el número de iteraciones varía entre 2 y 17, siendo la media 13.52, (ver tabla 4.2).

En la tabla 4.3 se muestra el número de iteraciones para los casos $\Delta d_{ini} = \{0.1, 0.01\}$ y $\Delta d_\epsilon = 0.001$. Las profundidades obtenidas son las mismas que en el ejemplo anterior, debido a que la precisión con que se obtienen los resultados es la misma. Se puede observar que a medida que se reduce $\Delta d_{ini}$ aumenta el número de iteraciones. Sin embargo, con $\Delta d_{ini} = 0.1$ el número de iteraciones medio es 12.26, ligeramente menor que con $\Delta d_{ini} = 0.5$. Esto es debido a que dado un $\Delta d_\epsilon$ constante, al reducir $\Delta d$, la etapa de acercamiento es más larga, pero la etapa de refinamiento es más corta. En este ejemplo, en el caso $\Delta d_{ini} = 0.1$, lo uno compensa lo otro, mientras que en el caso $\Delta d_{ini} = 0.01$, el incremento de iteraciones en la etapa de acercamiento es mucho mayor que el decremento de iteraciones en la etapa de refinamiento.

En la tabla 4.4 se muestran los resultados con $\Delta d_{ini} = 0.5$ y $\Delta d_\epsilon = 0.1$. En este caso las profundidades obtenidas son ligeramente peores (error relativo medio 3.88 %) debido al aumento en $\Delta d_\epsilon$, pero el número de iteraciones se reduce considerablemente, respecto a los casos anteriores.

| Número de iteraciones (niveles 7 a 0, respectivamente) | | |
|---|---|---|
| Región | $\Delta d_{ini} = 0.1$ $\Delta d_{\epsilon} = 0.001$ | $\Delta d_{ini} = 0.01$ $\Delta d_{\epsilon} = 0.001$ |
| $R_0$ | 0,0,15,13,13,14,15,11 | 0,0,46,24,14,17,32,15 |
| $R_1$ | -,0,1,13,10,13,17,11 | -,0,1,42,6,35,74,10 |
| $R_2$ | -,0,14,10,12,16,14,13 | -,0,44,14,22,52,29,39 |
| $R_3$ | -,0,14,12,12,12,13,14 | -,0,32,16,13,32,14,22 |
| $R_4$ | -,-,0,9,10,12,13,15 | -,-,0,7,6,16,21,26 |
| $R_5$ | -,-,12,15,14,12,9,13 | -,-,31,34,15,22,7,33 |
| fondo | 0,1,15,8,15,12,18,11 | 0,1,46,62,18,10,72,8 |

**Tabla 4.3:** Número de iteraciones en el ejemplo sintético *pirámide*, $\Delta d_{ini} = \{0.1, 0.01\}$, $\Delta d_{\epsilon} = 0.001$.

| Región | $R_0$ | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | fondo |
|---|---|---|---|---|---|---|---|
| Z obtenida | 81.5707 | 84.7294 | 90.524 | 98.409 | 105.386 | 113.882 | 116.986 |
| # iteraciones (niveles 7 a 0) | 0,0,6,5, 5,5,5,5 | -,0,1,6, 4,5,6,4 | -,0,6,6, 5,7,6,6 | -,0,2,2, 5,5,5,5 | -,-,0,4, 4,5,5,5 | -,-,5,5, 5,5,4,5 | 0,1,6,6, 5,4,6,4 |

**Tabla 4.4:** Resultados del ejemplo sintético *pirámide*, $\Delta d_{ini} = 0.5$, $\Delta d_{\epsilon} = 0.1$.

### 4.5.3 Evaluación del esquema multiescala

En las gráficas de la figura 4.9 se puede observar la función $F_z$ de la región $R_0$ del ejemplo *pirámide*. Obsérvese que ya en el nivel 4, el paso por cero está muy cerca de la solución, y que el paso por cero del nivel 0 está muy cerca de la solución verdadera. En realidad, la profundidad obtenida para esta región es la que más se parece a su profundidad verdadera. Esto es debido a que esta región es la más cercana y por tanto no sufre el problema de las oclusiones.

En la figura 4.10 se muestran las gráficas relativas a la región $R_3$. Como se puede observar, hay una diferencia mayor entre el paso por cero y la solución verdadera, indicada con la línea vertical. También se observa que hay desplazamientos similares en las curvas que describen las funciones $F_z$ de las demás regiones (figura 4.11).

En la figura 4.10 se pueden observar dos pasos por cero en el nivel 0. En este caso, un $\Delta d_{ini}$ grande podría haber causado la obtención de un paso por cero distinto. Sin embargo, debido a que el paso por cero del nivel 1 se encuentra en la zona positiva del paso por cero de más a la izquierda (el más lejano a la solución verdadera), es éste el que se obtiene sea cual sea el $\Delta d_{ini}$ que se utilice.

Por último, la figura 4.12 muestra con detalle la función $F_z$ de la región $R_2$. En este caso, se puede observar que hay dos pasos por cero también, pero en este caso el paso

(a) Niveles 5 a 0 de la escala, $\Delta d = 0.2$.

(b) Niveles 3 a 0 de la escala, $\Delta d = 0.2$.

(c) Nivel 0 de la escala, $\Delta d = 0.2$.

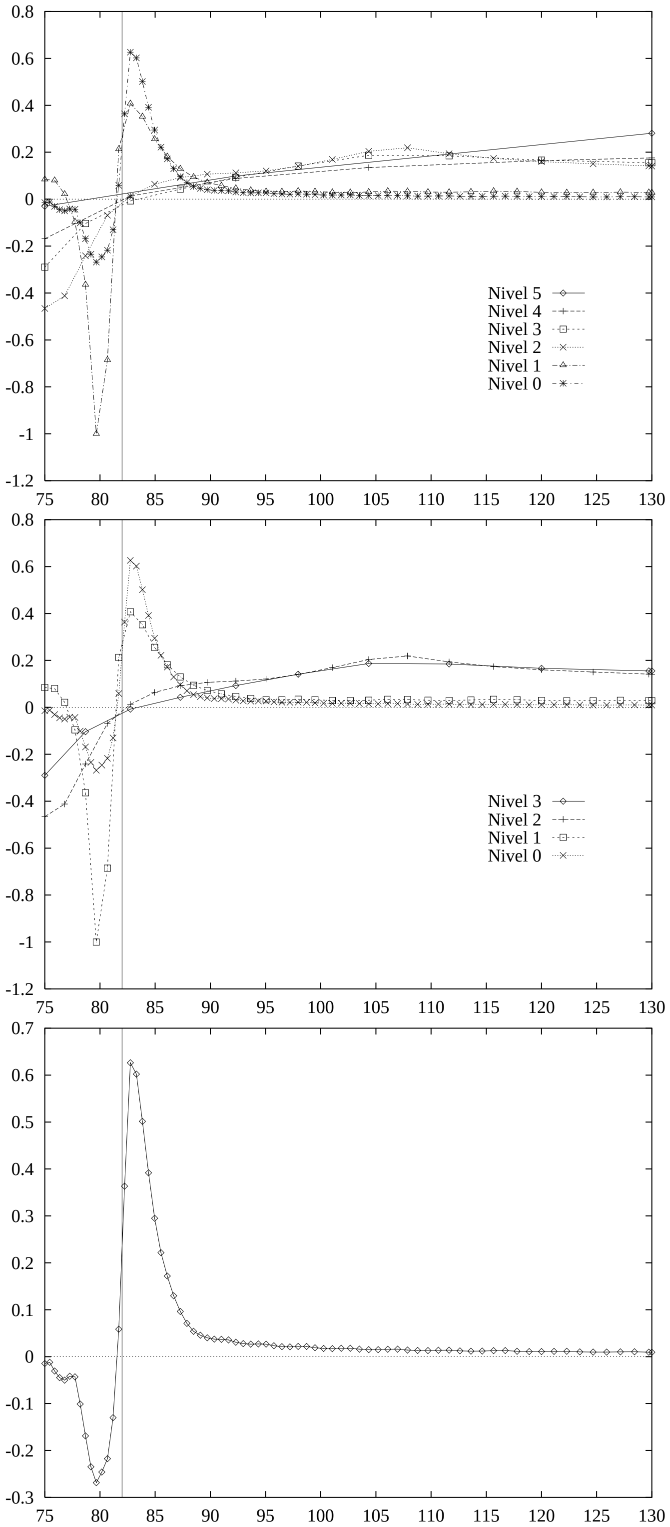
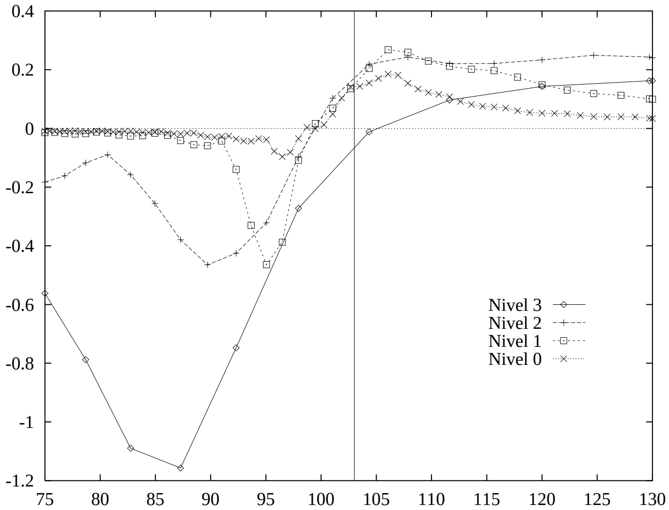
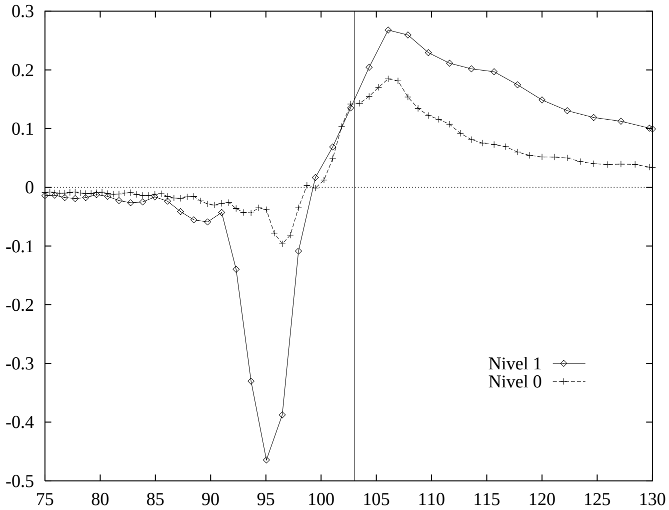



**Figura 4.9:** Evolución de $F_z, \forall Z$, de la región $R_0$ del ejemplo *pirámide*. La línea vertical indica la profundidad verdadera.

(a) Niveles 3 a 0 de la escala, $\Delta d = 0.2$.




(b) Niveles 1 y 0 de la escala, $\Delta d = 0.2$.




(c) Nivel 0 de la escala, $\Delta d = 0.2$.

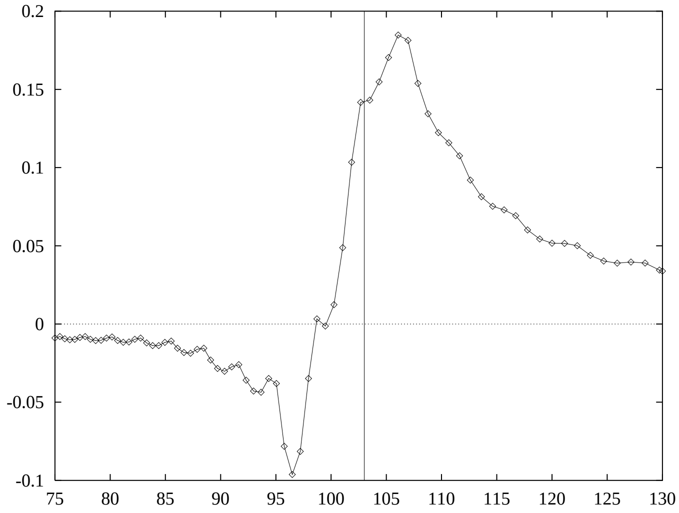
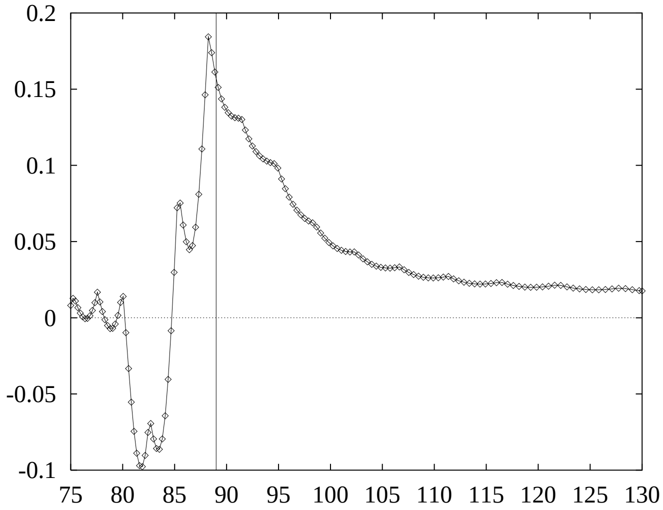
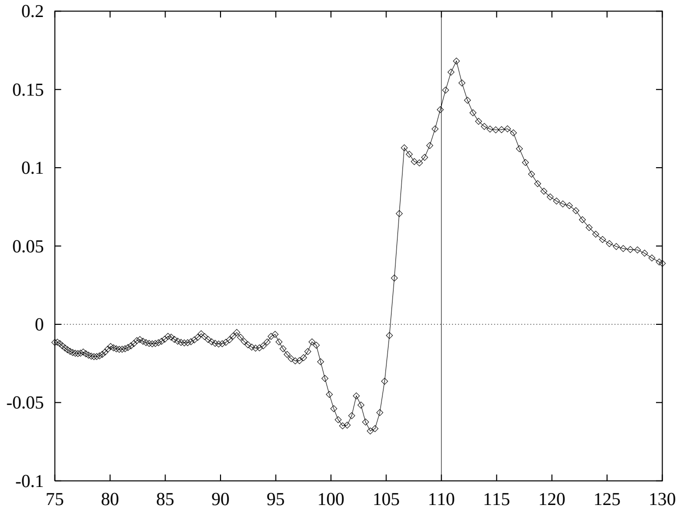



**Figura 4.10:** Evolución de $F_z, \forall Z$, de la región $R_3$ del ejemplo *pirámide*. La línea vertical indica la profundidad verdadera.

(a) Nivel 0, region $R_1$, $\Delta d = 0.1$.




(b) Nivel 0, region $R_4$, $\Delta d = 0.1$.




(c) Nivel 0, region $R_5$, $\Delta d = 0.1$.

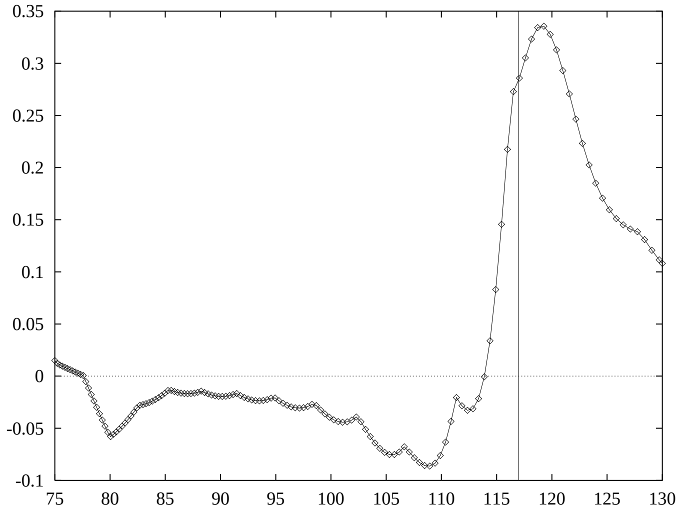



**Figura 4.11:** Evolución de $F_z, \forall Z$, de varias regiones del ejemplo *pirámide*. La línea vertical indica la profundidad verdadera.

por cero del nivel 1 dirige la solución hacia el paso por cero de más a la derecha (el más cercano a la solución verdadera). Nótese también, que el paso por cero del nivel 3 se encuentra todavía bastante lejos de la solución. Sin embargo, el paso por cero del nivel 2 se acerca ya lo suficiente, de modo que llegar desde ahí al paso por cero del nivel 1 y del nivel 0 se realiza en muy pocas iteraciones.

### 4.5.4 Otros ejemplos sintéticos

Se han realizado dos tipos de experimentos, respecto al modo en que se calculan las regiones de los diferentes niveles de escala de la imagen de referencia.

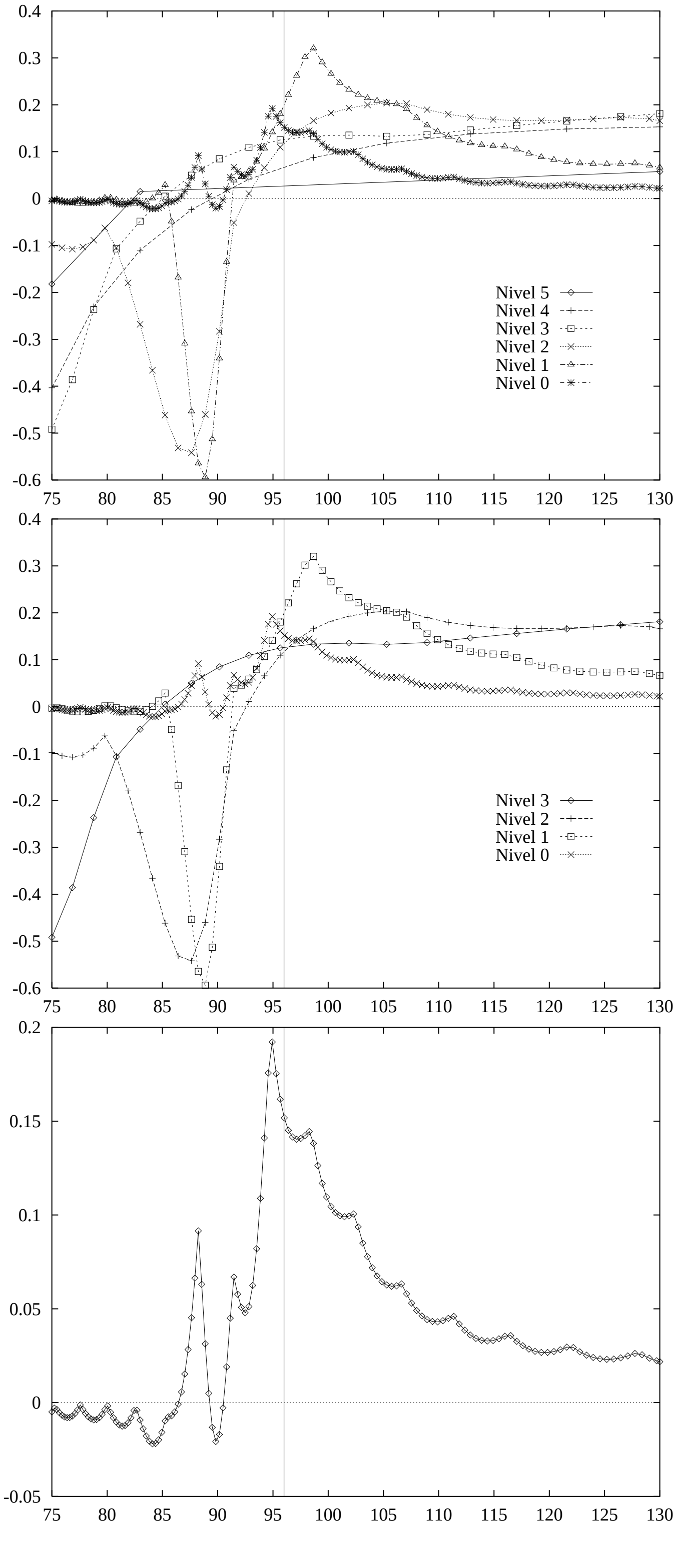
- segmentando la imagen de mayor resolución y creando versiones escaladas del resultado, y
- segmentando todos y cada uno de los niveles de la imagen de referencia.

Los experimentos han mostrado que la segunda opción proporciona resultados mucho mejores que la primera, tal como cabía esperar, tanto en los experimentos con imágenes sintéticas como con imágenes reales. Por tanto, los resultados que se muestran en este capítulo corresponden al segundo caso.

A continuación se muestran algunos resultados con distintos métodos de segmentación. En la figura 4.13, se muestra la imagen que se ha tomado de referencia en el par sintético de imágenes *corridor* [Froehlinghaus, 1997], segmentado mediante la técnica de agrupamiento [Badenas *et al.*, 1997]. En este ejemplo, la restricción de profundidad constante en las regiones no es aplicable a la escena, puesto que se trata principalmente de planos inclinados y algunas superficies curvas (esfera y cono).

Nótese que las regiones de las paredes y el techo del pasillo son las que peor representan a las superficies reales, debido a que la suposición de profundidad constante no es aplicable en estos casos. Aunque aparecen discontinuidades en la profundidad en todas las fronteras de las regiones debido a la restricción impuesta, en general las profundidades cambian poco de una región a la siguiente, salvo en el caso de las discontinuidades esperadas, como son el cono y la esfera, que claramente destacan respecto a sus regiones vecinas, y algunas regiones grandes que representan planos inclinados, como la región central del techo. El nivel 0, que consta de 141 regiones, tarda en ejecutarse 32.86 segundos en un Pentium 100MHz con 64 Mbytes de memoria RAM. El número de iteraciones máximo en cada nivel es: 0, 0, 3, 14, 17, 23, 30, 25. Para que se pueda apreciar mejor las profundidades de los objetos más cercanos, se muestra también el mapa de disparidades extraído a partir del mapa de profundidades.

En la figura 4.14, se muestra la misma imagen segmentada mediante la técnica de unión de regiones [Rosenfeld y Kak, 1982], y el mapa de profundidades obtenido con $\Delta d_{ini} = 0.1$ y $\Delta d_{\epsilon} = 0.001$. El número de iteraciones medio del nivel 7 al 0, respectivamente, es: 0, 2, 1, 1, 5.4, 16.4, 12.8, 13.9. El número de iteraciones máximo en cada



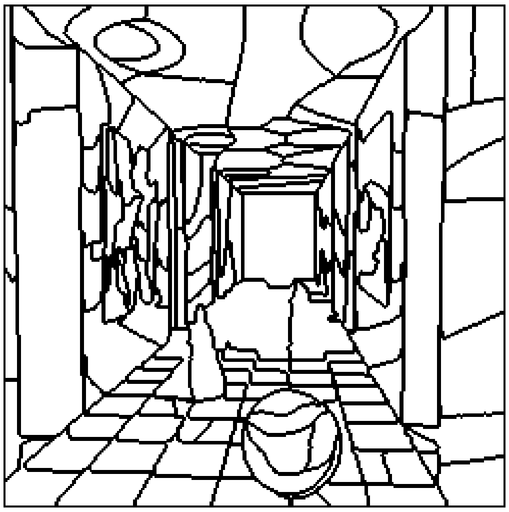

(a) Niveles 5 a 0, $\Delta d = 0.1$.

(b) Niveles 3 a 0, $\Delta d = 0.1$.

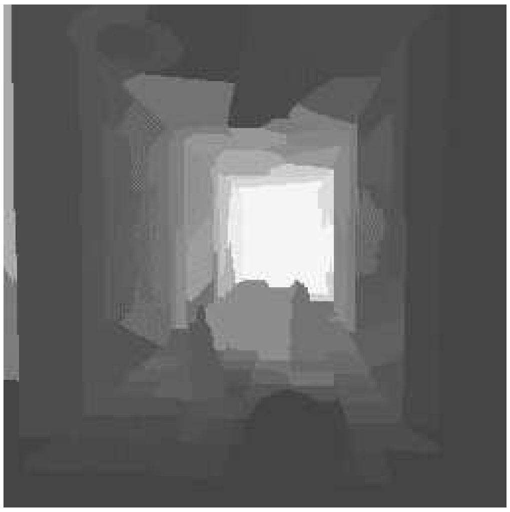
(c) Nivel 0, $\Delta d = 0.1$.

**Figura 4.12:** Evolución de $F_z, \forall Z$, de la región $R_2$ del ejemplo *pirámide*. La línea vertical indica la profundidad verdadera.

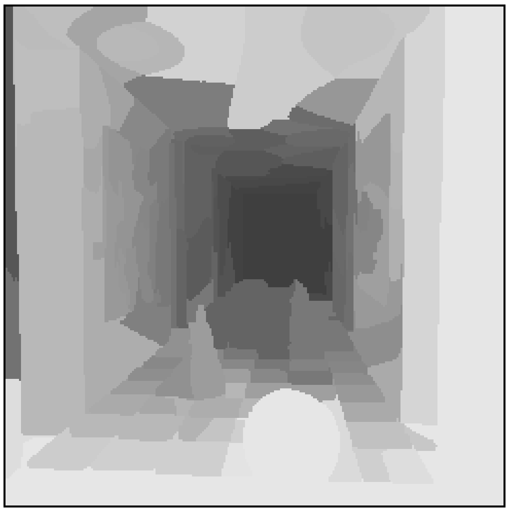
(a) Imagen de referencia



(b) Segmentación por agrupamiento



(c) Mapa de profundidades



(d) Mapa de disparidades

**Figura 4.13:** Resultados para el par de imágenes sintéticas *corridor*, segmentación mediante la técnica de agrupamiento, $\Delta d_{ini} = 0.1$ y $\Delta d_{\epsilon} = 0.001$.

nivel es: 0, 2, 1, 1, 8, 24, 36, 37. Para que se pueda apreciar mejor las profundidades de los objetos más cercanos, se muestra también el mapa de disparidades extraido a partir del mapa de profundidades, y el mapa de disparidades real, para poder comparar con él los resultados.

Estos resultados, comparados con los obtenidos en el capítulo anterior, tienen las siguientes ventajas:

- Las disparidades obtenidas tienen precisión subpixel. Esto da lugar a que las disparidades varíen más suavemente entre regiones vecinas. Por ejemplo, a cada lado del pasillo, las disparidades de la región más cercana a la más lejana.
- Al no considerar la segmentación de la segunda imagen, se evitan los problemas generados por las diferencias en la segmentación de las regiones correspondientes. Esto da lugar a que haya menos imprecisión en las disparidades (véase la diferencia en la esfera en los mapas de disparidades de las figuras 3.11 y 4.13).
- El esquema multiescala, que propaga los resultados de un nivel a otro, permite obtener la correspondencia de algunas regiones que en ocasiones son difíciles de emparejar. Por ejemplo, algunas regiones sin correspondencia (por su escaso tamaño o por las dificultades en la segmentación) del mapa de disparidades de la figura 3.9 obtienen una disparidad aceptable en el mapa de disparidades de la figura 4.14.

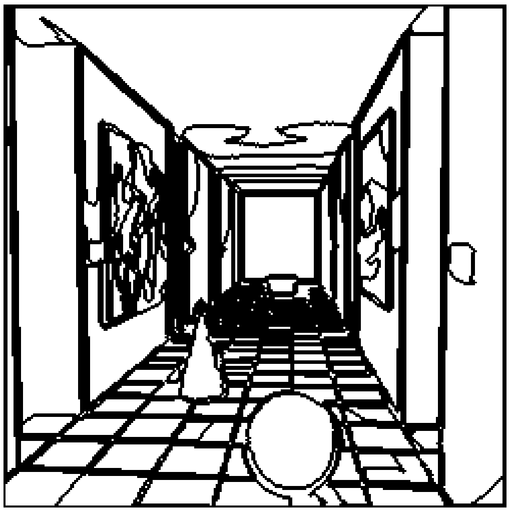
Estas diferencias también se pueden apreciar en los experimentos con pares de imágenes reales, como se muestra a continuación.

### 4.5.5 Experimentos con escenas reales

En la figura 4.15, se muestran los resultados para un par de imágenes reales, de tamaño 512x480, y cuya segmentación por unión de regiones proporciona 267 regiones en el nivel de mayor resolución. Aunque se ha aplicado la restricción de profundidad constante en las regiones, el mapa de profundidades obtenido (figura 4.15(c)) proporciona resultados satisfactorios y da una aproximación de la escena tridimensional.

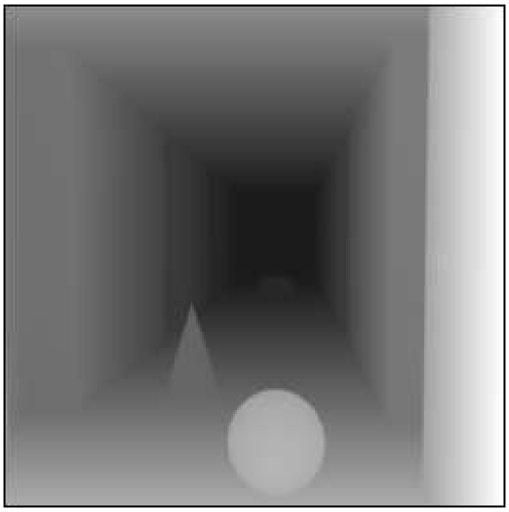
Cabe resaltar que el parquímetro está ligeramente más cercano que el matorral que hay alrededor suyo y que el matorral de más a la izquierda está más cerca que el parquímetro. Aunque hay algunas regiones delgadas verticales que adquieren profundidades erróneas, los resultados nos animan a generalizar el método a todo tipo de superficies planas.

En la figura 4.15(d) se puede observar la reconstrucción tridimensional de la escena, donde se ha añadido la textura de las regiones en la imagen. Aunque algunas áreas no obtienen la profundidad esperada (algunas regiones negras y una región blanca, que indican la profundidad mínima y máxima, respectivamente), el método obtiene resultados

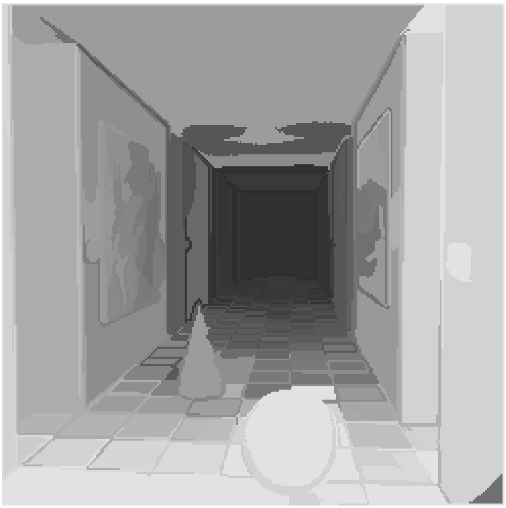


(a) Segmentación por unión de regiones

(b) Mapa de profundidades obtenido



(c) Mapa de disparidades real



(d) Mapa de disparidades obtenido

**Figura 4.14:** Resultados para el par de imágenes sintéticas *corridor*, segmentación por unión de regiones, con $\Delta d_{ini} = 0.1$ y $\Delta d_{\epsilon} = 0.001$.

(a) Imagen izquierda (512x480)

(b) Imagen derecha

(c) Mapa de profundidades

(d) Reconstrucción 3D

**Figura 4.15:** Resultados para un par de imágenes reales, *parking meter*.

satisfactorios incluso en el caso de las discontinuidades de la profundidad, como se puede observar en las fronteras del parquímetro y del coche.

En la figura 4.16 se muestran otros resultados con imágenes del *Calibrated Imaging Laboratory* (CIL)[2] de la Universidad Carnegie Mellon (CMU). En el ejemplo *cil0002* las disparidades del azulejo varían ligeramente de arriba a abajo, de más lejos a más cerca. Esta variación es en global de aproximadamente un pixel. Cabe notar que se detecta esta variación a nivel subpixel, como se puede observar tanto en el mapa de disparidades como en el mapa de profundidades.

## 4.6 Escenas con superficies planas

### 4.6.1 Restricción de superficies planas

Bajo esta restricción, la profundidad de cada región $R$ puede ser cualquier plano del espacio 3D (figura 4.17), que puesto en coordenadas normalizadas de la imagen sería

$$au + bv + cz = d \tag{4.28}$$

de modo que, despejando $z$ obtenemos que ésta depende de tres variables independientes,

$$\begin{aligned} z &= \alpha u + \beta v + \gamma & (4.29)\\ \alpha &= -\frac{a}{c} & (4.30)\\ \beta &= -\frac{b}{c} & (4.31)\\ \gamma &= +\frac{d}{c} & (4.32) \end{aligned}$$

La restricción de planos fronto-paralelos no es más que un caso particular de ésta, donde se ha fijado $\alpha = \beta = 0$.

Por tanto, la región correspondiente a una región cualquiera $R$ es:

$$R' = g_{12}(R, Z(R)) = \{m' = f_{12}((u, v), \alpha(R)u + \beta(R)v + \gamma(R)), \forall (u, v) \in R\} \tag{4.33}$$

donde $\alpha(R), \beta(R), \gamma(R) \in \mathbb{R}$ determinan la ecuación del plano de la superficie 3D cuyas proyecciones son $R$ y $R'$.

En la figura 4.17 se muestra un ejemplo del planteamiento que se realiza. En el ejemplo, los planos sólo pueden estar inclinados en el eje $X$, que es paralelo a $u$. Por tanto, la región correspondiente $R'_b$ a una región cualquiera $R_b$ depende no sólo de la profundidad a la que se encuentra la región, sino también a la inclinación del plano. Si consideramos la inclinación en $Y$ el razonamiento es análogo, y considerar cualquier orientación del plano $S_b$ sería una combinación de ambos.

[2] `http://www.cs.cmu.edu/~cil/cil-ster.html`

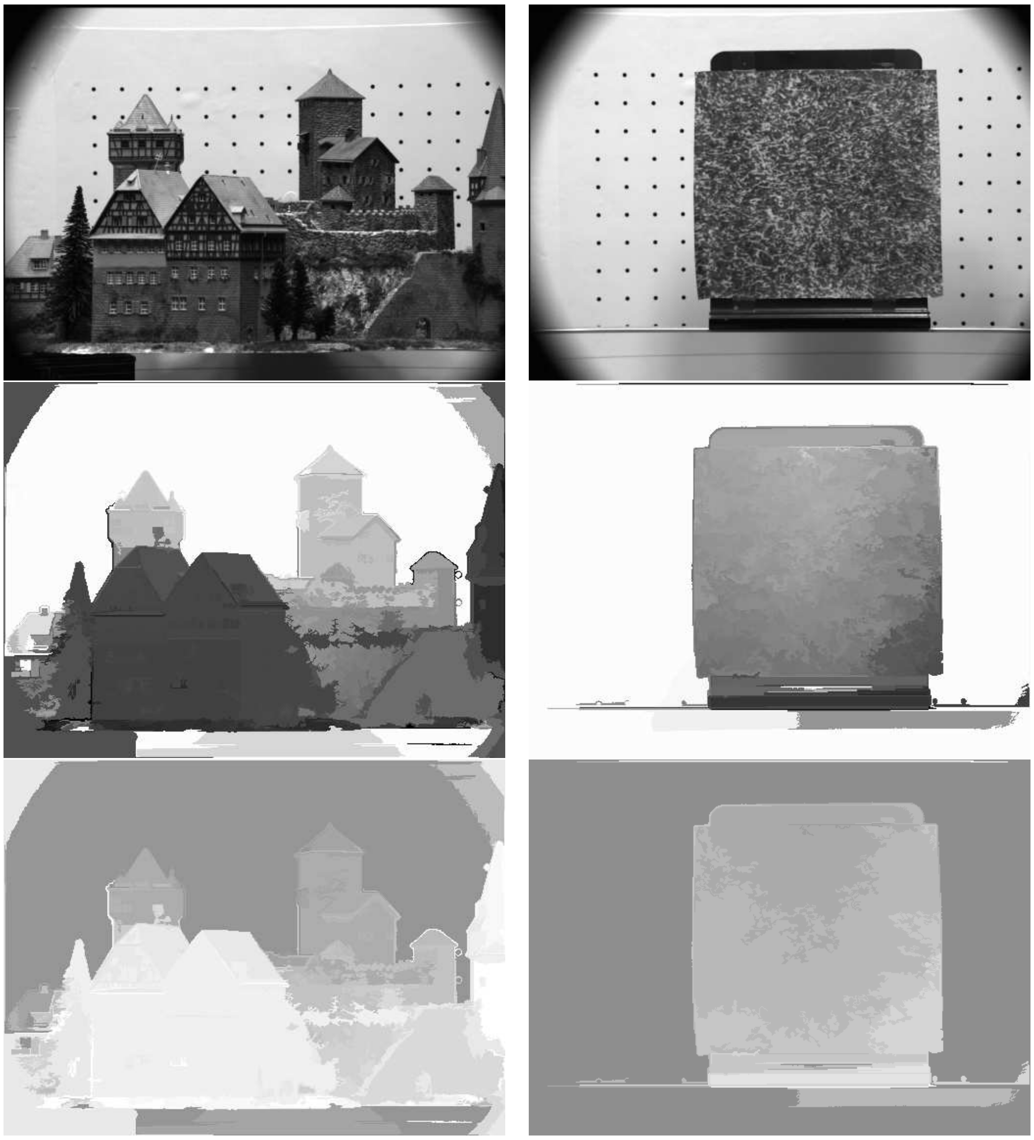


**Figura 4.16:** Resultados para un par de imágenes reales, *cil0001* y *cil0002*, del conjunto de imágenes CIL. De arriba a abajo: imágenes de referencia, mapas de profundidades y mapas de disparidades.

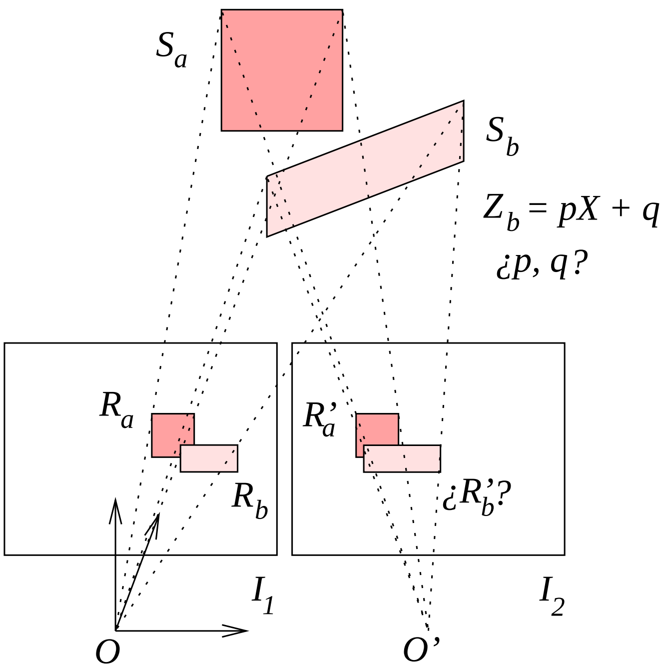



**Figura 4.17:** Planteamiento del problema: superficies planas.

### 4.6.2 Cálculo de la derivada de la función de energía

La principal diferencia con la restricción anterior es que ahora tenemos que calcular tres funciones $\alpha, \beta, \gamma$ en lugar de una, $Z$. Así pues, la ecuación de energía a minimizar (ecuación 4.9) se puede reescribir como,

$$E(\alpha, \beta, \gamma) = \int_u \int_v F(u, v, \alpha, \beta, \gamma, \frac{\partial \alpha}{\partial u}, \frac{\partial \alpha}{\partial v}, \frac{\partial \beta}{\partial u}, \frac{\partial \beta}{\partial v}, \frac{\partial \gamma}{\partial u}, \frac{\partial \gamma}{\partial v})\ du\ dv \tag{4.34}$$

Para obtener las condiciones necesarias de extremo de esta función de energía que depende de múltiples funciones de varias variables independientes, se varía sólo una de las funciones, dejando las demás invariables. Entonces, la función $E(\alpha, \beta, \gamma)$ se transforma en una función que depende sólo de una función variable, por ejemplo $\alpha$. Por consiguiente, la función $\alpha$ que realiza el extremo debe satisfacer la ecuación de Euler [Elsgoltz, 1977],

$$F_\alpha - \frac{\partial}{\partial u}\{F_{\alpha_u}\} - \frac{\partial}{\partial v}\{F_{\alpha_v}\} = 0 \tag{4.35}$$

donde $\alpha_u = \frac{\partial \alpha}{\partial u}, \alpha_v = \frac{\partial \alpha}{\partial v}$. Como este razonamiento es aplicable a cualquier función, se obtiene el sistema de ecuaciones diferenciales de segundo orden

$$F_\alpha - \frac{\partial}{\partial u}\{F_{\alpha_u}\} - \frac{\partial}{\partial v}\{F_{\alpha_v}\} = 0$$
$$F_\beta - \frac{\partial}{\partial u}\{F_{\beta_u}\} - \frac{\partial}{\partial v}\{F_{\beta_v}\} = 0$$

$$F_\gamma - \frac{\partial}{\partial u}\{F_{\gamma_u}\} - \frac{\partial}{\partial v}\{F_{\gamma_v}\} = 0$$

Sin embargo, la suposición de superficies planas hace que $\frac{\partial \alpha}{\partial u}, \frac{\partial \alpha}{\partial v}, \frac{\partial \beta}{\partial u}, \frac{\partial \beta}{\partial v}, \frac{\partial \gamma}{\partial u}, \frac{\partial \gamma}{\partial v}$ sean constantes dentro de la región. Además no imponemos ninguna restricción en las fronteras entre regiones. Por tanto, los términos $F_{\alpha_u} = F_{\alpha_v} = F_{\beta_u} = F_{\beta_v} = F_{\gamma_u} = F_{\gamma_v} = 0$, y el sistema de ecuaciones resultante es:

$$F_\alpha = \frac{\partial F}{\partial \alpha} = 0 \tag{4.36}$$

$$F_\beta = \frac{\partial F}{\partial \beta} = 0 \tag{4.37}$$

$$F_\gamma = \frac{\partial F}{\partial \gamma} = 0 \tag{4.38}$$

Hay que tener en cuenta que este sistema de ecuaciones se ha obtenido fijando dos de las funciones y variando una en cada caso, de modo que el funcionamiento del algoritmo que resuelva el sistema debe ser análogo. Por tanto, los incrementos/decrementos en cada una de las funciones se realizarán en iteraciones diferentes del algoritmo.

El desarrollo de estas ecuaciones es análogo al realizado en la sección 4.4, salvo que en lugar de $\frac{\partial m'_i}{\partial z}$ obtendremos $\frac{\partial m'_i}{\partial \alpha}, \frac{\partial m'_i}{\partial \beta}$ y $\frac{\partial m'_i}{\partial \gamma}$, respectivamente. Por tanto, éstas son las únicas derivadas que quedan por realizar.

### Derivadas parciales del punto correspondiente

Para calcular la derivada parcial del punto correspondiente respecto a $\alpha, \beta$ y $\gamma$, tenemos en cuenta la ecuación de cálculo de $m'_i$ a partir del pixel de referencia $m_i$ (ver ecuación A.33 y su forma desarrollada A.37 y A.38), y la ecuación de $z$ (4.29),

$$\frac{\partial}{\partial z}\{m'_i\} = \frac{\partial}{\partial z}\{u', v'\} = \frac{\partial}{\partial z}\left\{\frac{a_1 z + b_1}{cz + t_z} + u'_0, \frac{a_2 z + b_2}{cz + t_z} + v'_0\right\} \tag{4.39}$$

$$z = \alpha u + \beta v + \gamma \tag{4.40}$$

por tanto

$$\frac{\partial}{\partial \alpha}\{m'_i\} = \left(\frac{a_1 u t_z - b_1 u c}{(cz + t_z)^2}, \frac{a_2 t_z - b_2 c}{(cz + t_z)^2}\right) = \left(u\frac{a_1 t_z - b_1 c}{(cz + t_z)^2}, \frac{a_2 t_z - b_2 c}{(cz + t_z)^2}\right) \tag{4.41}$$

Análogamente

$$\frac{\partial}{\partial \beta}\{m'_i\} = \left(v\frac{a_1 t_z - b_1 c}{(cz + t_z)^2}, \frac{a_2 t_z - b_2 c}{(cz + t_z)^2}\right) \tag{4.42}$$

$$\frac{\partial}{\partial \gamma}\{m'_i\} = \left(\frac{a_1 t_z - b_1 c}{(cz + t_z)^2}, \frac{a_2 t_z - b_2 c}{(cz + t_z)^2}\right) \tag{4.43}$$

Finalmente, obtenemos que $\frac{\partial m_i'}{\partial\alpha} = u\frac{\partial m_i'}{\partial\gamma}$ y $\frac{\partial m_i'}{\partial\beta} = v\frac{\partial m_i'}{\partial\gamma}$. El valor de $\frac{\partial m_i'}{\partial\gamma}$ indica la dirección de la línea epipolar, mientras que $\frac{\partial m_i'}{\partial\alpha}$, $\frac{\partial m_i'}{\partial\beta}$ están ponderados por el valor de $u$ y $v$, respectivamente.

### Definición de los gradientes medios ponderados respecto a $\alpha, \beta$ y $\gamma$

Para cada derivada parcial necesitamos definir una función análoga a $\mathcal{H}$, tal que al sustituir los valores de $\frac{\partial m_i'}{\partial\alpha}$, $\frac{\partial m_i'}{\partial\beta}$ queda,

$$\mathcal{H}_\alpha(R') = \frac{1}{N}\sum_{m_i'\in R'} \mathcal{I}_2(m_i')I_2'(m_i')\frac{\partial m_i'}{\partial\alpha} = \frac{1}{N}\sum_{m_i'\in R'} \mathcal{I}_2(m_i')uI_2'(m_i')\frac{\partial m_i'}{\partial\gamma} \quad (4.44)$$

$$\mathcal{H}_\beta(R') = \frac{1}{N}\sum_{m_i'\in R'} \mathcal{I}_2(m_i')I_2'(m_i')\frac{\partial m_i'}{\partial\beta} = \frac{1}{N}\sum_{m_i'\in R'} \mathcal{I}_2(m_i')vI_2'(m_i')\frac{\partial m_i'}{\partial\gamma} \quad (4.45)$$

$$\mathcal{H}_\gamma(R') = \frac{1}{N}\sum_{m_i'\in R'} \mathcal{I}_2(m_i')I_2'(m_i')\frac{\partial m_i'}{\partial\gamma} \quad (4.46)$$

Hay que resaltar que los valores de $u$ y $v$ que aparecen en estas ecuaciones se refieren al punto $m_i$ de la imagen de referencia, mientras que los valores de intensidad se refieren al punto $m_i'$ de la otra imagen.

### Derivadas parciales de la función de energía

Siguiendo el mismo razonamiento que en la sección 4.4, el sistema de ecuaciones resultante es el siguiente:

$$\frac{1}{N}\sum_{m_i\in R} \mathcal{I}_1(m_i)\left(\mathcal{I}_2(m_i')\,\mathcal{H}_\alpha(R') - uI_2'(m_i')\frac{\partial m_i'}{\partial\gamma}\right) = 0 \quad (4.47)$$

$$\frac{1}{N}\sum_{m_i\in R} \mathcal{I}_1(m_i)\left(\mathcal{I}_2(m_i')\,\mathcal{H}_\beta(R') - vI_2'(m_i')\frac{\partial m_i'}{\partial\gamma}\right) = 0 \quad (4.48)$$

$$\frac{1}{N}\sum_{m_i\in R} \mathcal{I}_1(m_i)\left(\mathcal{I}_2(m_i')\,\mathcal{H}_\gamma(R') - I_2'(m_i')\frac{\partial m_i'}{\partial\gamma}\right) = 0 \quad (4.49)$$

El algoritmo de correspondencia debe buscar la solución en un espacio tridimensional formado por $\alpha$, $\beta$ y $\gamma$, donde $\alpha$ representa la inclinación del plano en la dirección de $u$, $\beta$ representa la inclinación del plano en la dirección de $v$, y $\gamma$ representa la profundidad global del plano. La solución es un paso por (0, 0, 0) de $(F_\alpha, F_\beta, F_\gamma)$ en dicho espacio. El algoritmo ha de detenerse cuando $F_\alpha$, $F_\beta$ y $F_\gamma$ alcanzan el valor 0 a la vez.

## 4.7 Discusión y conclusiones

En este capítulo se ha introducido un método de búsqueda de la correspondencia de las regiones de una imagen de un par estereoscópico. Este método se basa en la minimización de una función de energía que consiste en el error de correlación de la intensidad de los puntos de cada región con la intensidad de los puntos de su región correspondiente. La región correspondiente se calcula a cada paso, a partir de la profundidad de la región estimada hasta ese momento. Por tanto, correspondencia y reconstrucción se realizan en un solo proceso, obteniendo una representación poliédrica de la escena, que se puede traducir en un mapa de profundidades denso.

Los resultados experimentales permiten validar los diferentes aspectos del método desarrollado, como son el esquema multiescala y el algoritmo de correspondencia para cada nivel. Los resultados obtenidos con escenas reales muestran que la principal ventaja de esta metodología respecto a otros métodos de correspondencia de regiones consiste en que se evitan los problemas generados por las diferencias en la segmentación de las regiones correspondientes, y por tanto, las profundidades y disparidades obtenidas para cada región son más precisas.

El método se ha desarrollado e implementado bajo la suposición de que la escena está formada por planos paralelos al plano imagen, con un coste computacional teórico similar al de las técnicas basadas en áreas. Sin embargo, es posible la generalización del método a la suposición de que la escena está formada por planos de cualquier orientación, mediante una búsqueda, en un espacio tridimensional, del plano que proporcione un error mínimo entre sus dos proyecciones. El desarrollo, implementación y validación de los algoritmos referentes a este caso es la principal línea de trabajo futuro que parte de esta tesis doctoral.

La técnica propuesta permite una ampliación sencilla al uso de más de 2 imágenes, añadiendo a la función de energía los sumandos debidos a cada imagen. Otra extensión inmediata sería la utilización de pares estéreo de imágenes no rectificados, por ejemplo, calculando los incrementos de $Z$ a partir de un $\Delta u'$ fijo y un centroide de cada región.

Por último, cabe incluir como trabajo futuro la extensión del método a otros tipos de superficies como, por ejemplo, superficies cuadráticas.

## 4.A Apéndice: Derivadas parciales

### 4.A.1 Derivada parcial del punto correspondiente

Para calcular la derivada parcial del punto correspondiente respecto a $z$ tenemos en cuenta la ecuación de cálculo de $m_i'$ a partir del pixel de referencia $m_i$ (ver ecuación

A.33 y su forma desarrollada A.37 y A.38)

$$\frac{\partial}{\partial z}\{m_i'\} = \frac{\partial}{\partial z}\{u', v'\} = \frac{\partial}{\partial z}\left\{\frac{a_1 z + b_1}{cz + t_z'} + u_0', \frac{a_2 z + b_2}{cz + t_z'} + v_0'\right\} \tag{4.50}$$

donde $a_1 = \alpha_u'(k_{11}u + k_{12}v + k_{13})$, $b_1 = \alpha_u' t_x'$, $a_2 = \alpha_v'(k_{21}u + k_{22}v + k_{23})$, $b_2 = \alpha_v' t_y'$ y $c = k_{31}u + k_{32}v + k_{33}$, siendo los coeficientes $k_{ij}$ dependientes de los parámetros de calibración (ecuaciones A.34-A.36).

Por tanto,

$$\frac{\partial}{\partial z}\{m_i'\} = \left(\frac{a_1 t_z' - b_1 c}{(cz + t_z')^2}, \frac{a_2 t_z' - b_2 c}{(cz + t_z')^2}\right) \tag{4.51}$$

donde

$$\begin{aligned}
a_1 t_z' - b_1 c &= \alpha_u' t_z' \left(\frac{r_{11}'}{\alpha_u}(u - u_0) + \frac{r_{12}'}{\alpha_v}(v - v_0) + r_{13}'\right) \\
&- \alpha_u' t_x' \left(\frac{r_{31}'}{\alpha_u}(u - u_0) + \frac{r_{32}'}{\alpha_v}(v - v_0) + r_{33}'\right) \\
&= \alpha_u' \begin{bmatrix} \frac{u-u_0}{\alpha_u} & \frac{v-v_0}{\alpha_v} & 1 \end{bmatrix} \left( t_z' \begin{bmatrix} r_{11}' \\ r_{12}' \\ r_{13}' \end{bmatrix} - t_x' \begin{bmatrix} r_{31}' \\ r_{32}' \\ r_{33}' \end{bmatrix} \right) \\
a_2 t_z' - b_2 c &= \alpha_v' t_z' \left(\frac{r_{21}'}{\alpha_u}(u - u_0) + \frac{r_{22}'}{\alpha_v}(v - v_0) + r_{23}'\right) \\
&- \alpha_v' t_y' \left(\frac{r_{31}'}{\alpha_u}(u - u_0) + \frac{r_{32}'}{\alpha_v}(v - v_0) + r_{33}'\right) \\
&= \alpha_v' \begin{bmatrix} \frac{u-u_0}{\alpha_u} & \frac{v-v_0}{\alpha_v} & 1 \end{bmatrix} \left( t_z' \begin{bmatrix} r_{21}' \\ r_{22}' \\ r_{23}' \end{bmatrix} - t_y' \begin{bmatrix} r_{31}' \\ r_{32}' \\ r_{33}' \end{bmatrix} \right)
\end{aligned}$$

Por tanto $\frac{\partial}{\partial z}\{m_i'\}$ es un vector $(\frac{\partial}{\partial z}\{u'\}, \frac{\partial}{\partial z}\{v'\})$ cuyas coordenadas dependen del pixel de referencia $(u, v)$, su profundidad, $z$, y los parámetros de calibración. En el caso particular de geometría paralela, donde sólo hay una traslación $t_x' = b$ en el eje de abscisas,

$$\begin{aligned}
a_1 t_z' - b_1 c &= -\alpha_u' b \\
a_2 t_z' - b_2 c &= 0 \\
cz + t_z' &= z
\end{aligned}$$

de modo que

$$\frac{\partial}{\partial z}\{m_i'\} = \left(\frac{-\alpha_u' b}{z^2}, 0\right) \tag{4.52}$$

El significado de estas ecuaciones es el siguiente: el diferencial de $u'$ y $v'$ que se produce cuando se modifica $z$ es tal que $m' = (u', v')$ se mueve a lo largo de la línea epipolar de $m$, $l'_m$. En el caso de geometría paralela, sólo se modifica $u'$ porque las líneas epipolares son paralelas al eje de abscisas y por tanto el incremento en $v'$ es 0.

El valor de $z$ al cuadrado en el denominador indica que un mismo incremento $\Delta z$ produce un incremento en $m'$, $\Delta m'$, que es mayor cuanto menor es $z$. Es decir, cuanto más cerca está un objeto de la cámara, un $\Delta z$ fijo produce un $\Delta m'$ mayor en la segunda imagen.

### 4.A.2 Derivada parcial de la intensidad

La derivada parcial de la intensidad en un punto $m'_i$ es

$$\frac{\partial}{\partial z}\{I_2(m'_i)\} = I'_2(m'_i)\frac{\partial}{\partial z}\{m'_i\} \tag{4.53}$$

Como ya se ha visto, los incrementos en $z$ producen que $m'_i$ se traslade a lo largo de la línea epipolar correspondiente a $m_i$. Por tanto, esta ecuación significa que la primera derivada de la intensidad (gradiente de la intensidad) debe calcularse en la dirección de la línea epipolar que pasa por $m'_i$.

### 4.A.3 Derivada parcial de la intensidad media de la región

La derivada de la intensidad media de una región se calcula a partir de la derivada de la intensidad en cada pixel de la región

$$\mu_2(R') = \frac{1}{N}\sum_{m'_i \in R'} I_2(m'_i)$$

Aplicando la derivada parcial de la intensidad (ecuación 4.53) obtenemos

$$\frac{\partial}{\partial z}\{\mu_2(R')\} = \frac{1}{N}\sum_{m'_i \in R'}\frac{\partial}{\partial z}\{I_2(m'_i)\} = \frac{1}{N}\sum_{m'_i \in R'} I'_2(m'_i)\frac{\partial}{\partial z}\{m'_i\} \tag{4.54}$$

Finalmente, podemos reescribir esta expresión con la siguiente notación, a fin de simplificar las fórmulas posteriores

$$\frac{\partial}{\partial z}\{\mu_2(R')\} = \mu'_2(R') \tag{4.55}$$

Se podría considerar que $\mu'_2(R')$ es la media del gradiente de intensidad en la región $R'$, de tal modo que el gradiente es calculado en la dirección de la línea epipolar que corresponde a cada pixel de la región.

### 4.A.4 Derivada parcial de la varianza de la intensidad de la región

Para obtener la derivada parcial de la desviación típica, comenzamos con la derivada parcial de la varianza respecto a $z$.

$$\frac{\partial}{\partial z}\left\{\sigma_2^2(R')\right\} = \frac{1}{N}\sum_{m_i'\in R'}\frac{\partial}{\partial z}\left\{\left[I_2(m_i') - \mu_2(R')\right]^2\right\}$$

Con las derivadas de los apartados anteriores (4.53, y 4.55), obtenemos

$$\frac{\partial}{\partial z}\left\{\sigma_2^2(R')\right\} = \frac{2}{N}\sum_{m_i'\in R'}\left(I_2(m_i') - \mu_2(R')\right)\left(I_2'(m_i')\frac{\partial m_i'}{\partial z} - \mu_2'(R')\right)$$

### 4.A.5 Derivada parcial de la desviación típica de la intensidad

La desviación típica es la raíz cuadrada de la varianza,

$$\sigma_2(R') = \sqrt{\sigma_2^2(R')} \tag{4.56}$$

y por tanto,

$$\frac{\partial}{\partial z}\left\{\sigma_2(R')\right\} = \frac{1}{2\sqrt{\sigma_2^2(R')}}\frac{\partial}{\partial z}\left\{\sigma_2^2(R')\right\} = \frac{1}{2\sigma_2(R')}\frac{\partial}{\partial z}\left\{\sigma_2^2(R')\right\}$$

Con la derivada del apartado anterior obtenemos

$$\frac{\partial}{\partial z}\left\{\sigma_2(R')\right\} = \frac{1}{N\sigma_2(R')}\sum_{m_i'\in R'}\left(I_2(m_i') - \mu_2(R')\right)\left(I_2'(m_i') - \mu_2'(R')\right) \tag{4.57}$$

Poniendo la desviación típica, $\sigma_2(R')$, junto con la diferencia de intensidades, obtenemos la intensidad normalizada de media cero

$$\frac{\partial}{\partial z}\left\{\sigma_2(R')\right\} = \frac{1}{N}\sum_{m_i'\in R'}\mathcal{I}_2(m_i')\left(I_2'(m_i')\frac{\partial m_i'}{\partial z} - \mu_2'(R')\right)$$

Esta ecuación calcula la suma de intensidades normalizadas de media cero multiplicadas por sus gradientes direccionales de media cero en los pixeles de $R'$. Puede dividirse en dos sumatorios

$$\frac{\partial}{\partial z}\left\{\sigma_2(R')\right\} = \frac{1}{N}\sum_{m_i'\in R'}\mathcal{I}_2(m_i')I_2'(m_i')\frac{\partial m_i'}{\partial z} - \frac{1}{N}\sum_{m_j'\in R'}\mathcal{I}_2(m_j')\mu_2(R')$$

Nótese que $\mu_2(R')$ no depende del pixel $m'_j$ del sumatorio, ya que es un cálculo que se realiza sobre el total de la región. Por tanto, se puede sacar del segundo sumatorio, quedando

$$\frac{\partial}{\partial z}\left\{\sigma_2(R')\right\} = \frac{1}{N}\sum_{m'_i \in R'} \mathcal{I}_2(m'_i) I'_2(m'_i)\frac{\partial m'_i}{\partial z} - \frac{1}{N}\mu_2(R')\sum_{m'_j \in R'} \mathcal{I}_2(m'_j) \tag{4.58}$$

La suma de todas las intensidades normalizadas de media cero del total de los pixeles de la región es cero por definición,

$$\sum_{m'_j \in R'} \mathcal{I}_2(m'_j) = 0 \tag{4.59}$$

de modo que el segundo término de la ecuación 4.58 es también nulo. Por tanto, la derivada parcial de la desviación típica es

$$\frac{\partial}{\partial z}\left\{\sigma_2(R')\right\} = \frac{1}{N}\sum_{m'_i \in R'} \mathcal{I}_2(m'_i) I'_2(m'_i)\frac{\partial m'_i}{\partial z} \tag{4.60}$$

### 4.A.6 Definición del *gradiente medio ponderado*

La ecuación 4.60 es la suma de los productos de las intensidades de los pixeles de la región por sus gradientes en la dirección de la línea epipolar. Las intensidades están normalizadas, de modo que sirven como peso que pondera cada uno de los gradientes direccionales de los pixeles de la región $R'$.

Llamaremos a este término el *gradiente medio ponderado*, y lo denotaremos por $\mathcal{H}$.

$$\mathcal{H}(R') = \frac{1}{N}\sum_{m'_i \in R'} \mathcal{I}_2(m'_i) I'_2(m'_i)\frac{\partial m'_i}{\partial z} \tag{4.61}$$

Este término, que se calcula sobre la región correspondiente $R'$, proporciona un valor de gradiente normalizado por las intensidades relativas de los pixeles de $R'$.

### 4.A.7 Derivada parcial de la función de energía

A partir de la definición anterior y de todas las derivadas parciales calculadas en los apartados previos, podemos obtener la derivada de $\mathcal{I}_2(m'_i)$,

$$\begin{aligned}
\frac{\partial}{\partial z}\left\{\mathcal{I}_2\left(m'_i\right)\right\} &= \frac{\partial}{\partial z}\left\{\frac{I_2(m'_i) - \mu_2(R')}{\sigma_2(R')}\right\} = \\
&= \frac{\frac{\partial}{\partial z}\left\{I_2(m'_i) - \mu_2(R')\right\}\sigma_2(R') - \frac{\partial}{\partial z}\left\{\sigma_2(R')\right\}\left(I_2(m'_i) - \mu_2(R')\right)}{\sigma_2^2(R')} = \\
&= \frac{1}{\sigma_2(R')}\left(I'_2(m'_i)\frac{\partial m'_i}{\partial z} - \mu'_2(R') - \mathcal{H}(R')\mathcal{I}_2(m'_i)\right)
\end{aligned}$$

Sustituyendo esta derivada en la ecuación 4.17 obtenemos

$$F_z(R, Z(R)) = \frac{1}{N\sigma_2(R')} \sum_{m_i \in R} \mathcal{I}_1(m_i) \left( \mathcal{I}_2(m_i') \mathcal{H}(R') - I_2'(m_i') \frac{\partial m_i'}{\partial z} + \mu_2'(R') \right) = 0$$

De nuevo, la definición de $\mathcal{I}_1$ (ver equación 4.59) permite simplificar esta ecuación. El término $\mu_2'(R')$ no depende del pixel $m_i$ del sumatorio, de modo que puede eliminarse por la propiedad de $\mathcal{I}_1$,

$$\sum_{m_i \in R} \mathcal{I}_1(m_i) \mu_2'(R') = \mu_2'(R') \sum_{m_i \in R} \mathcal{I}_1(m_i) = 0$$

por tanto

$$F_z(R, Z(R)) = \frac{1}{N\sigma_2(R')} \sum_{m_i \in R} \mathcal{I}_1(m_i) \left( \mathcal{I}_2(m_i') \mathcal{H}(R') - I_2'(m_i') \frac{\partial m_i'}{\partial z} \right) = 0 \qquad (4.62)$$

Finalmente, podemos simplificar la división por $\sigma_2(R')$ (también la división por $N$, pero no lo haremos de momento)

$$\frac{1}{N} \sum_{m_i \in R} \mathcal{I}_1(m_i) \left( \mathcal{I}_2(m_i') \mathcal{H}(R') - I_2'(m_i') \frac{\partial m_i'}{\partial z} \right) = 0 \qquad (4.63)$$

donde $\mathcal{H}(R')$ es el *gradiente medio ponderado* (ecuación 4.61) de la región $R'$.

## 4.B Apéndice: Relación entre incrementos de profundidad y de disparidad

Dado un valor de $z_c = z$, y el valor siguiente $z_c = z + \Delta z$, obtenemos los puntos correspondientes $m_1' = (u_1', v_1')$ y $m_2' = (u_2', v_2')$ mediante la ecuación A.33[3] tales que

$$u_1' = \alpha_u' \frac{k_{11}uz + k_{12}vz + k_{13}z + t_x'}{k_{31}uz + k_{32}vz + k_{33}z + t_z'} + u_0' \qquad (4.64)$$

$$u_2' = \alpha_u' \frac{k_{11}u(z+\Delta z) + k_{12}v(z+\Delta z) + k_{13}(z+\Delta z) + t_x'}{k_{31}u(z+\Delta z) + k_{32}v(z+\Delta z) + k_{33}(z+\Delta z) + t_z'} + u_0' \qquad (4.65)$$

donde los valores $k_{ij}$ dependen de los parámetros de calibración (ecuaciones A.34-A.35).

El incremento de $u'$ producido por el incremento de $z$ es,

$$\begin{aligned} u_2' - u_1' = {} & \alpha_u' \frac{k_{11}u(z+\Delta z) + k_{12}v(z+\Delta z) + k_{13}(z+\Delta z) + t_x'}{k_{31}u(z+\Delta z) + k_{32}v(z+\Delta z) + k_{33}(z+\Delta z) + t_z'} - \\ & \alpha_u' \frac{k_{11}uz + k_{12}vz + k_{13}z + t_x'}{k_{31}uz + k_{32}vz + k_{33}z + t_z'} \end{aligned} \qquad (4.66)$$

[3]La versión desarrollada para el caso general y para el caso de geometría paralela se puede encontrar en las ecuaciones A.37 y A.48.

Llamando $a = k_{11}u + k_{12}v + k_{13}$ y $c = k_{31}u + k_{32}v + k_{33}$, obtenemos:

$$\begin{aligned} \Delta u' &= \alpha'_u \frac{az + a\Delta z + t'_x}{cz + c\Delta z + t'_z} - \alpha'_u \frac{az + t'_x}{cz + t'_z} \\ &= \alpha'_u \frac{(at'_z - ct'_x)\,\Delta z}{(cz + t'_z)^2 + c\,(cz + t'_z)\,\Delta z} \end{aligned} \tag{4.67}$$

De esta ecuación podemos despejar $\Delta z$ de modo que se puede calcular a partir de $\Delta u'$. Fijar $\Delta u'$ significa que podemos calcular el incremento de $z$ en cada iteración del algoritmo, de modo que $\Delta u'$ sea constante,

$$\Delta z = \frac{\Delta u'\,(cz + t'_z)^2}{\alpha'_u\,(at'_z - ct'_x) - \Delta u' c\,(cz + t'_z)}$$

Esto nos permitirá imponer un límite a los incrementos, de modo que $\Delta u'$ no pueda ser mayor que un valor dado, por ejemplo, 1 pixel.

En el caso particular de la geometría paralela, se simplifica enormemente esta operación, ya que $t'_z = 0$ y $c = 1$

$$\Delta z = \frac{\Delta u' z^2}{-\alpha'_u t'_x - \Delta u' z}. \tag{4.68}$$

de modo que el incremento de $z$ ya sólo depende de la profundidad actual, del $\Delta u'$ que se desea fijar (que coincide con el incremento de disparidad $\Delta d$), y de los parámetros de calibración.

## 4.C Apéndice: Límites de la profundidad

Supongamos que el sistema de coordenadas del mundo coincide con el sistema de coordenadas de la cámara de la imagen de referencia, de modo que ambos ejes $X$ coinciden, y la escena se encuentra en el semiespacio de $Z$ positiva. Si suponemos también que los ejes $U$ y $U'$ de ambas cámaras son paralelos a ellos y en la misma dirección, entonces $\alpha_u > 0$. Supongamos que el sistema binocular es paralelo y que además la imagen de referencia es la imagen izquierda. En este caso, $t'_x < 0$. Por tanto, la disparidad, $d = u' - u$, es negativa[4],

$$d = u' - u = \frac{\alpha_u t'_x}{Z} < 0.$$

Nótese que el incremento en la disparidad coincide con el incremento en $u'$,

$$\Delta d = d_2 - d_1 = (u'_2 - u) - (u'_1 - u) = u'_2 - u'_1 = \Delta u',$$

[4]En la literatura a veces se define $d = u - u'$, o se considera como imagen de referencia la imagen derecha, de modo que la disparidad es positiva.

y que un incremento positivo en la disparidad supone que se está reduciendo su valor absoluto, esto es:

$$\Delta d > 0 \rightarrow |d + \Delta d| < |d|.$$

Vamos a comprobar la relación entre los incrementos de disparidad (o de $u'$) y los de profundidad.

Si fijamos $\Delta d < 0$, de la ecuación 4.68 se extrae que

$$\begin{aligned}\Delta Z < 0 \quad &\leftrightarrow \quad -\frac{\Delta d\ Z^2}{\Delta d\ Z + \alpha'_u t'_x} < 0 \leftrightarrow \frac{1}{\Delta d\ Z + \alpha'_u t'_x} < 0 \\ &\leftrightarrow \quad \Delta d\ Z + \alpha'_u t'_x < 0 \leftrightarrow \Delta d < \frac{-\alpha'_u t'_x}{Z}.\end{aligned}$$

Como $-\alpha'_u t'_x > 0$, esta condición se cumple siempre, ya que,

$$\Delta d < 0 < \frac{-\alpha'_u t'_x}{Z}.$$

Por tanto, resumimos que, todo incremento negativo de disparidad produce un decremento de profundidad,

$$\Delta d < 0 \longrightarrow \Delta Z < 0$$

Si fijamos $\Delta d > 0$, de la ecuación 4.68 se extrae que

$$\Delta Z > 0 \quad \leftrightarrow \quad -\frac{\Delta d\ Z^2}{\Delta d\ Z + \alpha'_u t'_x} > 0 \leftrightarrow \frac{1}{\Delta d\ Z + \alpha'_u t'_x} < 0 \leftrightarrow \Delta d < \frac{-\alpha'_u t'_x}{Z}.$$

En este caso, como $-\alpha'_u t'_x > 0$, esta condición se puede incumplir cuando

$$\Delta d > \frac{-\alpha'_u t'_x}{Z} > 0.$$

La cuestión entonces es saber cuándo se puede dar este caso. Incluso con $\Delta d < 1$, puede ocurrir que haya una $Z$ lo suficientemente grande $Z > -\alpha'_u t'_x$ tal que

$$1 > \Delta d > \frac{-\alpha'_u t'_x}{Z} > 0,$$

y por tanto, no proporciona un $\Delta Z$ positivo, tal como se desea.

Nótese que a medida que $Z$ aumenta, $d$ tiende a 0, y viceversa, a medida que disminuye $d$, $Z$ tiende al infinito. La cuestión es que cuando $Z$ sea muy grande, el $\Delta d$ que se ha fijado puede ser mayor que $\frac{-\alpha'_u t'_x}{Z}$ y podemos obtener una profundidad negativa.

Como $Z$ no se puede incrementar hasta el infinito, una solución práctica es imponer el límite en el rango de profundidades. Por tanto, cuando $\Delta d\ Z > -\alpha'_u t'_x$, la solución siempre será el valor máximo de profundidad, $Z_{max}$.

De las ecuaciones anteriores se extrae que dado un $\Delta d < 0$ se obtiene un $\Delta Z < 0$, siempre. Sin embargo, dado un $\Delta d > 0$, sólo se obtiene $\Delta Z > 0$ si $\Delta d\ Z > -\alpha'_u t'_x$.

De manera análoga, se puede comprobar el razonamiento es válido también para el caso en el que se tome la imagen derecha como imagen de referencia ($t'_x > 0$, $d = u' - u > 0$).

### 4.C.1 Incremento de profundidad

Para obtener $Z^{t+1}$ a partir de $Z^t$ dado un $\Delta d$ cualquiera,

**si** $\Delta d\ Z^t > -\alpha'_u t'_x$ **entonces**
  $Z^{t+1} \leftarrow \max\left(Z^t + \Delta Z, Z_{min}\right)$
**sino**
  $Z^{t+1} \leftarrow Z_{max}$
**fin si**

donde $Z^t + \Delta Z$ puede calcularse directamente,

$$Z^t + \Delta Z = Z^t + \frac{-\Delta d\left(Z^t\right)^2}{\alpha'_u t'_x + \Delta d\ Z^t} = \frac{\alpha'_u t'_x Z^t}{\alpha'_u t'_x + \Delta d\ Z^t}$$

tanto si $\Delta d$ es positivo como negativo.

# Capítulo 5

# Detección de oclusiones

En el capítulo anterior se introdujo un nuevo método de correspondencia de regiones basado en la minimización de una función de energía, que permite la presencia de discontinuidades en la profundidad. Tal como se comentó en la introducción, uno de los objetivos de esta tesis es detectar también las oclusiones, y es deseable que esta detección sea parte integrante del proceso de correspondencia. Por tanto, aunque la suposición sobre la forma de la profundidad en cada región de la imagen es muy restrictiva todavía, es preciso realizar un estudio de cómo se puede detectar la presencia de oclusiones en este método, antes incluso de considerar otras extensiones posibles del mismo.

## 5.1 Introducción

La necesidad de la detección de las oclusiones queda de manifiesto en los resultados experimentales del capítulo anterior. En los resultados obtenidos para el ejemplo sintético *pirámide*, podemos observar que la profundidad más parecida a la profundidad verdadera es la de la región $R_0$ (véase la tabla 4.1). Esto es debido a que el resto de las regiones contienen pixeles que en la otra imagen están ocultos. Estos pixeles se están emparejando con pixeles en la otra imagen que no son sus correspondientes, de modo que se introduce cierto error en la ecuación de cálculo de $F_z$ y el resultado es un desplazamiento del paso por cero respecto a la solución real. En este capítulo vamos a comprobar esta influencia de las oclusiones y vamos a proponer un modo de detectarlas, de modo que se obtengan mejores correspondencias.

El método obtiene la profundidad de las regiones de una imagen para cada nivel de la escala, donde la profundidad de cada región se calcula independientemente. La consideración de las oclusiones durante el proceso la correspondencia incluye un factor importante en el cálculo de la profundidad de las regiones, ya que un pixel de una región estará oculto o no dependiendo de la profundidad de la región o regiones vecinas. Esto significa que la profundidad de unas regiones influye en el la profundidad de las regiones próximas y por tanto, el proceso de cálculo de la profundidad de cada región ya no es un proceso independiente del de las demás regiones.

En la sección 5.2 se muestra el planteamiento del problema, así como la solución que se propone. En la sección 5.3 se detalla el nuevo algoritmo de correspondencia que incluye la consideración de las oclusiones. En la sección 5.4 se muestran los resultados experimentales del algoritmo propuesto. Finalmente, en la sección 5.5 se resumen las conclusiones que se obtienen del trabajo expuesto en las secciones anteriores.

## 5.2 Planteamiento del problema

En el algoritmo que se ha descrito en el capítulo anterior, la profundidad de cada región se calcula independientemente, de modo que las regiones correspondientes intersectan en las áreas de oclusión. En esta sección vamos a abordar el problema de detectar estas intersecciones entre regiones correspondientes y asignar cada pixel a la región más adecuada. Para ilustrar el problema que se plantea, la figura 5.1 muestra un ejemplo donde dos regiones correspondientes, $R'_a$ y $R'_b$, intersectan en un área común.

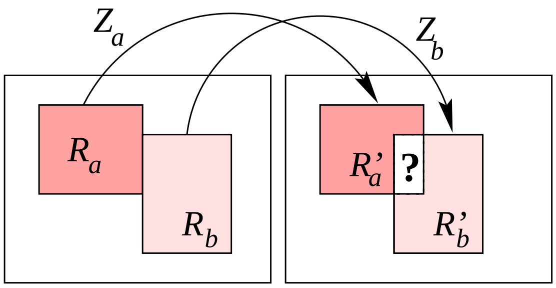



**Figura 5.1:** Planteamiento del problema de la detección de oclusiones.

Por la restricción de unicidad, los pixeles del área de intersección sólo pueden corresponderse con una de las dos regiones, mientras que la otra región tendrá una zona sin correspondencia. La información de profundidad de que disponemos en todo momento, es decir, $Z_a$ y $Z_b$, nos puede ayudar a decidir cuál de las dos regiones contiene un área sin correspondencia.

El algoritmo que se propone en esta sección consiste en establecer la correspondencia de estas áreas con la región más cercana ($Z$ menor), e indicar que los pixeles correspondientes de las otras regiones están ocultos en la segunda imagen.

El algoritmo de correspondencia que además detecte las oclusiones deberá tener en cuenta las siguientes características:

- Cuando dos regiones tienen correspondencias cuya intersección no es vacía, el área de intersección corresponde a la región más cercana, mientras que los pixeles de la región más lejana no tienen correspondencia, porque el área correspondiente está oculta.

- Aunque no es habitual, puede haber más de dos regiones cuyas regiones correspondientes intersecten en un mismo pixel[1] de $I_2$. A menudo las regiones que ocultan a otras son adyacentes a ellas, pero no es así necesariamente. Por lo tanto, no impondremos ninguna restricción en estos dos sentidos.

- La comparación entre regiones no debería tener en cuenta los pixeles de $R$ que se corresponden con pixeles ocultos en $R'$. A estos pixeles los marcaremos con la etiqueta *oculto*.

- La comparación entre regiones no debería tener en cuenta los pixeles de $R$ que se correspondan con puntos fuera de los límites de $I_2$. A estos pixeles los marcaremos con la etiqueta *fuera-de-límites*.

- Los pixeles etiquetados como *ocultos* y *fuera-de-límites* pueden variar de una iteración a otra.

Para poder manejar todos los casos, definimos un *estado del pixel* tal que en cada iteración del algoritmo todos los pixeles de la imagen de referencia se marcan como *activo*, *oculto* o *fuera-de-límites*. Solamente los pixeles activos serán considerados en los cálculos estadísticos de la región (intensidades normalizadas de media cero, gradiente medio ponderado, etc.) y por tanto, en el cálculo de los incrementos/decrementos de profundidad.

Por lo tanto, el nuevo algoritmo de correspondencia ya no puede calcular las profundidades de las regiones independientemente, puesto que las profundidades de las regiones dependen de las posibles oclusiones que puedan ejercer sobre ellas las regiones vecinas. Las posibles oclusiones dependen a su vez de las profundidades estimadas, y por lo tanto, el algoritmo que detecte correspondencias y oclusiones ha de ser un algoritmo cooperativo que modifique las profundidades lentamente para converger al conjunto de profundidades y oclusiones de mínima energía.

## 5.3 Algoritmo de correspondencia

El objetivo del algoritmo sigue siendo buscar el paso por cero de la función $F_z$ para cada región $R$ de la imagen $I_1$, de modo que se detenga cuando $F_z(R) = 0, \forall R \subset I_1$. El esquema multiescala se mantiene igual, pero el algoritmo de correspondencia de cada nivel ha de tener en cuenta las oclusiones para mejorar en lo posible la correspondencia. Este algoritmo se podría dividir en dos pasos principales, como se muestra en la figura 5.2:

[1] Por ejemplo, podemos pensar en una escena con tres columnas en línea a distintas distancias, dos de las cuales se ocultan tras la columna más cercana en una de las imágenes.

- *Transformar* la segunda imagen en una réplica $T$ de la imagen de referencia, dada una profundidad $Z$.
- *Incrementar* (o decrementar) $Z$ a partir de la comparación de las imágenes $I_1$ y $T$.

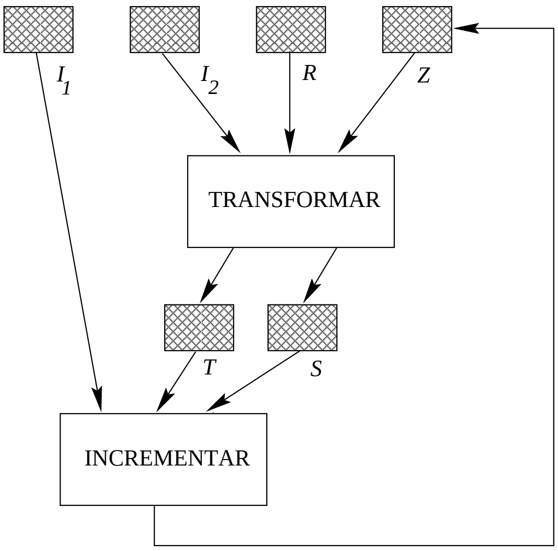



**Figura 5.2:** Proceso de correspondencia en cada nivel.

Los pixeles de $I_1$ cuyo pixel correspondiente está oculto en $I_2$, introducen imprecisiones en el cálculo del error de similitud entre regiones. Por tanto, los pixeles de una región que hayan sido marcados como ocultos, no se tienen en cuenta en los cálculos, para mejorar los resultados. Debido a esto, la posición del paso por cero puede variar ligeramente de una iteración a otra.

Mediante este procedimiento, la profundidad de una región sí que influye en las profundidades de otras regiones, a través de las oclusiones. Por tanto, todas las profundidades deben calcularse al mismo tiempo, en lugar de secuencialmente, y deben incrementarse o decrementarse muy poco en cada iteración para que el algoritmo converja a una solución global. El proceso, que se detalla en el algoritmo 5.1, se detendrá cuando ninguna profundidad sea incrementada o decrementada, es decir, cuando $Z$ converja.

Sea $S$ el *mapa de estados* de los pixeles de la imagen de referencia. El estado puede calcularse mientras se *transforma* la imagen $I_2$ en $T$ de acuerdo con la profundidad actual. El método para reconocer los pixeles cuyos puntos correspondientes están ocultos dado un mapa de profundidades se lleva a cabo manteniendo una tabla de pixeles visitados. El pixel correspondiente (en $I_2$) de cada pixel de una región se marca como visitado mientras que se realiza la operación *transformar*. Si un pixel es visitado varias veces, todos los pixeles "visitantes" excepto uno deben marcarse como *ocultos*.

La operación *transformar* se muestra en el algoritmo 5.2, aunque no se incluyen los detalles de control de los pixeles visitados, para mayor claridad. En el siguiente apartado

**Algoritmo 5.1 CORRESPONDENCIA2**. Cálculo de $Z$ del nivel $k$, con detección de oclusiones.

**Entradas:** $Z_{min}, Z_{max}, \Delta d_{ini}, \Delta d_{\epsilon}$, nivel $k$ de $I_1$ e $I_2$, y $Z$ inicial del nivel $k$.
**Salidas:** $Z$ final del nivel $k$.
Sea $\mathcal{R}$ el conjunto de regiones resultante de la segmentación de $I_1$.
TRANSFORMAR $I_2$ en $T$ y $S$, de acuerdo con $Z$
**para cada** región $R_i \in R$ **hacer**
Calcular $F_z^1(R_i)$ a partir de $I_1, T$, usando los pixeles de $R_i$ activos en $S$
Determinar la dirección y magnitud de $\Delta d_i$
$Z_i^1 \leftarrow$ IncZ($Z^0, \Delta d_i$, parámetros de calibración, $Z_{min}, Z_{max}$)
**fin para**
**mientras** $Z^t$ no converja **hacer**
TRANSFORMAR $I_2$ en $T$ y $S$, de acuerdo con $Z^t$
$Z^{t+1} \leftarrow$ INCREMENTAR$^t$ $Z^t$ de acuerdo con $I_1, T, S$
**fin mientras**

**Algoritmo 5.2 TRANSFORMAR**. Obtener $T, S$ a partir de $I_2$, dada $Z$.

**Entradas:** nivel $k$ de la pirámide de $I_2$, $Z$ actual, lista de regiones $\mathcal{R}$.
**Salidas:** réplica $T$, y mapa de estados $S$.
Ordenar la lista de regiones $\mathcal{R}$ en orden creciente de $Z$
**para cada** región $R_i \in \mathcal{R}$ **hacer**
**para cada** pixel $m \in R_i$ **hacer**
$m' \leftarrow f_{12}(m, Z(m))$
$S(m) \leftarrow$ estado de $m'$
**si** $S(m) =$ ACTIVO **entonces**
$T(m) \leftarrow I_2(m')$
**fin si**
**fin para**
**fin para**

se detalla el control de los pixeles visitados, así como la obtención del estado de los pixeles "visitantes".

Por otro lado, la operación *incrementar* se detalla en el algoritmo 5.3. Esta operación consiste en recalcular $F_z$ en cada iteración (como se explicó en la sección 4.4), y aplicar los nuevos incrementos de profundidad mediante la función IncZ, que también controla que no se sobrepasen los límites de la profundidad, $Z_{min}$ y $Z_{max}$.

La operación *incrementar* además, determina cuándo converge la profundidad de cada región. El algoritmo 5.1 se detiene cuando todas las profundidades convergen a la vez, pero hay dos alternativas para establecer esta convergencia global:

- Cuando la $Z$ de una región converge, no se vuelve a procesar.

**Algoritmo 5.3 INCREMENTAR**. Calcular nueva $Z$ dados $I_1, T, S$ (iteración $t$)

**Entradas:** $Z_{min}, Z_{max}, \Delta d_\epsilon$, niveles $k$ de $I_1$, $T$ y $S$, y $Z_i^{t-1}, Z_i^{t-2}, \Delta d_i^{t-1}, \forall R_i \in \mathcal{R}$.
**Salidas:** $Z$ de la iteración $t$.

```
para cada región R_i ∈ 𝓡 hacer
   Recalcular F_z^t(R_i) a partir de los pixeles activos de R_i
   si F_z^t(R_i) = 0 entonces
      R_i converge
   sino si sign(F_z^t(R_i)) ≠ sign (F_z^{t-1}(R_i)) entonces { Etapa de refinamiento }
      si Δd ≥ Δd_ε entonces
         Δd_i^t ← Δd_i^{t-1}/2
         Z_i^t ← IncZ(Z_i^{t-2}, Δd_i^t, parámetros de calibración, Z_min, Z_max)
         si Z_i^t = Z_i^{t-2} entonces
            R_i converge
         fin si
      sino
         Z_i^t ← Interpolar(Z^{t-1}, F_z^{t-1}, Z^t, F_z^t)
         R_i converge
      fin si
   sino { Etapa de acercamiento }
      Δd_i^t ← Δd_i^{t-1}
      Z_i^t ← IncZ(Z_i^{t-1}, Δd_i^t, parámetros de calibración, Z_min, Z_max)
      si Z_i^t = Z_i^{t-1} entonces
         R_i converge
      fin si
   fin si
fin para
```

- Cuando la $Z$ de una región converge, si en alguna de las siguientes iteraciones ocurre que $F_z(R_i) \neq 0$, se vuelve a recalcular $\Delta d_i$ y $Z(R_i)$, y se continua su procesamiento.

Ambas opciones tienen inconvenientes. Supongamos que una región $R_i$ converge muy rápido y sus regiones adyacentes tardan más iteraciones que ella. Con la primera opción, las profundidades obtenidas de sus regiones adyacentes puede producir oclusiones en $R_i$ que probablemente no se hayan tenido en cuenta. Para evitar esto se puede reducir $\Delta d$, pero entonces crece demasiado el número de iteraciones.

La segunda opción, además de que genera un número de iteraciones mucho mayor, tiene el inconveniente de que se pueden producir ciclos debido a la influencia de unas regiones sobre otras, y por tanto no se puede asegurar su convergencia. Aunque se han probado soluciones a este problema, como es parar cuando el error global[2] de $F_z$ sea menor que un valor dado, con resultados bastante satisfactorios en cuanto a la convergencia, sería preferible poder controlar el nivel de precisión en todas las profundidades de todas las regiones.

Por tanto, queremos modificar el algoritmo de manera que la convergencia se garantice y que podamos obtener la solución con la precisión que se requiera, de modo que el número de iteraciones no sea demasiado grande.

El algoritmo que se propone consiste en repetir el bucle **mientras** del algoritmo 5.1 varias veces, aplicando la primera opción de la lista en cada iteración. Por tanto, tenemos dos bucles anidados: el más exterior se repite un número determinado de veces, y el más interior converge cuando todas las $Z$ de todas las regiones hayan convergido, de modo que cuando la $Z$ de una región converge, ya no se sigue procesando dicha región.

Este algoritmo tiene la ventaja de que se pueden aplicar $\Delta d_\epsilon$ diferentes en cada iteración del bucle exterior, realizando así una aproximación *coarse-to-fine* a la convergencia global. En este caso, el bucle exterior debe ejecutarse tantas veces como valores de $\Delta d$ queramos hasta obtener la solución con la precisión requerida, $\Delta d_\epsilon$. Nuestra propuesta es empezar con el $\Delta d$ inicial, y reducirlo en cada iteración hasta llegar a un valor menor que $\Delta d_\epsilon$ (algoritmo 5.4). De este modo tendremos que la primera iteración es una etapa de acercamiento, mientras que las demás iteraciones son sucesivas etapas de refinamiento.

Por tanto, al trasladar el control de la precisión al algoritmo principal de correspondencia de cada nivel, la operación *incrementar* se simplifica, tal como se muestra en el algoritmo 5.5.

---

[2]El error global puede ser calculado como suma de valores absolutos de $F_z$ siempre y cuando el valor de $Z$ de todas las regiones se encuentre en las inmediaciones del paso por cero, es decir, que al menos se haya iniciado la etapa de refinamiento.

**Algoritmo 5.4 CORRESPONDENCIA3**. Cálculo de $Z$ del nivel $k$, con detección de oclusiones.

**Entradas:** $Z_{min}, Z_{max}, \Delta d_{ini}, \Delta d_\epsilon$, nivel $k$ de $I_1$ e $I_2$, y $Z$ inicial del nivel $k$.
**Salidas:** $Z$ final del nivel $k$.
  Sea $\mathcal{R}$ el conjunto de regiones resultante de la segmentación de $I_1$.
  Sea $\Delta d \leftarrow \Delta d_{ini}$
  **mientras** $\Delta d \geq \Delta d_\epsilon$ **hacer**
    TRANSFORMAR $I_2$ en $T$ y $S$, de acuerdo con $Z$
    **para cada** región $R_i \in R$ **hacer**
      Calcular $F_z^1(R_i)$ a partir de $I_1, T$, usando los pixeles de $R_i$ activos en $S$
      Determinar la dirección y magnitud de $\Delta d_i$ a partir de $\Delta d$ y $F_z^1(R_i)$
      $Z_i^1 \leftarrow$ IncZ($Z^0, \Delta d_i$, parámetros de calibración, $Z_{min}, Z_{max}$)
    **fin para**
    **mientras** $Z^t$ no converja **hacer**
      TRANSFORMAR $I_2$ en $T$ y $S$, de acuerdo con $Z^t$
      $Z^{t+1} \leftarrow$ INCREMENTAR2$^t$ $Z^t$ de acuerdo con $I_1, T, S$
    **fin mientras**
    $\Delta d \leftarrow \Delta d/2$
  **fin mientras**

**Algoritmo 5.5 INCREMENTAR2**. Calcular nueva $Z$ dados $I_1, T, S$ (iteración $t$)

**Entradas:** $Z_{min}, Z_{max}, \Delta d_\epsilon$, niveles $k$ de $I_1$, $T$ y $S$, y $Z_i^{t-1}, Z_i^{t-2}, \Delta d_i^{t-1}, \forall R_i \in \mathcal{R}$.
**Salidas:** $Z$ de la iteración $t$.
  **para cada** región $R_i \in \mathcal{R}$ **hacer**
    Recalcular $F_z^t(R_i)$ a partir de los pixeles activos de $R_i$
    **si** $F_z^t(R_i) = 0$ **entonces**
      $R_i$ converge
    **sino si** sign($F_z^t(R_i)$) $\neq$ sign ($F_z^{t-1}(R_i)$) **entonces** { *Paso por cero alcanzado* }
      $Z_i^t \leftarrow$ Interpolar($Z^{t-1}, F_z^{t-1}, Z^t, F_z^t$)
      $R_i$ converge
    **sino** { *Etapa de acercamiento* }
      $\Delta d_i^t \leftarrow \Delta d_i^{t-1}$
      $Z_i^t \leftarrow$ IncZ($Z_i^{t-1}, \Delta d_i^t$, parámetros de calibración, $Z_{min}, Z_{max}$)
      **si** $Z_i^t = Z_i^{t-1}$ **entonces**
        $R_i$ converge
      **fin si**
    **fin si**
  **fin para**

### 5.3.1 La tabla de pixeles visitados

Necesitamos una tabla de pixeles visitados para controlar si un pixel es visitado varias veces. El pixel "visitante" de $Z$ menor es el que establece la correspondencia (estado *activo*) y el resto de "visitantes" serán etiquetados con el estado *oculto*.

Una forma fácil de resolver este problema sin tener que registrar todos los pixeles visitantes es ordenar las regiones por orden creciente de $Z$, de modo que se procesen primero las regiones más cercanas. De este modo, el primer pixel de $I_1$ que marca cada pixel de $I_2$ como visitado establece la correspondencia y consigue un estado *activo* en $S$. A continuación, los estados de los pixeles de otras regiones que intenten emparejarse con el mismo pixel, se marcan como pixeles *ocultos* en $S$.

Dado que los pixeles correspondientes en $I_2$ se obtienen en coordenadas no enteras, la resolución de la tabla de pixeles visitados influye en la precisión del método. La resolución más burda es mantener un elemento de la tabla por cada pixel de la imagen $I_2$. La marcación de un pixel como "visitado" se realiza por redondeo de las coordenadas reales de $m'$ a enteras, e igualmente la obtención del estado de un pixel.

La resolución se puede refinar tanto como se quiera, manteniendo un array de $n \times m$ elementos por pixel. La marcación como "visitado" y la obtención del estado de un pixel en este caso, se realiza considerando los $n \times m$ elementos del array alrededor del elemento más cercano al punto real. En la figura 5.3 se muestra un ejemplo de cómo se marcarían $5 \times 5$ secciones por pixel.

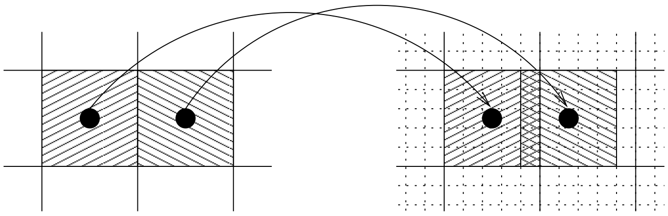


**Figura 5.3:** Resolución de la tabla de pixeles visitados.

En los experimentos se han probado varios métodos para determinar si un pixel está *oculto* o no a partir de la tabla de pixeles visitados, cuando la resolución de la tabla es de $n \times m$ secciones por pixel.

**Condición 1:** Un pixel se considera oculto si al menos el 50% de las $n \times m$ secciones han sido visitadas.

**Condición 2:** Un pixel se considera oculto si al menos 1 de las $n \times m$ secciones han sido visitadas.

Los numerosos experimentos realizados mostraron que la segunda condición es más efectiva y proporciona profundidades más cercanas a las reales, mientras que la primera condición proporcionaba resultados que, aunque con alguna mejoría, se parecían más a los resultados sin detección de oclusiones.

Sin embargo, al término del proceso de correspondencia, muchos de los pixeles marcados como ocultos lo son en un porcentaje muy pequeño y, como también es deseable obtener un mapa de oclusiones de la imagen, este hecho proporciona muchas falsas oclusiones. Por tanto, para construir el mapa de oclusiones final se puede utilizar la primera condición.

### 5.3.2 Análisis del coste

Empezaremos por calcular el coste de las dos operaciones principales del algoritmo que son la operación *transformar* (algoritmo 5.2) y la operación *incrementar* (algoritmo 5.5).

La operación *transformar* es lineal respecto al número de pixeles de la imagen, es decir, que dada una imagen de $N_u \times N_v$ pixeles, es $O(N_u N_v)$. La marcación del estado de cada pixel depende, no obstante, del número de secciones por pixel ($n \times m$) que se utilicen en la representación de la tabla de pixeles visitados. Por tanto, el coste total de la operación es, $O(N_u N_v nm)$.

El cálculo más costoso de la operación *incrementar* es recalcular para cada región su $F_z$ a partir de los pixeles activos de la región, que en el peor caso serán todo los pixeles de la región. Por tanto, la única diferencia de este cálculo con el cálculo de $F_z$ que se hacía en el capítulo anterior, es el chequeo del estado de cada pixel, que es $O(nm)$ en el peor caso. Por tanto, el cálculo de $F_z$ de una región es $O(Nnm)$, donde $N$ es el número de pixeles de la región.

Dado que las regiones forman una partición de la imagen de referencia y el cálculo de $F_z$ se realiza en el peor caso para todas las regiones, el coste en el peor caso de los cálculos del valor de $F_z$ depende de la suma del número de pixeles de todas las regiones ($N_u N_v$), y del número de secciones por pixel $nm$, es decir, $O(N_u N_v nm)$.

Estas dos operaciones son las que se repiten sucesivamente en el algoritmo 5.4 hasta alcanzar la convergencia. El bucle más exterior de este algoritmo se repite tantas veces como veces se puede reducir $\Delta d_{ini}$ hasta alcanzar un valor menor que $\Delta d_\epsilon$. Si asumimos que la reducción consiste en dividir $\Delta d_{ini}$ por dos, el número de iteraciones es

$$n_1 = \log_2 \frac{\Delta d_{ini}}{\Delta d_\epsilon} + 1$$

De los dos bucles interiores, el bucle **para** tiene un coste del mismo orden que la operación *incrementar*. El bucle **mientras**, sin embargo, no se detiene hasta que convergen todas las profundidades de todas las regiones. Como el incremento de disparidad es idéntico para todas, el bucle tendrá tantas iteraciones como incrementos necesite la región que más tarde en converger. En el caso peor, existirá alguna región que comienza con un valor de $Z$ extremo, por ejemplo $Z = Z_{min}$ (y equivalentemente $d = d_{max}$), y llega al otro extremo, $Z = Z_{max}$ (y $d = d_{min}$), a intervalos regulares de $|\Delta d|$. Es decir, que el número de iteraciones que precisa esa región es

$$n_2 = \frac{d_{max} - d_{min}}{|\Delta d|} = \frac{|\mathcal{D}|}{|\Delta d|}$$

El valor absoluto de $\Delta d$ depende de la iteración $i$ del bucle más exterior en que nos encontremos. Es decir, que en cada iteración $i$

$$|\Delta d| = \frac{\Delta d_{ini}}{2^i}$$

donde $i$ varía entre 0 y $n_1$, calculado previamente.

Por tanto, el número total de veces que se ejecutan las operaciones *transformar* e *incrementar* del bucle más interior es

$$n_T = \sum_{i=1}^{n_1} (n_2 + 1) = n_1 + \sum_{i=1}^{n_1} \frac{|\mathcal{D}| 2^i}{\Delta d_{ini}}$$

La suma de los $n$ primeros términos de una serie geométrica

$$a_1 + a_1 r + a_1 r^2 + ... + a_1 r^{n-1} + ...$$

es $(a_1 - a_1 r^n)/(1 - r)$ [Díaz Hernando, 1991]. Por tanto,

$$n_T = n_1 + \frac{|\mathcal{D}|}{\Delta d_{ini}} (2^{n_1} - 1)$$

donde

$$2^{n_1} = 2^{\log_2 \frac{\Delta d_{ini}}{\Delta d_\epsilon} + 1} = 2 \ 2^{\log_2 \frac{\Delta d_{ini}}{\Delta d_\epsilon}} = 2 \frac{\Delta d_{ini}}{\Delta d_\epsilon}$$

Dados los valores habituales de las variables de que depende $n_T$, el término dominante es el que tiene $2^{n_1}$. Por tanto, despreciando los términos de menor orden, obtenemos

$$n_T \approx 2 \frac{|\mathcal{D}|}{\Delta d_\epsilon}$$

Sabiendo que las operaciones *transformar* e *incrementar* se repiten $n_T$ veces, el coste total del algoritmo en el caso peor sería

$$O\left(2 \frac{|\mathcal{D}|}{\Delta d_\epsilon} N_u N_v n m\right)$$

Este es el coste para el nivel de mayor resolución del esquema multiescala. Para calcular el coste de los demás niveles se pueden hacer las mismas consideraciones que se hicieron en el análisis del coste en el capítulo anterior respecto a $|\mathcal{D}|$, $N_u$ y $N_v$.

Por último, queda señalar que para calcular este coste teórico se está considerando el peor caso que consistiría que en cada iteración del bucle exterior del algoritmo hay alguna región que recorre todo el intervalo de profundidades de un extremo a otro. Según esto, la iteración más costosa del bucle exterior sería la última, donde los incrementos son más pequeños. Esto no ocurre así en la práctica, donde la primera iteración suele ser la más costosa (más iteraciones del bucle interior) ya que corresponde a la etapa de *acercamiento*, mientras que las demás iteraciones suelen ser menos costosas (pocas iteraciones del bucle interior) por tratarse de la etapa de *refinamiento*. Se podrá observar este comportamiento del algoritmo en los experimentos que se muestran a continuación.

## 5.4 Resultados experimentales

### 5.4.1 Resultados con imágenes sintéticas

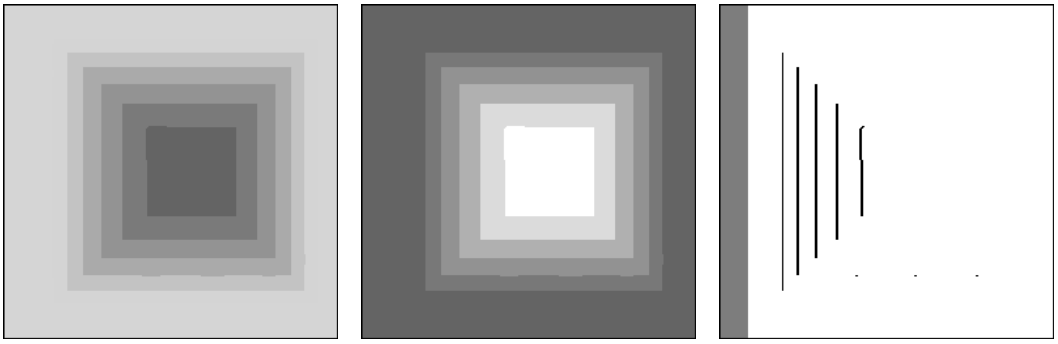
De nuevo nos centramos en el ejemplo sintético *pirámide*, para comparar los resultados de este nuevo método con los resultados del algoritmo anterior. En la tabla 5.1 se pueden observar los resultados obtenidos utilizando distintas resoluciones en la tabla de pixeles visitados. Nótese que ya en la columna relativa a $1 \times 1$ secciones por pixel en la tabla de pixeles visitados, los resultados mejoran en casi todas las regiones respecto al algoritmo que no tiene en cuenta las oclusiones, salvo la región $R_0$ (que por no tener oclusiones ya daba un buen resultado) y las regiones $R_1$ y $R_2$. Gracias a esta mejora el error relativo medio ($\overline{e}_r$) obtenido es del 2.97%. En las columnas de mayor resolución se mejoran los resultados de las regiones $R_1$ y $R_2$ de modo que el error relativo medio disminuye hasta el 0.62%.

| Regiones | $Z$ real | Profundidades ($Z$) y oclusiones ($O$) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Sin ocl. | $1 \times 1$ | | $3 \times 1$ | | $5 \times 1$ | |
| | | $Z$ | $Z$ | $O$ | $Z$ | $O$ | $Z$ | $O$ |
| $R_0$ | 82 | 81.5722 | 81.5722 | 0 | 81.5722 | 0 | 81.5722 | 0 |
| $R_1$ | 89 | 84.7316 | 84.4689 | 69 | 86.4655 | 0 | 88.2683 | 138 |
| $R_2$ | 96 | 90.5292 | 86.4655 | 0 | 88.2682 | 242 | 96.3125 | 208 |
| $R_3$ | 103 | 98.9442 | 103.938 | 665 | 103.938 | 665 | 103.604 | 266 |
| $R_4$ | 110 | 105.386 | 111.203 | 324 | 111.203 | 324 | 111.203 | 324 |
| $R_5$ | 117 | 113.892 | 116.579 | 182 | 116.579 | 182 | 116.579 | 182 |
| fondo | | 116.994 | 116.995 | 0 | 116.994 | 0 | 116.894 | 0 |
| $\overline{e}_r$ (%) | | 3.78 | 2.97 | | 2.31 | | 0.62 | |

**Tabla 5.1:** Resultados del ejemplo sintético *pirámide*: comparación del algoritmo con oclusiones con el algoritmo sin oclusiones, $\Delta d = 0.5$, $\Delta d_\epsilon = 0.001$. $Z$ indica profundidad y $O$ indica número de pixeles ocultos.

Los resultados con $\Delta d = 0.5$, $\Delta d_\epsilon = 0.001$ y $5 \times 1$ secciones por pixel en la tabla de pixeles visitados se muestran en la figura 5.4. Las disparidades resultantes varían entre

21.9 y 31.4 pixeles. En los mapas de oclusiones se puede observar el estado final de los pixeles de la imagen izquierda: *activo* (blanco), *oculto* (negro) y *fuera-de-límites* (gris).



(a) Mapa de profundidades (b) Mapa de disparidades (c) Mapa de oclusiones

**Figura 5.4:** Mapas de profundidades, disparidades y oclusiones del par de imágenes *pirámide*. El mapa de oclusiones indica el estado final de cada pixel: *activo* (blanco), *oculto* (negro) y *fuera-de-límites* (gris).

En la figura 5.5 se pueden comparar las gráficas con y sin oclusiones de una de las regiones del ejemplo *pirámide*. Los pasos por cero están más cerca de la solución real cuando se incorpora la detección de oclusiones.

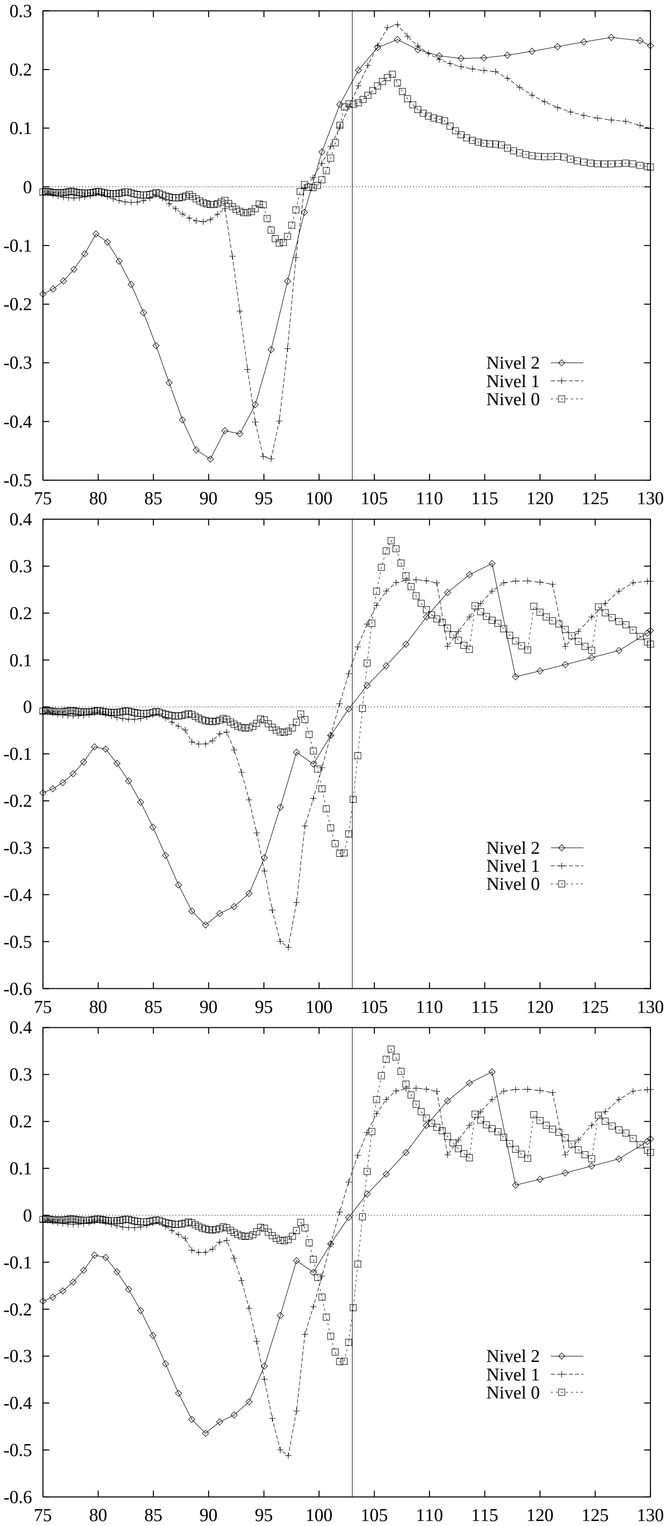
Dada una región, para poder calcular las gráficas de $F_z$ considerando las oclusiones, hay que asignar profundidades a las otras regiones para poder calcular en cada caso las oclusiones pertinentes. Para ello, hemos considerado dos casos:

- las profundidades de las otras regiones han sido fijadas a su profundidad real, que llamaremos *caso ideal*.
- las profundidades de las otras regiones han sido fijadas a la profundidad obtenida por el algoritmo, que llamaremos *caso real*.

En las gráficas con detección de oclusiones se puede observar que hay un patrón repetitivo más marcado, que es debido a la utilización de redondeo de las coordenadas reales de $m'$ a enteras en la tabla de pixeles visitados. En la figura 5.6 se muestra la forma de la función $F_z$ de otras dos regiones en el caso ideal.

Los resultados mejoran mucho, cuando se aumenta la resolución de la tabla de pixeles visitados. Esto es debido principalmente a dos regiones, $R_1$ y $R_2$, que obtienen un paso por cero diferente de la solución. En las figuras 5.7 y 5.8 se pueden ver las gráficas relativas a la región $R_2$, con y sin detección de oclusiones, y con distintas resoluciones.

Como se puede observar, hay más de un paso por cero en el nivel 0 de la $F_z$ de la región $R_2$, uno de los cuales se ajusta bastante a la solución real (línea vertical). El




(a) Sin detectar oclusiones, $\Delta d = 0.1$.

(b) Detectando oclusiones, $\Delta d = 0.1$. Caso ideal.

(c) Detectando oclusiones, $\Delta d = 0.1$. Caso real.

**Figura 5.5:** Evolución de $F_z, \forall Z$, de la región $R_3$ del ejemplo *pirámide*: Comparación de resultados con y sin detección de oclusiones ($1 \times 1$ secciones).

(a) $R_4$, $1 \times 1$ secciones/pixel, $\Delta d = 0.1$. Caso ideal.

(b) $R_5$, $1 \times 1$ secciones/pixel, $\Delta d = 0.1$. Caso ideal.

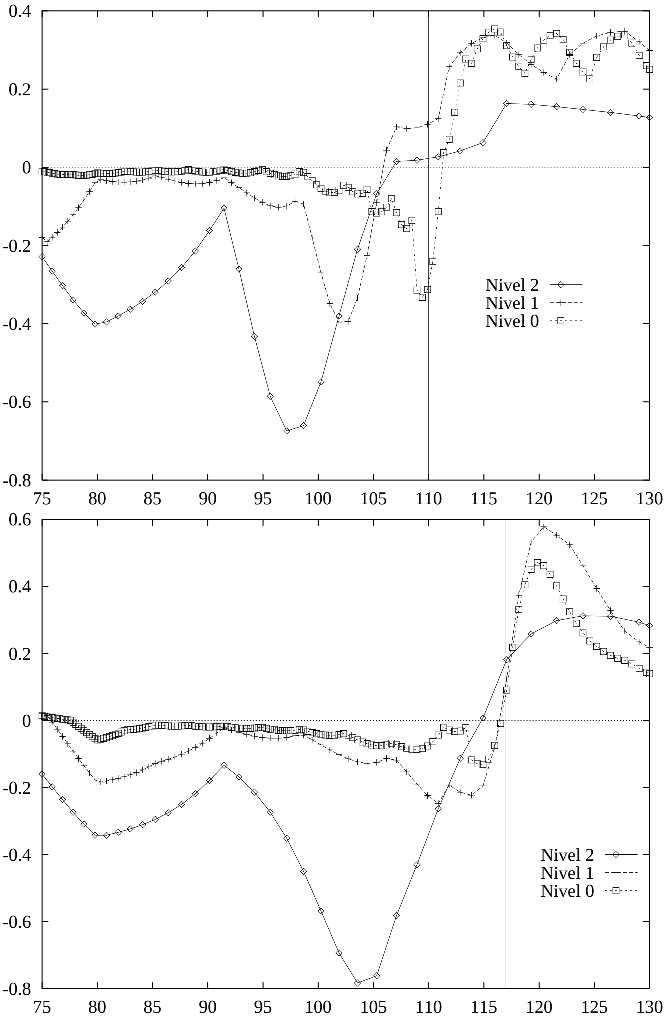



**Figura 5.6:** Evolución de $F_z, \forall Z$, de las regiones $R_4$ y $R_5$ del ejemplo *pirámide* con detección de oclusiones.

(a) Sin detectar oclusiones, $\Delta d = 0.1$.

(b) Detectando oclusiones, $1 \times 1$ secciones/pixel, $\Delta d = 0.1$. Caso ideal.

(c) Detectando oclusiones, $3 \times 1$ secciones/pixel, $\Delta d = 0.1$. Caso ideal.

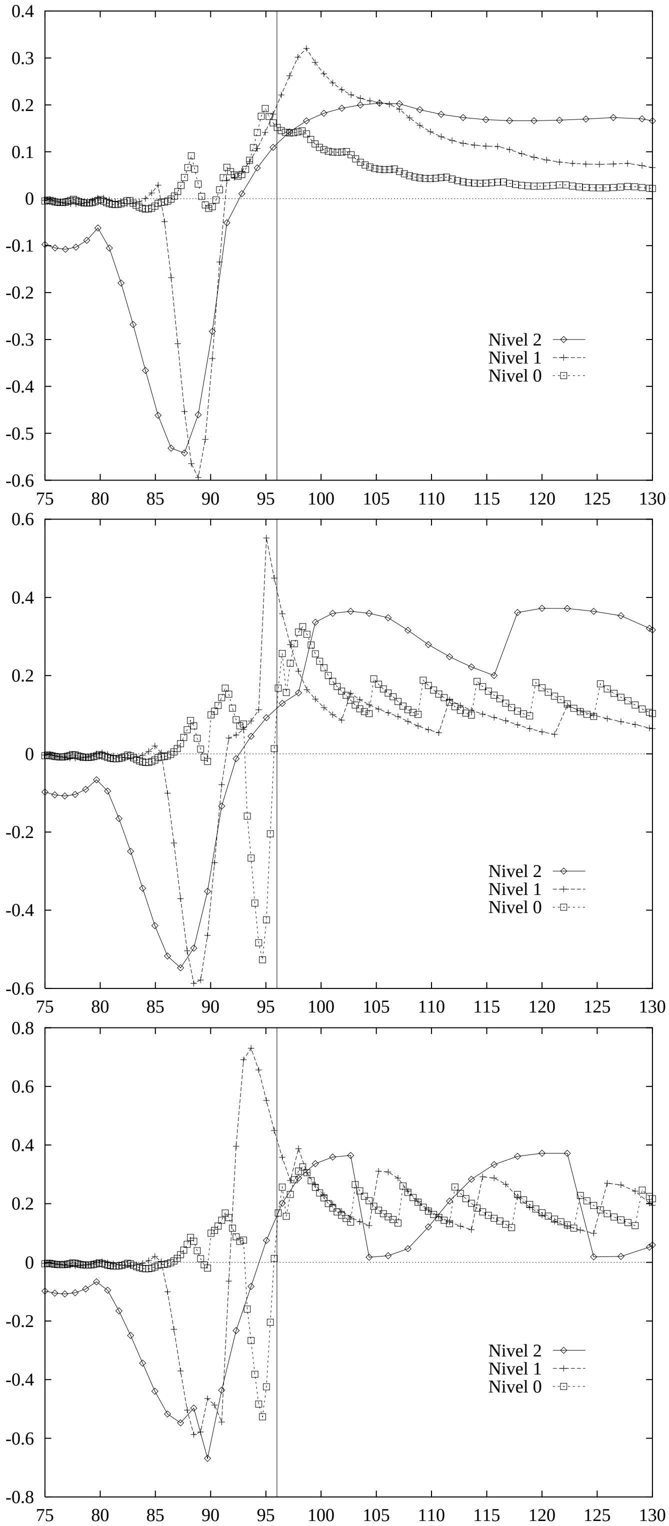



**Figura 5.7:** Evolución de $F_z, \forall Z$, de la región $R_2$ del ejemplo *pirámide*, con y sin detección de oclusiones.

problema estriba en que, con $1 \times 1$ secciones por pixel, el paso por cero en el nivel 1 dirige la solución hacia el paso por 0 de más a la izquierda. Sin embargo, a medida que se aumenta la resolución en la tabla de pixeles visitados, el paso por cero del nivel 1 se mueve ligeramente hacia la derecha.

En cuanto a la convergencia del algoritmo se puede apreciar un aumento considerable de iteraciones con respecto al algoritmo sin oclusiones, como ya se esperaba. En la tabla 5.2 se muestra el número de iteraciones de cada nivel para los casos $\Delta d = \{0.5, 0.1\}$. Cada valor representa el número total de iteraciones del bucle más interior. Por ejemplo, en el nivel 7 del caso $\Delta d = 0.5$ hay 9 iteraciones del bucle exterior y 1 iteración del bucle interior en cada una, lo que suma un mínimo de 9 iteraciones por nivel.

Aunque el número de iteraciones total es similar en todos los casos, se puede apreciar que con $\Delta d = 0.5$ hay menos iteraciones en el nivel 0, por lo que el tiempo de ejecución total es menor.

| Número de iteraciones ($\Delta d_\epsilon = 0.001$) (niveles 7 a 0, respectivamente) | | |
|---|---|---|
| secciones/pixel | $\Delta d = 0.5$ | $\Delta d = 0.1$ |
| $1 \times 1$ | 9, 10, 19, 34, 18, 18, 19, 24 | 7, 8, 17, 38, 15, 18, 24, 33 |
| $3 \times 1$ | 9, 10, 19, 34, 26, 18, 20, 25 | 7, 8, 17, 39, 15, 20, 25, 33 |
| $5 \times 1$ | 9, 10, 18, 33, 18, 19, 19, 22 | 7, 8, 17, 39, 15, 19, 24, 32 |

**Tabla 5.2:** Número de iteraciones en el ejemplo sintético *pirámide* con detección de oclusiones, $\Delta d = \{0.5, 0.1\}$, $\Delta d_\epsilon = 0.001$.

La tabla 5.3 muestra los tiempos relativos de ejecución de todos los casos mostrados en la tabla 5.2. Debido a que la implementación del método no se ha dirigido hacia una implementación eficiente, sino hacia una evaluación de la validez del mismo, los valores que se muestran en la tabla se han calculado dividiendo todos los tiempos por el más pequeño.

| Tiempos relativos ($\Delta d_\epsilon = 0.001$) | | |
|---|---|---|
| secciones/pixel | $\Delta d = 0.5$ | $\Delta d = 0.1$ |
| $1 \times 1$ | 1.000756 | 1.1934587 |
| $3 \times 1$ | 1.03206 | 1.2132377 |
| $5 \times 1$ | 1 | 1.1878999 |

**Tabla 5.3:** Tiempos relativos del ejemplo *pirámide* con detección de oclusiones, y parámetros $\Delta d = \{0.5, 0.1\}$, $\Delta d_\epsilon = 0.001$.

Los resultados para el ejemplo *corridor* son muy similares a los obtenidos sin detección de oclusiones, como se puede observar en la figura 5.9. El mapa de disparidades con detección de oclusiones ha sido obtenido con $\Delta d = 0.5$, $\Delta d_\epsilon = 0.001$ y $11 \times 1$ secciones

(a) Detectando oclusiones, $5 \times 1$ secciones/pixel, $\Delta d = 0.1$. Caso ideal.

(b) Detectando oclusiones, $9 \times 1$ secciones/pixel, $\Delta d = 0.1$. Caso ideal.

(c) Detectando oclusiones, $15 \times 1$ secciones/pixel, $\Delta d = 0.1$. Caso ideal.

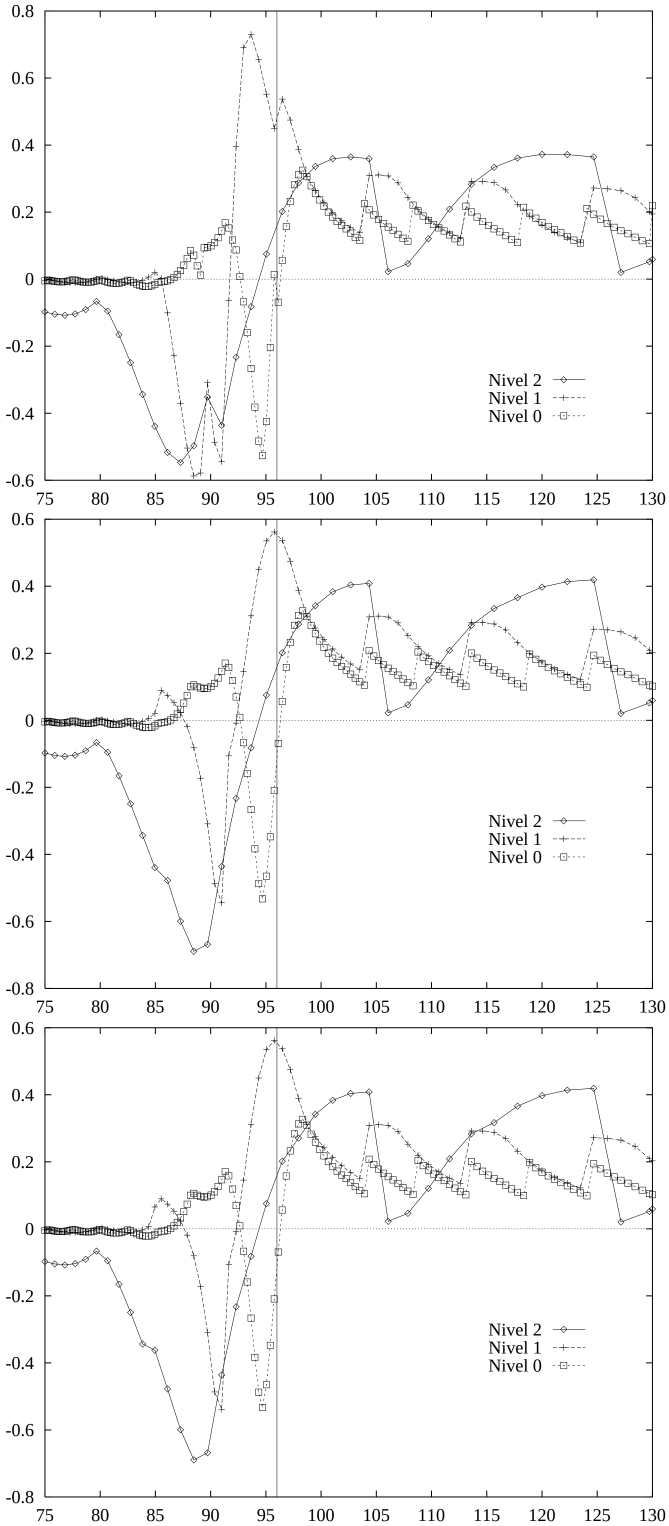
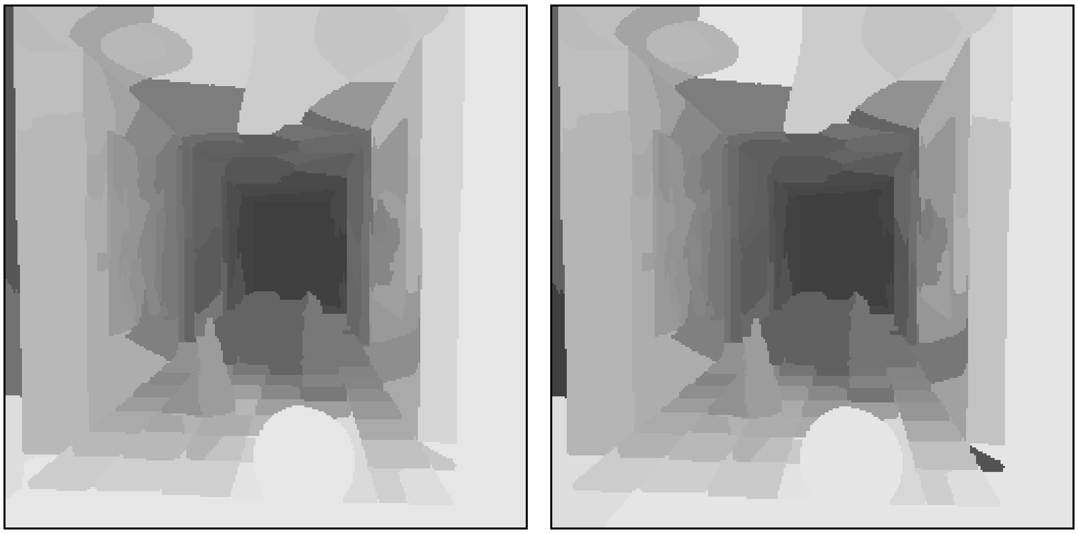



**Figura 5.8:** Evolución de $F_z, \forall Z$, de la región $R_2$ del ejemplo *pirámide*, con detección de oclusiones.

por pixel. Cabe destacar la mejora de la profundidad de la región alrededor del vértice del cono, gracias a la detección de la oclusión que ejerce el cono sobre ella.



(a) Sin detección de oclusiones (b) Con detección de oclusiones

**Figura 5.9:** Mapas de disparidades del ejemplo *corridor*, con y sin detección de oclusiones.

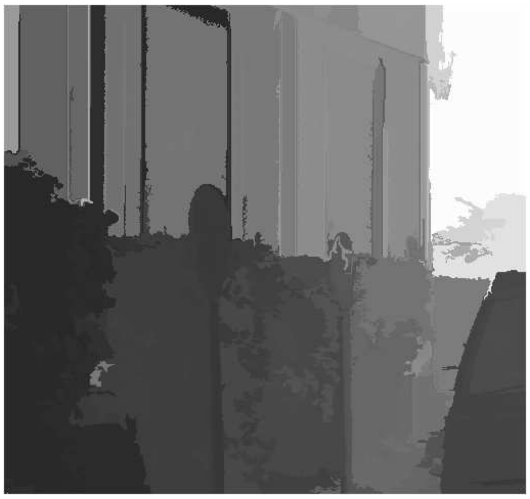
Aunque el número total de iteraciones es elevado (9, 9, 10, 21, 269, 69, 63, 34) la mayor parte de las iteraciones se realizan en los niveles intermedios (nivel 3), de modo que en los niveles de mayor resolución (nivel 0), que son los más costosos, se realiza un número de iteraciones menor. El rango de disparidades obtenido es $[1.28, 7.11111]$ y el tiempo de ejecución es de 1.5234309, según las unidades empleadas en la tabla 5.3.

### 5.4.2 Resultados con imágenes reales

La figura 5.10 muestra los resultados del par de imágenes *parking meter*, del conjunto de pruebas JISCT [Bolles *et al.*, 1993], obtenidos con $\Delta d = 0.5$ y $\Delta d_\epsilon = 0.001$. Las disparidades resultantes varían en el rango $[2.56, 14.2222]$. En este caso, sólo se han considerado 5 niveles, desde $32 \times 30$ hasta $512 \times 480$ pixeles. El número de iteraciones de cada nivel con estos parámetros es: 28, 31, 39, 52, 76.

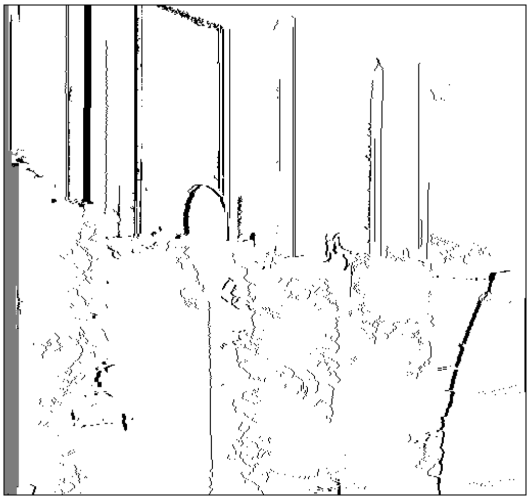
Los mapas de profundidades y disparidades se pueden comparar con otros obtenidos en los capítulos anteriores. Comparado con el capítulo 3 tiene una ventaja importante: las disparidades obtenidas tienen precisión subpixel. Esto da lugar a que se pueda observar claramente la diferencia entre la profundidad del parquímetro y del seto a la izquierda del mismo.

(a) Imagen de referencia



(b) Mapa de profundidades

(c) Mapa de disparidades



(d) Mapa de oclusiones

**Figura 5.10:** Mapas de profundidades, disparidades y oclusiones del *parking meter*.

Comparado con el capítulo 4 podemos observar que, aunque sigue habiendo algunos errores en ciertas regiones de la pared, en general, el resultado obtenido teniendo en cuenta la detección de oclusiones es mejor. Por ejemplo, el degradado de la profundidad entre regiones vecinas es más suave, los errores de la pared, aunque los hay, son menos numerosos, etc.

En cuanto al mapa de oclusiones, se puede observar que aparecen algunas oclusiones falsas en el seto y la pared, debido a la suposición de planos fronto-paralelos que se ha aplicado. Sin embargo, las oclusiones importantes se han obtenido bien, como son las del parquímetro y el coche.

La figura 5.11 muestra algunos resultados con imágenes del *Calibrated Imaging Laboratory* (CIL) de la Universidad Carnegie Mellon (CMU). Los parámetros empleados para obtener estos resultados se detallan en la tabla 5.4. Los resultados de profundidad, disparidad y oclusiones son similares en los dos casos que se detallan en la tabla, que solo se diferencian en el rango de profundidades que se proporciona al algoritmo.

Para ilustrar el comportamiento convergente del algoritmo en el bucle más exterior podemos tomar como ejemplo el par de imágenes *cil0001*, donde el nivel 0 ejecuta 158 iteraciones del bucle más interior, repartidas en 7 iteraciones del bucle más exterior como sigue: 89, 28, 12, 15, 5, 6, 3. Dado que $\Delta d$ se ha fijado inicialmente a un valor de 0.1, 89 iteraciones es un incremento de casi 9 pixeles, las 28 siguientes consisten en un incremento de $\frac{2.8}{2}$ pixeles, las 12 siguientes representan $\frac{1.2}{4}$ pixeles, etc.

| Imágenes CIL ($512 \times 384$) | | | | | | |
|---|---|---|---|---|---|---|
| Imagen | $Z_{min}$ | $\Delta d$ | $\Delta d_\epsilon$ | sec./pixel | # iteraciones | Tiempo rel. |
| cil0001 | 140 | 0.1 | 0.001 | $3 \times 1$ | 7, 7, 14, 20, 53, 62, 158 | 12.7 |
| cil0001 | 160 | 0.1 | 0.001 | $3 \times 1$ | 7, 7, 14, 20, 42, 52, 133 | 10.6 |
| cil0002 | 140 | 0.1 | 0.001 | $3 \times 1$ | 14, 13, 15, 18, 31, 147, 141 | 12.6 |
| cil0002 | 180 | 0.1 | 0.001 | $3 \times 1$ | 8, 8, 15, 16, 20, 43, 51 | 4.98 |

**Tabla 5.4:** Parámetros y resultados de correspondencia de los pares de imágenes CIL.

Si comparamos los resultados con los obtenidos en el capítulo anterior podemos observar que son muy similares. Sin embargo, se pueden apreciar algunas pequeñas mejoras debidas a la detección de oclusiones. Por ejemplo, en la torre de más a la izquierda del ejemplo *cil0001* parcialmente oculta por una casita, cabe notar una mejora en la profundidad de la región más próxima a dicha oclusión.

Por último, vamos a ver algunos resultados con imágenes adquiridas en el laboratorio de Visión por Ordenador con una cámara triestéreo Triclops. Este dispositivo dispone de tres cámaras dispuestas en L. A continuación se muestran los resultados de dos escenas que podrían constituir el entorno en el que trabaje un brazo robot o se desplace un robot móvil. La correspondencia se ha realizado sobre las imágenes rectificadas izquierda y derecha y utilizando la segmentación por unión de regiones.

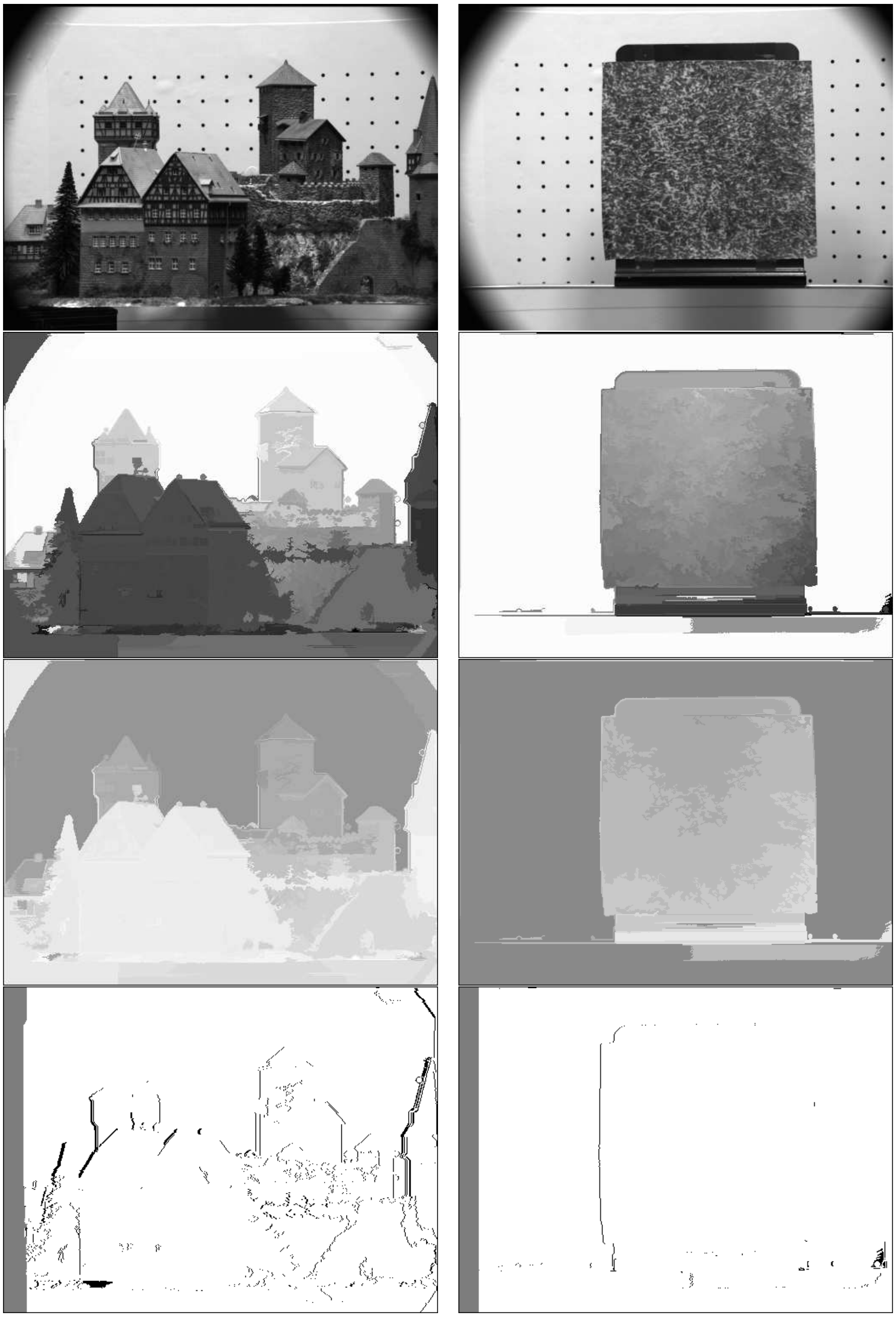


**Figura 5.11:** De izquierda a derecha: *cil0001* y *cil0002* del conjunto de imágenes CIL. De arriba a abajo: imagen de referencia, mapa de profundidades, mapa de disparidades y mapa de oclusiones.

En primer lugar, la figura 5.12 muestra los resultados para el par estéreo *despacho* donde se muestran el mapa de profundidades, el mapa de disparidades y el mapa de oclusiones obtenidos por el proceso de correspondencia. En este ejemplo hay pocas áreas en la izquierda que estén ocultas en la derecha. Entre las oclusiones resultantes, hay algunas correctas (la cajonera oculta un poco al armario) y otras falsas debidas a errores en la profundidad, como es la oclusión que aparece por encima de la cajonera. Esta oclusión es debida a la región que corresponde a un eje vertical que aparece con una profundidad más cercana de la esperada. Cabe resaltar que esto es debido a que la región que incluye dicho eje vertical, continúa por los bordes de los objetos situados encima de la cajonera y por el borde del armario, de modo que contiene pixeles de áreas de distinta profundidad.

Una posible solución a este problema podría ser la aplicación de la restricción de consistencia izquierda-derecha a este método, que consistiria en realizar no una sino dos veces la correspondencia, tomando cada vez una imagen del par estéreo como referencia. Esto permitiría integrar los resultados de ambos procesos, y modificar la segmentación inicial con el objetivo de evitar este tipo de errores. Esto significaría integrar un refinamiento de la segmentación en el proceso de correspondencia.

En la figura 5.13 se muestran varias vistas de la reconstrucción tridimensional de la escena *despacho*, donde se puede observar que los resultados son satisfactorios ya que proporcionan una representación poliédrica de la escena. Debido a la restricción impuesta, la representación del suelo de la escena es deficiente, ya que se muestra como un plano vertical. Aunque se podría estudiar el tratar el suelo, techo y paredes de una escena de manera especial, sería más interesante realizar la generalización del método a cualquier tipo de planos, de modo que el proceso de correspondencia por sí solo sea capaz de reconocer la geometría de estos elementos.

En la figura 5.14 se muestran los resultados para el par estéreo *libros*, donde se puede observar que en la imagen izquierda hay zonas de tamaño considerable que están ocultas en la imagen derecha. Tanto el mapa de profundidades como el de disparidades muestran buenos resultados debido a la existencia de menos errores de segmentación que en el ejemplo anterior. El mapa de oclusiones muestra todas las oclusiones importantes. La profundidad de la mesa, sin embargo, es un problema similar al del suelo en el ejemplo anterior. En la reconstrucción tridimensional de la escena (figura 5.15) se puede observar que se recupera satisfactoriamente la estructura tridimensional de todos los objetos de la escena, salvo la mesa, que aparece como un plano vertical por detrás de los libros.

## 5.5 Discusión y conclusiones

En este capítulo se ha mostrado un algoritmo cooperativo que encuentra las correspondencias y las oclusiones al mismo tiempo. Los experimentos con escenas sintéticas y reales muestran la mejora de los resultados gracias a la detección de las oclusiones. Los experimentos con escenas reales muestran que el método obtiene resultados razonables.

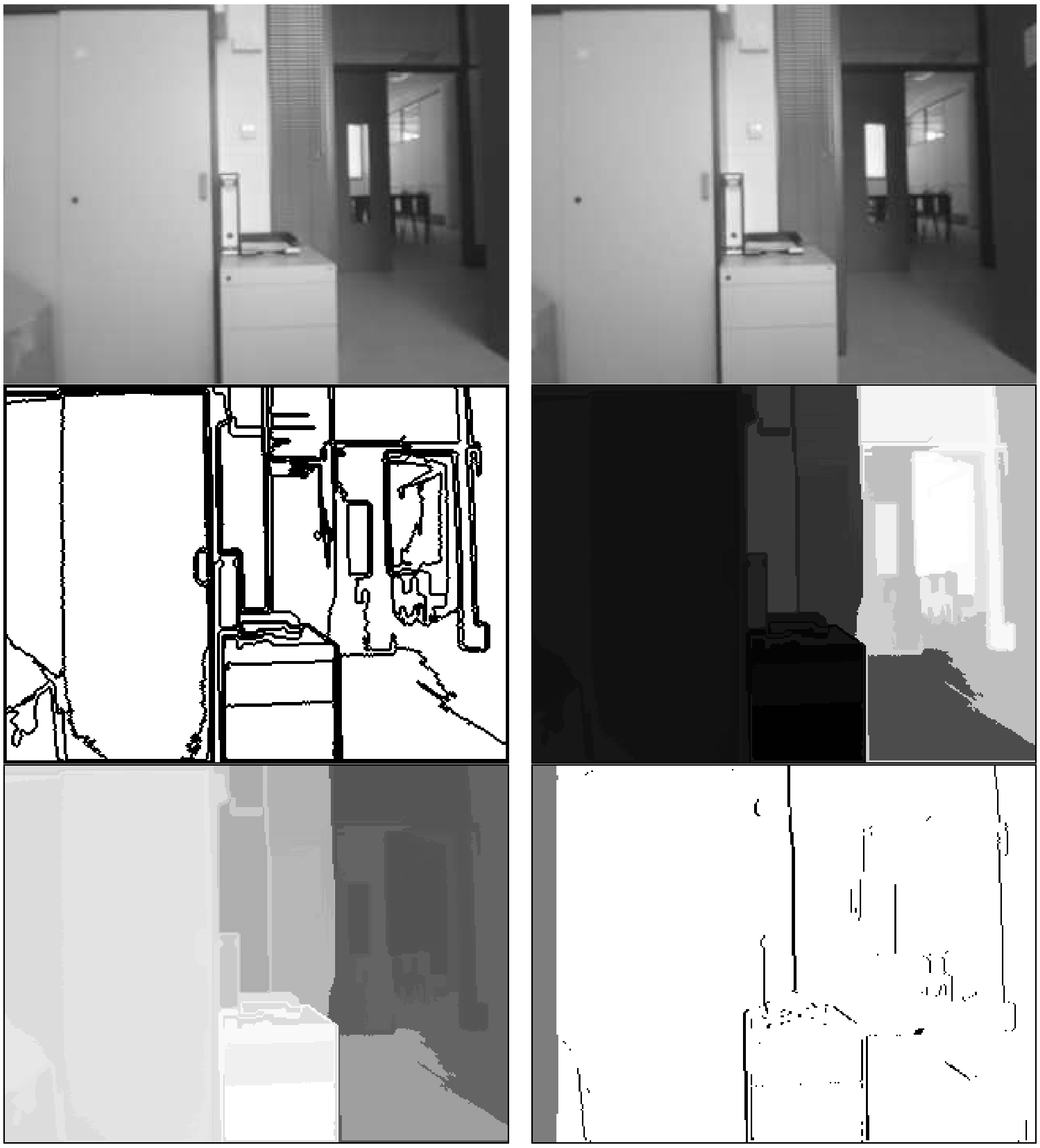


**Figura 5.12:** Resultados del par estéreo *despacho*. De izquierda a derecha y de arriba a abajo: imágenes izquierda y derecha, segmentación por unión de regiones, y mapas de profundidades, disparidades y oclusiones.

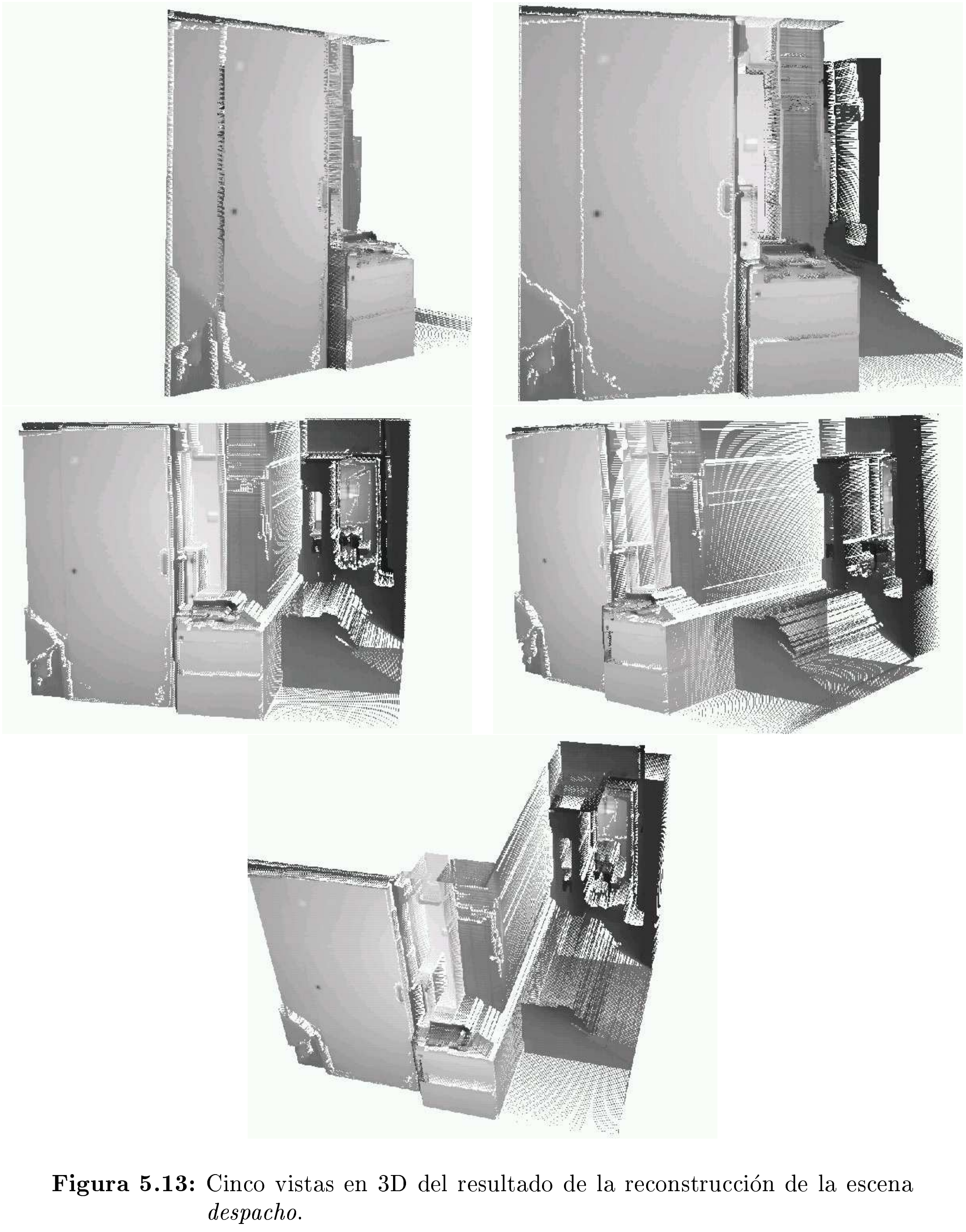


**Figura 5.13:** Cinco vistas en 3D del resultado de la reconstrucción de la escena *despacho*.

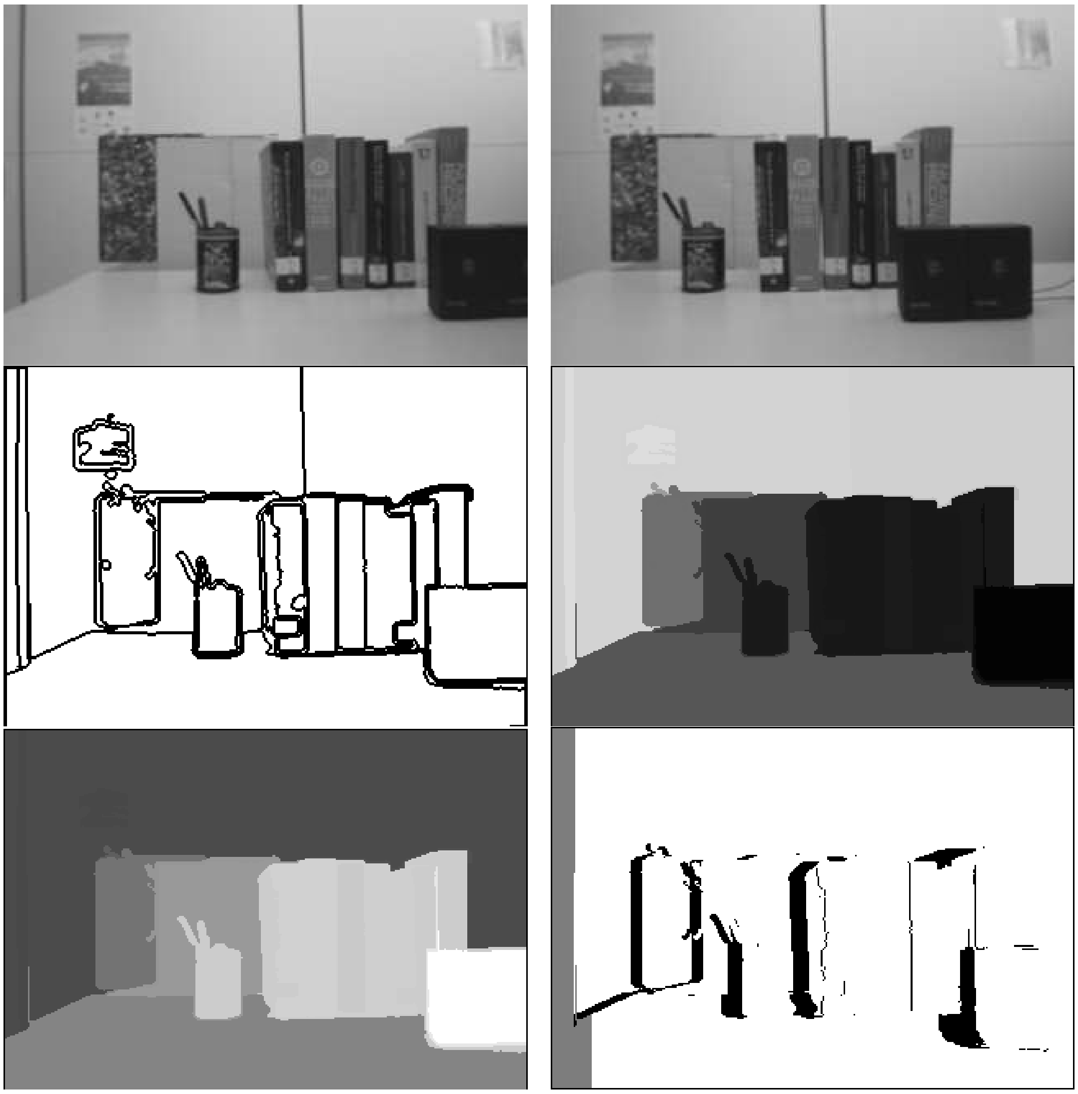


**Figura 5.14:** Resultados del par estéreo *libros*. De izquierda a derecha y de arriba a abajo: imágenes izquierda y derecha, segmentación por unión de regiones, y mapas de profundidades, disparidades y oclusiones.

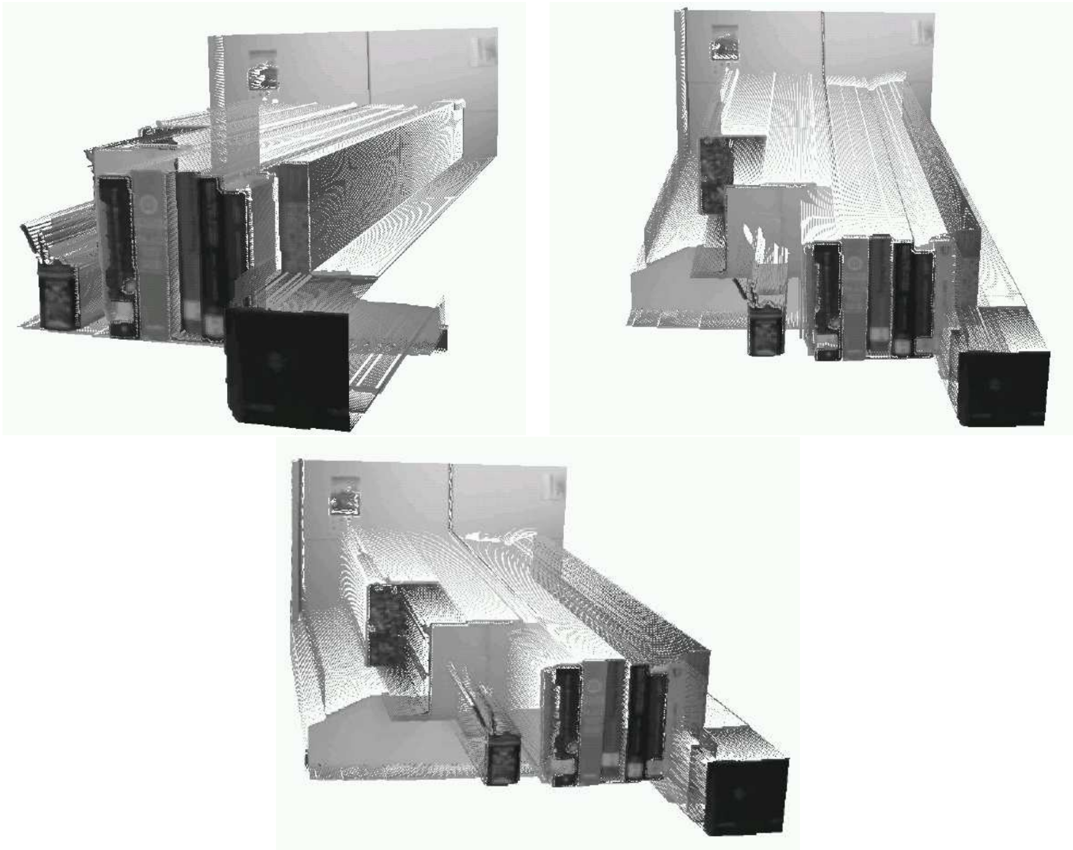


**Figura 5.15:** Tres vistas en 3D del resultado de la reconstrucción de la escena *libros*.

El método es aplicable a escenas donde la mayor parte de las superficies son planas, como son las escenas de interiores, y en las que la falta de textura y la presencia de discontinuidades en las profundidades y oclusiones son dificultades importantes. Otro tipo de escenas que encajan dentro de esta categoría son las escenas urbanas.

Las oclusiones forman parte de la información que se mantiene sobre la escena y ayudan a encontrar las correspondencias. La detección de oclusiones en el algoritmo cooperativo hace que la correspondencia sea más exacta.

Un posible trabajo futuro podría consistir en mejorar la detección de oclusiones, por ejemplo, mediante la definición de un porcentaje de oclusión para cada pixel, de modo que se pueda considerar este porcentaje en los cálculos estadísticos.

También se puede mejorar la eficiencia en el procesamiento de la tabla de pixeles visitados, realizando en cada iteración del algoritmo solamente las modificaciones necesarias debidas a aquellas profundidades de las regiones que hayan cambiado respecto a la iteración anterior.

# Capítulo 6

# Conclusiones y líneas de trabajo futuro

El objetivo principal de esta tesis ha sido contribuir al estudio y al desarrollo de técnicas que permitan conocer la estructura tridimensional de una escena observada por un sistema de cámaras binocular. Principalmente se ha estudiado el problema de la correspondencia, y con el objetivo de desarrollar una técnica de reconstrucción que sea aplicable a escenas poco texturadas que son las que generalmente corresponden al entorno de trabajo de un robot, se ha estudiado la correspondencia basada en regiones. A continuación enumeramos las contribuciones que aporta esta tesis doctoral a este problema, así como los proyectos futuros de investigación a los que ha dado lugar el trabajo realizado.

## 6.1 Conclusiones

En esta tesis se han abordado dos tipos de técnicas: la correspondencia basada en grafos, que ha sido ampliamente estudiada en la literatura, y la correspondencia basada en la minimización de una función de energía.

En primer lugar, se han estudiado las técnicas basada en grafos, donde se representan todas las posibles correspondencias y las compatibilidades entre ellas. Se han realizado aportaciones para la resolución de los problemas generados por las diferencias en la segmentación de las dos imágenes del par estereoscópico. Este trabajo está recopilado en el capítulo 3.

En segundo lugar, se ha desarrollado una técnica nueva, que obtiene directamente el mapa de profundidades de la escena. Esta técnica, basada en la minimización de una función de energía, ha sido evaluada bajo diferentes hipótesis. Se ha hecho especial incapié en la detección de oclusiones y discontinuidades en la profundidad, y se ha demostrado que tener en cuenta las oclusiones durante el proceso de correspondencia mejora los resultados.

Aunque este método se ha desarrollado y validado solamente bajo la restricción de una escena formada por planos paralelos al plano imagen, los resultados obtenidos nos animan a continuar el trabajo comenzado en esta tesis, generalizando a otras restricciones más laxas.

### 6.1.1 Aportaciones

A continuación se resumen las principales aportaciones del trabajo presentado en esta tesis. Parte de estas aportaciones han sido ya publicadas en diferentes foros (ver lista de publicaciones en la sección 6.1.2).

**Manejo de los errores en la segmentación.** En el capítulo 3 se propone un método para la resolución de los problemas generados por las diferencias en la segmentación de un par estéreo de imágenes. Se dan soluciones a los problemas no resueltos por la correspondencia basada en regiones clásica:

- *Preproceso de unión de regiones:* para la resolución de la sobresegmentación e infrasegmentación. El método que se propone está basado en grafos, y es de bajo coste adicional al método clásico. Se consideran pares de regiones susceptibles de emparejarse con la misma región en la otra imagen y si la unión de ambas se empareja mejor con ella que las regiones por separado, se establece un arco entre los dos nodos que representan a esos emparejamientos en el grafo de asociación de incompatibilidades. El conjunto de todos los cliques máximos de dicho grafo constituye el conjunto de uniones de regiones a realizar para mejorar la correspondencia entre las dos imágenes. El método consigue resolver satisfactoriamente la mayoría de los errores de segmentación.
- *Criterio de similitud basado en correlación:* El área de intersección entre dos regiones a distintas disparidades es variable, y por tanto, se precisa de dos criterios de similitud diferentes para cada fase de la correspondencia: uno para la creación de los nodos del grafo de asociación y otro para la asignación de los pesos de dichos nodos. La ventaja respecto a los criterios basados en las propiedades geométricas de las regiones reside en que la correlación permite manejar los casos en los que la segmentación proporciona regiones de forma diferente en cada imagen, y por lo tanto permite mejorar los resultados de la correspondencia. Basar el criterio de similitud entre regiones en técnicas de correlación supone una combinación de las técnicas de correspondencia basadas en áreas y basadas en características.

**Algoritmo alternativo a la búsqueda del mejor clique máximo.** La búsqueda del mejor clique máximo proporciona la mejor correspondencia en los métodos basados en grafos, pero es un problema NP-completo. Basándonos en que los nodos del grafo de asociación incluyen un peso que indica la similitud entre las dos regiones del nodo, y que este peso se modifica mediante un proceso de relajación que incorpora información estructural del grafo, proponemos una alternativa a este algoritmo que no garantiza la solución óptima, pero obtiene una solución muy cercana a ella con un coste cuadrático, lo que permite su aplicación práctica.

**Correspondencia basada en minimización** del error de correlación entre regiones: Se ha presentado un nuevo método de correspondencia que pretende encontrar la región correspondiente que más se parece a la región de referencia, sin tener que segmentar la otra imagen. Se basa en la minimización de una función de energía que consiste en el error de correlación de la intensidad de los puntos de cada región con la intensidad de los puntos de su región correspondiente, calculada a partir de la profundidad estimada hasta el momento.

Este es un problema *mal planteado* (*ill-posed*) que requiere de alguna restricción para poder abordar su resolución. El método se ha desarrollado bajo la suposición de superficies planas paralelas al plano imagen. Aunque esta restricción puede parecer de poca aplicabilidad, los experimentos realizados para este caso sencillo muestran su efectividad y han permitido hacer un estudio y una evaluación exhaustivos de aspectos numéricos y algorítmicos del método que permitan una extensión sencilla a casos más generales.

Correspondencia y reconstrucción se realizan en un sólo proceso, obteniendo una representación poliédrica de la escena, que a su vez se puede traducir en un mapa de profundidades denso. Por tanto, no es necesaria la segmentación de la información de rango, puesto que ya tenemos una representación tridimensional de la escena utilizable.

Este proceso permite la aparición de discontinuidades de la profundidad en las fronteras entre regiones, es decir, los lugares donde hay cambios bruscos de intensidad, que es precisamente donde pueden haber discontinuidades en la profundidad. La profundidad de las zonas homogéneas de la imagen, que están representadas en el interior de las regiones, está restringida según la suposición que se haga sobre la forma de las superficies de la escena. El método halla la correspondencia satisfactoriamente, con precisión subpixel, y con un coste teórico en el peor caso similar al de las técnicas basadas en áreas.

**Detección de oclusiones.** El método de minimización propuesto permite integrar la detección de las oclusiones en el proceso de correspondencia, de modo que las profundidades y las oclusiones se obtienen mediante un proceso cooperativo que permite mejorar la precisión de las correspondencias obtenidas. Como resultado adicional, se obtiene un mapa de oclusiones junto con el mapa de profundidades.

Esta mejora debida a la consideración de las oclusiones durante la correspondencia es más importante si cabe de cara a la generalización del método a cualquier tipo de superficies planas o de orden superior.

### 6.1.2 Publicaciones

Las publicaciones originadas hasta el momento por el trabajo que se ha presentado en esta tesis doctoral son las siguientes:

"Solving Oversegmentation Errors in Graph-Based Region Matching"
Angeles López y Filiberto Pla.
*Proc. of the 8th. Portuguese Conference on Pattern Recognition*, 387-394.
Guimaraes (Portugal), Marzo 1996.

"Dealing with Segmentation Errors in Region-Based Stereo Vision"
Angeles López y Filiberto Pla.
*Proc. of the 7th. Conferencia de la Asociación Española Para la Inteligencia Artificial*, 317–326. Torremolinos (Malaga), Noviembre 1997.

"Robust Region-Based Stereo Vision to Build Environment Maps for Robotics Applications"
Angeles López y Filiberto Pla.
Engineering Applications of Artificial Intelligence and Expert Systems, 365-374. Benicssim, Spain, Julio 1998.
*Lecture Notes in Artificial Intelligence* 1416. Tasks and Methods in Applied Artificial Intelligence.
Ed. A.P. del Pobil *et al.* Springer-Verlag, Vol. II. ISBN 3-540-64574-8.

"Dealing with Segmentation Errors in Region-based Stereo Matching"
Angeles López y Filiberto Pla.
*Pattern Recognition*,
aceptado en Mayo 1999.

"Region-based Stereo Vision using a Minimization Approach"
Angeles López y Filiberto Pla.
*VIII Simposio Nacional de Reconocimiento de Formas y Análisis de Imagen*, 11-18. Bilbao, Mayo 1999.
ISBN 84-95120-81-X.

"A Minimization Approach for 3D Recovery in Region-based Stereo Vision"
Angeles López y Filiberto Pla.
*Proc. of the IEE Conf. on Image Processing and its Applications*,
Manchester, Julio 1999.
aceptado en Marzo 1999.

## 6.2 Líneas de trabajo futuro

Hay varias extensiones posibles de este trabajo, sobre todo en relación con la técnica presentada en el capítulo 4, aunque algunas de ellas son aplicables también a la correspondencia de regiones basada en grafos.

- La línea de trabajo futuro más importante que parte del trabajo mostrado en esta tesis doctoral es la **generalización del método de minimización** a superficies planas de cualquier orientación y a superficies de otros tipos como son las superficies cuadráticas.

  La generalización a superficies planas cualesquiera nos parece que es crucial para una aplicabilidad completa del método y por tanto, constituye el objetivo inmediato. El interés de esta generalización reside sobre todo en su utilidad práctica en multitud de aplicaciones en escenarios de interiores y urbanos.

- Otra posibilidad de continuación de este trabajo está relacionada con la **segmentación** de la imagen de referencia. Hay varios aspectos que quedan pendientes de ser estudiados:
  - Incluir la segmentación de la imagen de referencia en el proceso de correspondencia, o un refinamiento de la segmentación inicial. Por ejemplo, se podría adaptar el principio de la consistencia izquierda-derecha a esta técnica, realizando dos cálculos de la correspondencia con diferentes imágenes de referencia, de modo que sirva para refinar la segmentación.
  - Utilizar no una, sino varias segmentaciones de la imagen de referencia en cada nivel del esquema multiescala, quizá un sistema jerárquico de regiones, de modo que buscando la correspondencia de las regiones desde las que pertenecen al nivel más alto en la jerarquía, se va refinando hasta las regiones más pequeñas.

  Ambas posibilidades se pueden considerar conjuntamente, de modo que se mejore lo máximo posible la representación tridimensional final de la escena.

- En cuanto a la **detección de oclusiones** durante la correspondencia, se puede profundizar en el estudio del papel de los pixeles ocultos en la validación de las correspondencias obtenidas.
  - Incluir un *porcentaje de oclusión* que refleje para cada pixel la probabilidad de que esté oculto, de modo que un pixel semi-oculto no sea eliminado completamente de los cálculos relativos a su región, sino que sea considerado en la porción no oculta.
  - Mejorar la eficiencia en el procesamiento de la tabla de pixeles visitados, realizando, en cada iteración del algoritmo, solamente las modificaciones necesarias debidas a aquellas profundidades de las regiones que hayan cambiado respecto a la iteración anterior.
  - Valorar la bondad de la oclusión de cada pixel marcado como oculto, esto es, el error de similitud del pixel con su hipotético correspondiente (si no hubiera una región ocultándole). Esta información se podría utilizar al nivel de la región para valorar la bondad del área oculta de la región.

Este último punto puede ser interesante sobre todo de cara a las distintas generalizaciones que se puedan realizar en el futuro, ya que ante la posibilidad de dirigir la búsqueda en varias direcciones diferentes puede ayudar a decidir empezar por la dirección que, por ejemplo, deshaga la oclusión menos "buena" o favorezca la oclusión de mayor índice de bondad.

- Otra línea de futura investigación podría consistir en explorar **alternativas a la correlación** como medida de similitud entre regiones, como pueden ser las técnicas basadas en *información mutua* [Viola, 1997], con el objetivo de obtener información de las regiones, incluso cuando la variación de la intensidad es pequeña.

- Por último, cabe nombrar la extensión del método de minimización a **sistemas trinoculares** o de múltiples imágenes.

# Apéndice A

# Geometría de un sistema de visión estereoscópico

Para una mejor comprensión de este apéndice es necesario tener algunas nociones básicas de geometría proyectiva. Puede encontrarse una buena introducción a la geometría proyectiva aplicada a la visión por computador en el libro de Faugeras [1993] y el de Mundy y Zisserman [1992].

## A.1 La proyección en perspectiva

En este apéndice se describen las nociones geométricas fundamentales de la visión estereoscópica, basándonos en el modelo de cámara *pinhole*, que es el más utilizado en los sistemas de visión por computador (figura A.1).

### A.1.1 La matriz de proyección

La calibración consiste en obtener la relación entre un punto del mundo, $M$, y su proyección normalizada en el plano imagen, $m$. Podemos dividir esta relación en dos pasos principales.

El primer paso es un simple cambio del sistema de coordenadas: Ponemos $M$ en coordenadas de la cámara mediante una rotación $R$ y una traslación $t$.

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R \begin{bmatrix} x \\ y \\ z \end{bmatrix} + t \tag{A.1}$$

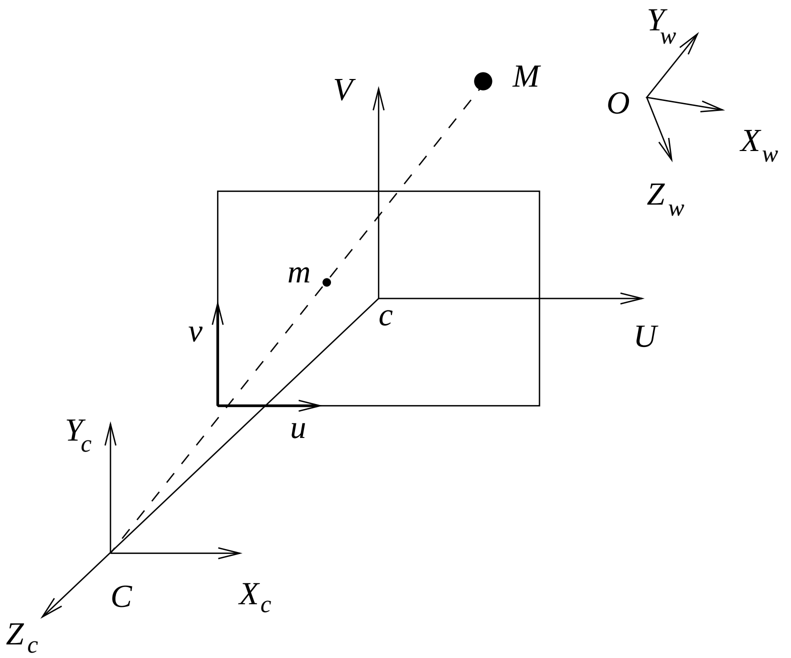



**Figura A.1:** El modelo de proyección en perspectiva *pinhole*

Esta ecuación puede reescribirse en coordenadas homogéneas como

$$\tilde{M}_c = \begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix} \tilde{M} \tag{A.2}$$

donde $0_3$ es el vector cero de tamaño 3. La rotación $R$ y la traslación $t$ son los *parámetros extrínsecos* de la calibración. Hay 6 parámetros, 3 ángulos de rotación para $R$ y 3 valores de traslación para $t$.

El segundo paso consiste en calcular $m$ en el sistema de coordenadas del plano imagen, y cambiar las coordenadas de $m$ al sistema de coordenadas normalizado. Las coordenadas en el plano imagen de $m$ son fáciles de calcular,

$$u_s = \frac{x_c f}{z_c} \quad \text{y} \quad v_s = \frac{y_c f}{z_c}$$

donde $f$ es la *longitud focal*, esto es, la distancia entre el centro óptico y el plano imagen. Este cálculo no es lineal, pero se puede hacer que lo sea utilizando álgebra proyectiva:

$$s \begin{bmatrix} u_s \\ v_s \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \tag{A.3}$$

donde $s$ es un factor de escala.

Para cambiar $m$ de coordenadas del plano imagen a coordenadas normalizadas, tenemos que considerar las coordenadas normalizadas del origen de la imagen $c$, $(u_0, v_0)$, y las unidades de los ejes $u$ y $v$, $(k_u, k_v)$, que pueden ser diferentes debido a la electrónica de la adquisición. Esto es, deben añadirse una traslación y un escalado a la ecuación anterior, obteniendo:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} fk_u & 0 & u_0 & 0 \\ 0 & fk_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \tag{A.4}$$

Por tanto, hay cuatro parámetros que calibrar, $\alpha_u = fk_u$, $\alpha_v = fk_v$, $u_0$ y $v_0$. Estos son los *parámetros intrínsecos* de la calibración.

Finalmente, consideramos tanto los parámetros intrínsecos como los extrínsecos para obtener la relación global,

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{A.5}$$

esto es,

$$s\tilde{m} = P\tilde{M} \tag{A.6}$$

donde $P$ es una matriz 3×4 llamada la *matriz de proyección*:

$$P = \begin{bmatrix} A & 0_3 \end{bmatrix} \begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix} \tag{A.7}$$

La matriz $P$ tiene 10 grados de libertad: 4 parámetros intrínsecos y 6 parámetros extrínsecos, que pueden representarse como dos matrices separadas o como la matriz de proyección completa.

Dado que la matriz $A$ tiene características especiales, la matriz $P$ puede descomponerse fácilmente en sus parámetros intrínsecos y extrínsecos. En la sección A.1.2 se muestra el cálculo de esta descomposición.

Finalmente, cuando es posible que haya una desviación del eje óptico, puede introducirse un nuevo parámetro, $\theta$, que represente el ángulo entre los ejes $U$ y $V$.

$$A = \begin{bmatrix} \alpha_u & -\alpha_u \cot\theta & u_0 \\ 0 & \alpha_v / \sin\theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{A.8}$$

En este caso, la matriz de proyección tiene 11 grados de libertad.

### A.1.2 Parámetros de calibración

La *matriz de proyección* relaciona cualquier punto del mundo 3D con su punto correspondiente en el plano imagen, esto es, su proyección en el plano retínico de la cámara. Dado que relaciona un punto en coordenadas 3D (coordenadas del mundo) con su punto proyectado en coordenadas 2D (coordenadas normalizadas), es una matriz 3×4.

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}$$

Contiene todos los parámetros involucrados en la proyección:

- *Parámetros intrínsecos*: características internas de la cámara.
- *Parámetros extrínsecos*: posición y orientación de la cámara.

La matriz de proyección puede reescribirse como dos matrices separadas, una para cada tipo de parámetros, como se muestra en la ecuación A.5. Es decir,

$$P = \begin{bmatrix} A & 0_3 \end{bmatrix} D \tag{A.9}$$

donde $A$ es una matriz 3×3 que contiene los parámetros intrínsecos (ecuación A.8) y $D$ es una matriz 4×4 que define la rotación y la traslación de la cámara respecto a las coordenadas del mundo:

$$D = \begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Las técnicas de calibración conocidas proporcionan los parámetros separadamente, o la matriz de proyección como un conjunto. Podemos necesitar descomponer la matriz de proyección en parámetros intrínsecos y extrínsecos para algunos usos.

#### Cálculo de los parámetros intrínsecos

En el capítulo 3 de [Faugeras, 1993], se muestran las restricciones sobre la matriz $P$ y se demuestra que existen 4 conjuntos de parámetros intrínsecos y extrínsecos tales que $P$ puede escribirse como A.6. Suponiendo que el origen de coordenadas está delante de la cámara ($t_z > 0$), sólo hay dos soluciones que corresponden a invertir los ejes $u$ y $v$.

El parámetro $\theta$ de la ecuación A.8 modela una posible imprecisión en la orientación de los ejes. Si asumimos que la orientación de los ejes es perfecta[1], los parámetros intrínsecos se reducen a 4 y su matriz es más simple.

$$A = \begin{bmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{A.10}$$

Los parámetros intrínsecos pueden calcularse directamente a partir de los elementos de la matriz de proyección,

$$u_0 = q_1^T q_3 \tag{A.11}$$
$$v_0 = q_2^T q_3 \tag{A.12}$$
$$\alpha_u = \varepsilon\sqrt{q_1^T q_1 - u_0^2} \tag{A.13}$$
$$\alpha_v = \varepsilon\sqrt{q_2^T q_2 - v_0^2} \tag{A.14}$$

donde $\varepsilon$ representa el signo (por tanto, hay dos soluciones posibles para $\alpha_u$ y $\alpha_v$), y $q_1$, $q_2$ y $q_3$ son vectores extraidos de $P$ como sigue:

$$P = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} q_1 & p_{14} \\ q_2 & p_{24} \\ q_3 & p_{34} \end{bmatrix}$$

### Cálculo de los parámetros extrínsecos

Podemos reescribir la ecuación A.9 como

$$P = A \begin{bmatrix} R & t \end{bmatrix}$$

donde $\begin{bmatrix} R & t \end{bmatrix}$ es una matriz 3×4 que puede denotarse con tres vectores $r_1$, $r_2$ y $r_3$, como sigue

$$\begin{bmatrix} R & t \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}$$

Una vez que se conocen los parámetros intrínsecos, podemos calcular el resto de parámetros multiplicando la inversa de la matriz $A$ en ambos lados, esto es,

$$\begin{bmatrix} R & t \end{bmatrix} = A^{-1} P$$

[1]Es decir, que el eje de vista es ortogonal a los ejes de coordenadas del plano imagen, $\theta = \frac{\pi}{2}$.

La inversa de la matriz $A$ es

$$A^{-1} = \begin{bmatrix} \frac{1}{\alpha_u} & 0 & -\frac{u_0}{\alpha_u} \\ 0 & \frac{1}{\alpha_v} & -\frac{v_0}{\alpha_v} \\ 0 & 0 & 1 \end{bmatrix} \tag{A.15}$$

Por tanto, podemos calcular los elementos de $\begin{bmatrix} R & t \end{bmatrix}$ como

$$r_1 = (p_1 - u_0 p_3)/\alpha_u \tag{A.16}$$

$$r_2 = (p_2 - v_0 p_3)/\alpha_v \tag{A.17}$$

$$r_3 = p_3 \tag{A.18}$$

La condición para seleccionar el signo en las ecuaciones A.13 y A.14 (que influirá en las ecuaciones A.16 y A.17) es que $\det(R)$ debe ser 1, para que sea una matriz ortogonal que conserva la orientación.

## A.2 Geometría epipolar

En esta sección se estudia la geometría de la adquisición de un par estereoscópico de imágenes. En el libro [Xu y Zhang, 1996] se puede encontrar una revisión más detallada sobre la geometría epipolar.

Consideremos dos cámaras que toman vistas diferentes de la misma escena (figura A.2). ¿Cuál es la relación entre dos proyecciones, $m$ y $m'$, de un punto del mundo, $M$?

Nótese que todos los puntos de la línea recta entre $M$ y $C$ se proyectan sobre el mismo punto $m$. Todos estos puntos se proyectan en la segunda cámara en una línea llamada la *línea epipolar* de $m$. Esta línea es la intersección entre dos planos: el plano definido por $m$, $C$ y $C'$, llamado el *plano epipolar*, y el plano imagen $I'$.

De manera similar, todos los puntos de la línea entre $M$ y $C'$ se proyectan sobre el mismo punto $m'$ en la segunda cámara, y sobre la línea epipolar de $m'$ en la primera cámara. Por tanto, si $m$ y $m'$ corresponden al mismo punto de la escena, entonces $m$, $m'$, $C$ y $C'$ deben estar sobre el mismo plano. Esta restricción se le llama la *restricción de coplanaridad*.

La figura A.3 muestra los planos imagen y el plano epipolar de $m$ y $m'$. Todos los planos epipolares de una imagen forman un haz de planos que contiene la línea $CC'$.

Más aún, todas las líneas epipolares de la primera imagen pasan sobre el mismo punto, llamado el *epipolo*. Cada epipolo, $e$ y $e'$, es la intersección entre la línea $CC'$ y cada plano imagen, $I$ e $I'$, respectivamente. Todas las líneas epipolares de un plano imagen forman un haz de líneas que contienen el epipolo.

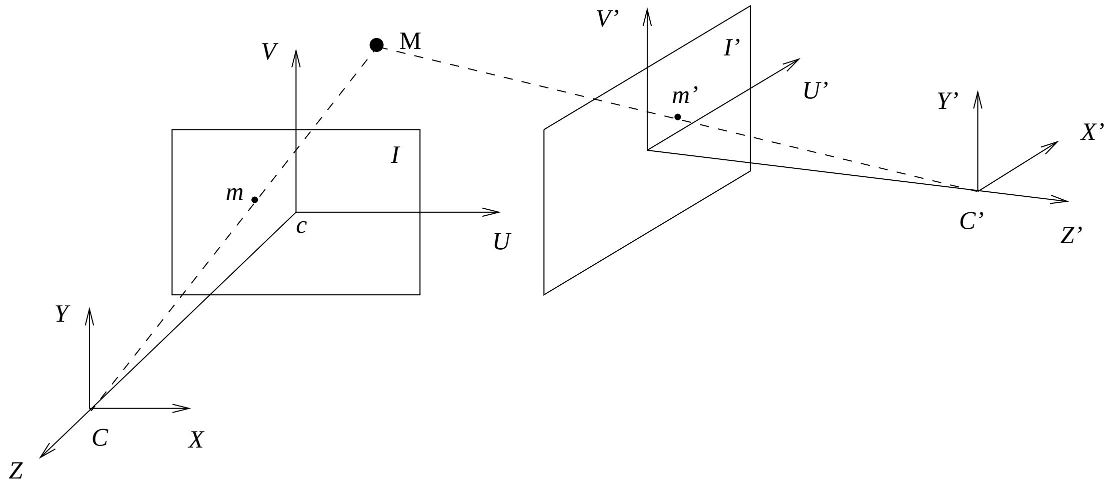



**Figura A.2:** Geometría de dos vistas de la escena.

Llamemos $l'_m$ a la línea epipolar de $m$ en la segunda imagen y $l_{m'}$ a la línea epipolar de $m'$ en la primera imagen. Para un punto $m$ de la primera imagen, su punto correspondiente debe estar en $l'_m$ (y simétricamente, el punto correspondiente de $m'$ debe estar en $l_{m'}$). Esta es una restricción muy importante en la búsqueda de las correspondencias en la visión estereoscópica, y se llama la *restricción epipolar*: el espacio de búsqueda del punto correspondiente se reduce de 2D (el plano imagen) a 1D (la línea epipolar).

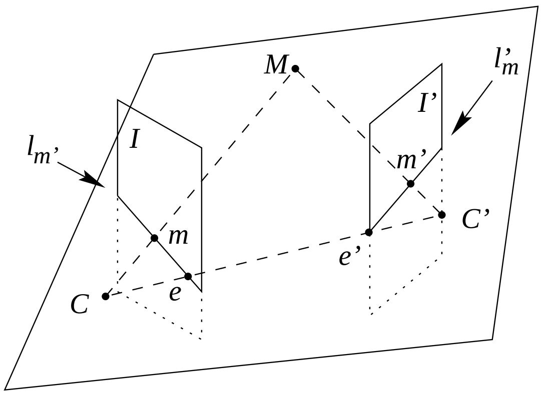



**Figura A.3:** Geometría epipolar.

La relación entre un punto y su línea epipolar es proyectiva lineal, y puede escribirse como

$$F\tilde{m} = l'_m \tag{A.19}$$

donde $F$ es una matriz 3×3 llamada la *matriz fundamental* [Luong y Faugeras, 1996]. Esta matriz contiene información sobre los parámetros intrínsecos y la posición relativa entre las cámaras.

Dado que $m'$ está en $l'_m$ por definición, se puede deducir que:

$$\tilde{m}'^T F \tilde{m} = 0 \tag{A.20}$$

y simétricamente,

$$\tilde{m}^T F^T \tilde{m}' = 0 \tag{A.21}$$

Si conocemos los parámetros de calibración de cada imagen, esto es, las matrices de proyección,

$$s\tilde{m} = P\tilde{M} \tag{A.22}$$

$$s'\tilde{m}' = P'\tilde{M} \tag{A.23}$$

entonces $F$ puede calcularse fácilmente a partir de $P$ y $P'$ como se muestra en la subsección A.2.2 (véase [Luong y Faugeras, 1996] para más detalles).

Por otro lado, $F$ puede calcularse directamente a partir de las dos vistas mediante una serie de métodos [Zhang, 1996b] sin una calibración previa. Este tipo de calibración se llama *calibración débil* [Robert y Faugeras, 1995, Robert *et al.*, 1995, Zhang *et al.*, 1995].

### A.2.1 Geometría epipolar paralela

Un caso particular de la geometría epipolar es el caso en el que los ejes ópticos de ambas cámaras son paralelos. En este caso se dice que el sistema estereoscópico tiene geometría paralela.

Con este tipo de geometría, todas las líneas epipolares son paralelas entre sí y los epipolos se encuentran en el infinito. Si además la posición relativa entre las cámaras sólo difiere en el eje $X_c$, las líneas epipolares son paralelas a las filas de la imagen (figura A.4) y por tanto los puntos correspondientes se encuentran en la misma fila.

Con esta configuración de las cámaras, la disparidad se calcula como la diferencia entre los valores de la coordenada horizontal de cada punto $m = (u, v)$ y su punto correspondiente, $m' = (u', v)$. Es decir, $d = u' - u$.

La disparidad de un par de puntos correspondientes en un sistema de cámaras paralelas es inversamente proporcional a la profundidad del punto 3D proyectado,

$$d = \frac{bf}{z_c} \tag{A.24}$$

donde $f$ es la longitud focal y $b$ es la distancia entre las cámaras (*baseline*).

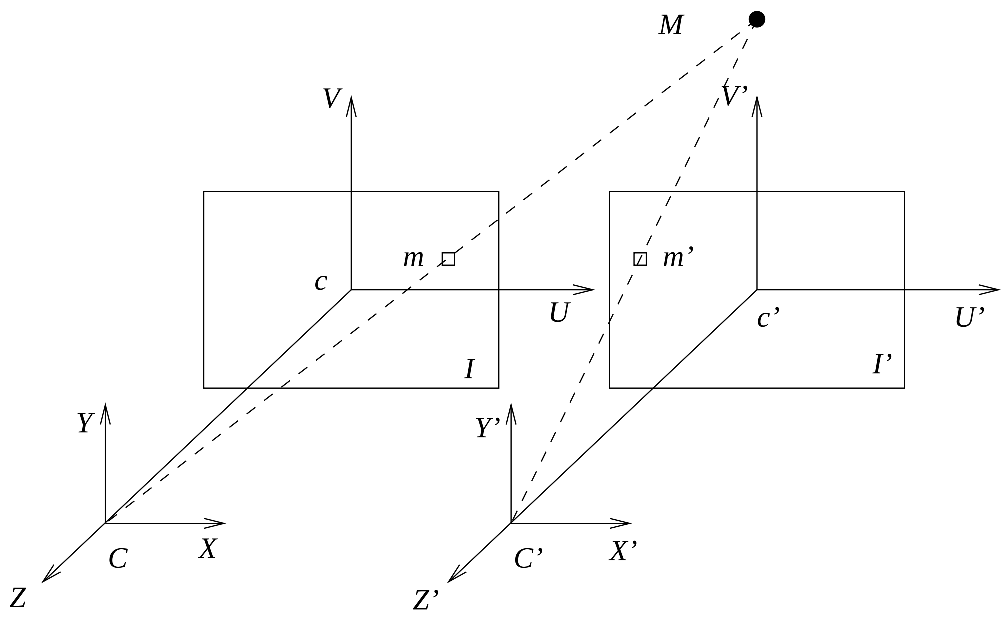



**Figura A.4:** Geometría epipolar paralela.

### A.2.2 De las matrices de proyección a la matriz fundamental

La matriz de proyección de cada imagen puede reescribirse como:

$$P = \left[ \begin{array}{cc} Q & q \end{array} \right]$$

donde $Q$ es una matriz 3×3 de rango 3 y $q$ es un vector 3×1. El centro óptico de cada cámara satisface la ecuación $P\tilde{C} = 0$ de lo que podemos concluir

$$C = -Q^{-1}q$$

Podemos asumir sin pérdida de generalidad que $C$ no está en el infinito,

$$\tilde{C} = \left[ \begin{array}{c} -Q^{-1}q \\ 1 \end{array} \right]$$

Dado un par estéreo, el epipolo de la segunda imagen es la proyección del centro óptico de la primera cámara en la segunda cámara, esto es:

$$e' = P'\tilde{C} = P' \left[ \begin{array}{c} -Q^{-1}q \\ 1 \end{array} \right] = \left[ \begin{array}{cc} Q' & q' \end{array} \right] \left[ \begin{array}{c} -Q^{-1}q \\ 1 \end{array} \right] = q' - Q'Q^{-1}q$$

La línea epipolar $l'_m$ de un punto $m$ en la primera imagen es la proyección del rayo óptico $CM$ en la segunda cámara. El rayo óptico puede definirse mediante el centro óptico $\tilde{C}$ (que se proyecta en el epipolo $e'$) y el punto en el infinito de $CM$ que es $[M\ 0]^T = [Q^{-1}m\ 0]^T$. Este punto se proyecta a:

$$P' \begin{bmatrix} Q^{-1}m \\ 0 \end{bmatrix} = Q'Q^{-1}m$$

La representación proyectiva de la línea epipolar $l'_m$ se obtiene tomando el producto vectorial de estos dos puntos,

$$l'_m = e' \times Q'Q^{-1}m = [e']_\times Q'Q^{-1}m = Fm$$

resultando

$$F = [e']_\times Q'Q^{-1} = [q' - Q'Q^{-1}q]_\times Q'Q^{-1}$$

## A.3 Rectificación de imágenes

La rectificación de un par de imágenes estéreo permite simplificar la búsqueda de la correspondencia a lo largo de la línea epipolar. Esta simplificación consiste en cambiar la geometría epipolar a geometría epipolar paralela, de modo que el recorrido de las líneas epipolares se reduce a recorrer las filas de la imagen. Los epipolos se envían al infinito, de modo que las líneas epipolares se hacen paralelas y horizontales.

Esta simplificación se consigue mediante una simple transformación geométrica.

- Si asumimos cámaras en perspectiva lineales, la transformación consiste en la proyección de ambas imágenes en el mismo plano, llamado *plano de rectificación* [Faugeras, 1993], que es un plano paralelo a la línea recta entre los dos centros ópticos de las cámaras, $CC'$;
- si las cámaras son en perspectiva pero no lineales, ha de tenerse en cuenta la distorsión de la cámara antes de realizar la proyección en el plano de rectificación;
- en cualquier otro caso, los métodos para rectificar el par de imágenes estéreo son completamente diferentes, dependiendo del modelo de cámara.

Asumimos que utilizamos cámaras en perspectiva lineal. Proyectar ambas imágenes al plano de rectificación hace que los epipolos sean proyectados al infinito. La matriz fundamental rectificada tiene la forma

$$F = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \tag{A.25}$$

Dado un par de imágenes estéreo, hay un conjunto de transformaciones posibles para obtener el par de imágenes rectificado. Este conjunto es una familia de *matrices de rectificación*, con 9 grados de libertad. El lector podrá encontrar un estudio de esta familia en [Devernay, 1996], así como algunos métodos para calcular el mejor par de matrices de rectificación.

## A.4 Correspondencia sin rectificación

Aunque la rectificación permite reducir las líneas epipolares a filas de la imagen, hay una serie de métodos que permiten establecer correspondencias sin necesidad de rectificar. En las técnicas que calculan las disparidades intermedias, basta con guardar la disparidad en los dos ejes del sistema de coordenadas normalizado. En las técnicas que calculan directamente la profundidad, como es el método desarrollado en el capítulo 4 de esta tesis, el objetivo es encontrar una función que, dada una proyección de un punto de la escena y dada su profundidad, determine el punto correspondiente en la otra imagen.

Sean $I_1$ e $I_2$ dos vistas diferentes de la misma imagen, siendo $I_1$ la imagen de referencia, y sea $Z$ el mapa de profundidades de todos los pixeles de $I_1$. El objetivo es calcular el punto correspondiente $m'$ en la imagen $I_2$ de un punto dado $m$ en la imagen $I_1$ dada su profundidad, $Z(m)$.

$$m' = f_{12}(m, Z(m)) \tag{A.26}$$

Sea $M = (x_m, y_m, z_m)$ el punto en coordenadas del mundo que se proyecta en el punto $m = (u, v)$ de la imagen de referencia y en el punto $m' = (u', v')$ de la segunda vista. Si conocemos las matrices de ambas proyecciones $P$ y $P'$, conocemos la relación entre estos tres puntos (ecuaciones A.22 y A.23), de modo que si descomponemos las matrices de proyección en sus parámetros intrínsecos y extrínsecos tenemos

$$s\tilde{m} = \begin{bmatrix} A & 0_3 \end{bmatrix} D\tilde{M} \tag{A.27}$$

$$s'\tilde{m}' = \begin{bmatrix} A' & 0_3 \end{bmatrix} D'\tilde{M} \tag{A.28}$$

donde $A, A', D, D'$ contienen los parámetros de calibración intrínsecos y extrínsecos de ambas proyecciones, respectivamente,

$$s\tilde{m} = \begin{bmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0_3^T & 1 \end{bmatrix} \tilde{M} \tag{A.29}$$

$$s'\tilde{m}' = \begin{bmatrix} \alpha'_u & 0 & u'_0 & 0 \\ 0 & \alpha'_v & v'_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R' & t' \\ 0_3^T & 1 \end{bmatrix} \tilde{M} \tag{A.30}$$

Nótese que multiplicar $D$ y $D'$ por $\tilde{M}$ no es más que un cambio de coordenadas de $\tilde{M}$: de las coordenadas del mundo a las coordenadas de cada una de las cámaras (ver ecuaciones A.1 y A.2). El punto $\tilde{M}$ reescrito en coordenadas de cada una de las cámaras lo denotamos en coordenadas proyectivas,

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = D\tilde{M} \quad \text{y} \quad \begin{bmatrix} x'_c \\ y'_c \\ z'_c \\ 1 \end{bmatrix} = D'\tilde{M} \tag{A.31}$$

Los escalares $s$ y $s'$ de las ecuaciones A.27 y A.28 coinciden con la profundidad del punto de la escena en coordenadas de la cámara respectiva, $s = z_c$ y $s' = z'_c$ (ver ec. A.3).

Despejando $\tilde{M}$ en la ecuación A.27 obtenemos

$$\tilde{M} = D^{-1} \begin{bmatrix} z_c A^{-1} \tilde{m} \\ 1 \end{bmatrix} \tag{A.32}$$

y sustituyendo en la ecuación A.28 obtenemos

$$s'\tilde{m}' = \begin{bmatrix} A' & 0_3 \end{bmatrix} D' D^{-1} \begin{bmatrix} z_c A^{-1} \tilde{m} \\ 1 \end{bmatrix} \tag{A.33}$$

donde $z_c$ es la profundidad estimada del pixel $m$ en coordenadas de la cámara de la imagen de referencia.

### A.4.1 Caso general

Vamos a desarrollar la ecuación A.33 para el caso general de dos cámaras situadas en cualquier posición. Podemos asumir sin falta de generalidad que el sistema de coordenadas del mundo coincide con el sistema de coordenadas de la cámara de referencia. De este modo, $D = I$, y en consecuencia $D^{-1} = I$.

$$\begin{aligned} s'\tilde{m}' &= \begin{bmatrix} A' & 0_3 \end{bmatrix} D' \begin{bmatrix} (u-u_0)z_c/\alpha_u \\ (v-v_0)z_c/\alpha_v \\ z_c \\ 1 \end{bmatrix} = \\ &= \begin{bmatrix} A' & 0_3 \end{bmatrix} \begin{bmatrix} \frac{r'_{11}}{\alpha_u}(u-u_0)z_c + \frac{r'_{12}}{\alpha_v}(v-v_0)z_c + r_{13}z_c + t'_x \\ \frac{r'_{21}}{\alpha_u}(u-u_0)z_c + \frac{r'_{22}}{\alpha_v}(v-v_0)z_c + r_{23}z_c + t'_y \\ \frac{r'_{31}}{\alpha_u}(u-u_0)z_c + \frac{r'_{32}}{\alpha_v}(v-v_0)z_c + r_{33}z_c + t'_z \\ 1 \end{bmatrix} \end{aligned}$$

Llamando $k_{i1}, k_{i2}, k_{i3}$ a los coeficientes de $uz_c$, $vz_c$ y $z_c$, respectivamente,

$$k_{i1} = \frac{r'_{i1}}{\alpha_u} \tag{A.34}$$

$$k_{i2} = \frac{r'_{i2}}{\alpha_v} \tag{A.35}$$

$$k_{i3} = r'_{i3} - \frac{r'_{i1}}{\alpha_u}u_0 - \frac{r'_{i2}}{\alpha_v}v_0 \tag{A.36}$$

y llamando

$$\begin{aligned} w_1 &= k_{11}uz_c + k_{12}vz_c + k_{13}z_c + t'_x \\ w_2 &= k_{21}uz_c + k_{22}vz_c + k_{23}z_c + t'_y \\ w_3 &= k_{31}uz_c + k_{32}vz_c + k_{33}z_c + t'_z \end{aligned}$$

tenemos que

$$s' \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} A' & 0_3 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_u w_1 + u_0 w_3 \\ \alpha_v w_2 + v_0 w_3 \\ w_3 \end{bmatrix}$$

de modo que

$$u' = \frac{\alpha'_u w_1 + u'_0 w_3}{w_3} = \alpha'_u \frac{w_1}{w_3} + u'_0$$
$$v' = \frac{\alpha'_v w_2 + v'_0 w_3}{w_3} = \alpha'_v \frac{w_1}{w_3} + v'_0$$

Finalmente, $(u', v')$ depende de $(u, v)$ y de $z_c$ de la siguiente manera:

$$u' = \alpha'_u \frac{k_{11} u z_c + k_{12} v z_c + k_{13} z_c + t'_x}{k_{31} u z_c + k_{32} v z_c + k_{33} z_c + t'_z} + u'_0 \tag{A.37}$$

$$v' = \alpha'_v \frac{k_{21} u z_c + k_{22} v z_c + k_{23} z_c + t'_y}{k_{31} u z_c + k_{32} v z_c + k_{33} z_c + t'_z} + v'_0 \tag{A.38}$$

donde los valores $k_{ij}$ dependen únicamente de los parámetros de calibración (ecuaciones A.34, A.35 y A.36).

### A.4.2 Caso particular de geometría paralela

En el caso particular de la geometría epipolar paralela ocurre que las dos cámaras tienen una matriz de rotación idéntica respecto a las coordenadas del mundo y sólo difieren en la traslación. El caso más habitual consiste en una diferencia en la traslación respecto al eje $X_c$, que llamaremos $b$. De nuevo, se puede asumir sin falta de generalidad que las coordenadas del mundo se encuentran en el sistema de coordenadas de la cámara de referencia, y por tanto, $D$ coincide con la matriz identidad, mientras que $D'$ consiste en una traslación $b$ en el eje de abscisas.

$$r'_{ii} = 1, \forall i = 1..3 \tag{A.39}$$
$$r'_{ij} = 0, \forall i \neq j \tag{A.40}$$
$$t'_x = b \tag{A.41}$$
$$t'_y = t'_z = 0 \tag{A.42}$$

De la ecuación A.33 obtenemos

$$s'\tilde{m}' = \begin{bmatrix} A' & 0_3 \end{bmatrix} \begin{bmatrix} z_c A^{-1} \tilde{m} + \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix} \\ 1 \end{bmatrix} \tag{A.43}$$

Sustituyendo el valor de la matriz $A$ y su inversa (ecuaciones A.10 y A.15), obtenemos

$$s'\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha'_u & 0 & u'_0 & 0 \\ 0 & \alpha'_v & v'_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} z_c \begin{bmatrix} \frac{u-u_0}{\alpha_u} \\ \frac{v-v_0}{\alpha_v} \\ 1 \end{bmatrix} + \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix} \\ 1 \end{bmatrix} \tag{A.44}$$

de modo que finalmente se obtiene,

$$s'\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\alpha'_u}{\alpha_u} z_c(u-u_0) + z_c u'_0 + \alpha'_u b \\ \frac{\alpha'_v}{\alpha_v} z_c(v-v_0) + z_c v'_0 \\ z_c \end{bmatrix} \tag{A.45}$$

de modo que, $s' = z_c$, y los valores de $u'$ y $v'$ son

$$u' = \frac{\alpha'_u}{\alpha_u}(u-u_0) + u'_0 + \frac{\alpha'_u b}{z_c} \tag{A.46}$$

$$v' = \frac{\alpha'_v}{\alpha_v}(v-v_0) + v'_0 \tag{A.47}$$

Este resultado se puede obtener también de las ecuaciones A.37 y A.38 sabiendo que los valores de la matriz de rotación de $D'$ son $r_{11} = r_{22} = r_{33} = 1, r_{ij} = 0 \forall i \neq j$, y el vector de traslación es $t_x = b, t_y = t_z = 0$.

Si asumimos que ambas imágenes se han obtenido con la misma cámara o que las cámaras son idénticas, entonces $\alpha_u = \alpha'_u$, $\alpha_v = \alpha'_v$, $u_0 = u'_0$ y $v_0 = v'_0$, quedando

$$u' = u + \frac{\alpha'_u b}{z_c} \tag{A.48}$$

$$v' = v \tag{A.49}$$

La diferencia entre las abscisas de los puntos correspondientes, $u' - u = \frac{\alpha'_u b}{z_c}$, es la disparidad, que es inversamente proporcional a la distancia del punto tridimensional al plano imagen.

# Bibliografía


[Ayache y Faverjon, 1987] N. Ayache y B. Faverjon. Efficient registration of stereo images by matching graph descriptions of edge segments. *Int. J. of Computer Vision*, 1(2):107–131, Abril 1987.

[Ayache y Lustman, 1987] N. Ayache y F. Lustman. Fast and reliable passive trinocular stereovision. En *Proc. of the 1st Int. Conf. on Computer Vision*, páginas 422–427, 1987.

[Ayache, 1991] Nicholas Ayache. *Artificial Vision for Mobile Robots.* MIT Press, 1991.

[Badenas *et al.*, 1997] J. Badenas, M. Bober, y F. Pla. Motion and intensity-based segmentation and its applications to traffic monitoring. En del Bimbo [1997], páginas 502–509.

[Baker y Binford, 1981] H.H. Baker y T.O. Binford. Depth from edge- and intensity-based stereo. En *Proceedings 7th Joint Conference on Artificial Intelligence, Vancouver, Canada*, páginas 631–636, Agosto 1981.

[Balas y Yu, 1986] Egon Balas y Chang Sung Yu. Finding a maximum clique in an arbitrary graph. *SIAM J. Appl. Math.*, 15(4):126–135, Noviembre 1986.

[Barnard y Fishler, 1982] S. T. Barnard y M. A. Fishler. Computational stereo. *ACM Computing Surveys*, 14(4):553–572, Diciembre 1982.

[Barnard y Thompson, 1980] Stephen T. Barnard y William B. Thompson. Disparity analysis of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2(4):333–340, Julio 1980.

[Beardsley *et al.*, 1994] Paul Beardsley, Andrew Zisserman, y David Murray. Navigation using affine structure from motion. En Eklundh [1994], páginas 85–96.

[Belhumeur y Mumford, 1992] Peter N. Belhumeur y David Mumford. A bayesian treatment of the stereo correspondence problem using half-occluded regions. En CVPR'92 [1992].

[Belhumeur, 1993a] Peter N. Belhumeur. A binocular stereo algorithm for reconstructing sloping, creased, and broken surfaces in the presence of half-occlusion. En *Proc. of the*

*4th Int. Conf. on Computer Vision*, Berlin, Germany, Mayo 1993. IEEE Computer Society Press.

[Belhumeur, 1993b] Peter Neil Belhumeur. *A Bayesian Approach to the Stereo Correspondence Problem.* PhD thesis, Electrical Engineering, Yale University, Mayo 1993.

[Bensrhair *et al.*, 1996] A. Bensrhair, P. Miche, y R. Debrie. Fast and automatic stereo vision matching algorithm based on dynamic programming method. *Pattern Recognition Letters*, 17(5):457–466, 1996.

[Besl y Jain, 1988] P. J. Besl y R. C. Jain. Segmentation through variable-order surface fitting. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 10(2):167–192, Marzo 1988.

[Birchfield y Tomasi, 1998] Stan Birchfield y Carlo Tomasi. Depth Discontinuities by Pixel-to-Pixel Stereo. En *Proceedings of the IEEE International Conference on Computer Vision, Bombay, India*, 1998.

[Bolles *et al.*, 1993] R.C. Bolles, H.H. Baker, y M.J. Hannah. The JISCT stereo evaluation. En *Proc. of the ARPA Image Understanding Workshop*, páginas 263–274, Washington, DC, Abril 18-21 1993. Morgan Kaufmann.

[Bolles y Cain, 1982] Robert C. Bolles y Ronald A. Cain. Recognizing and locating partially visible objects: The local-feature-focus method. *Int. J. of Robotics Research*, 1(3):57–82, 1982.

[Bolles y Haraud, 1986] Robert C. Bolles y P. Haraud. 3DPO: A three dimensional part orientation system. *Int. J. of Robotics Research*, 5(3):3–26, 1986.

[Bolles, 1979] Robert C. Bolles. Robust feature matching through maximal cliques. *Imaging Applications for Automated Industrial Inspection and Assembly*, 182:140–149, 1979.

[Burt y Julesz, 1980] P. Burt y B. Julesz. Disparity gradient limit for binocular fusion. *Science*, 208:615–617, 1980.

[Castan y Shen, 1984] S. Castan y J. Shen. A stereo vision algorithm taking into account the perspective distortions. En *Proc. of the 7th Int. Conf. on Pattern Recognition*, volumen 1, páginas 444–446, Montreal, Canada, Julio 1984.

[Chabbi y Berger, 1996] Houda Chabbi y Marie Odile Berger. Using projective geometry to recover planar surfaces in stereovision. *Pattern Recognition*, 29(4):533–548, 1996.

[Chang *et al.*, 1993] Yuh-Lin Chang, Xavier Lebegue, y J. K. Aggarwal. Calibrating a mobile camera's parameters. *Pattern Recognition*, 26(1):75–88, 1993.

[Chen y Lin, 1997] Liang-Hua Chen y Wei-Chung Lin. Visual surface segmentation from stereo. *Image and Vision Computing*, 15:95–106, 1997.

[Christmas *et al.*, 1995] William J. Christmas, Josef Kittler, y Maria Petrou. Structural matching in computer vision using probabilistic relaxation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(8):749–764, Agosto 1995.

[Christmas, 1995] William J. Christmas. *Structural matching in computer vision using probabilistic reasoning.* PhD thesis, University of Surrey, Septiembre 1995.

[Cohen *et al.*, 1989a] L. Cohen, A. Gagalowicz, P. T. Sander, y L. Vinet. Cooperative segmentation and region matching. En *Topical Meeting on Image Understanding and Machine Vision*, páginas North Falmouth, Massachussets, Junio 1989.

[Cohen *et al.*, 1989b] L. Cohen, L. Vinet, P.T. Sander, y A. Gagalowicz. Hierarchical region based stereo matching. En *Proc. of the IEEE Conf. Computer Vision and Pattern Recognition*, páginas 416–421, San Diego, CA, Junio 1989. IEEE Computer Society Press.

[Cox *et al.*, 1992] Ingemar J. Cox, Sunita L. Hingorani, Satish B. Rao, y Bruce M. Maggs. Stereo without disparity gradient smoothing: A bayesian sensor fusion solution. En *Proc. of the 3th British Machine Vision Conference*, páginas 337–346, 1992.

[Cox *et al.*, 1996] Ingemar J. Cox, Sunita L. Hingorani, Satish B. Rao, y Bruce M. Maggs. A maximum likelihood stereo algorithm. *Computer Vision and Image Understanding*, 63(3):542–567, 1996.

[CVPR'91, 1991] Lahaina, Hawai, Junio 1991. IEEE Computer Society Press.

[CVPR'92, 1992] Urbana Champaign, IL, Junio 1992. IEEE Computer Society Press.

[Davies, 1991] E. R. Davies. The minimal match graph and its use to speed identification of maximal cliques. *Signal Processing*, 22(3):329–343, Marzo 1991.

[de Agapito, 1996] L. de Agapito. *Estrategias de correspondencia jerárquica y métodos directos de autocalibración para un sistema estereoscópico binocular.* PhD thesis, Instituto de Automática Industrial, Octubre 1996.

[del Bimbo, 1997] A. del Bimbo, editor. volumen 1310 de *Lecture Notes in Computer Science*, Florence, Italy, Mayo 1997. Springer Verlag.

[Deriche *et al.*, 1994] R. Deriche, Z. Zhang, Q.-T. Luong, y O. Faugeras. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. En Eklundh [1994], páginas 567–576, Vol. 1.

[Deriche, 1987] Rachid Deriche. Using canny's criteria to derive a recursively implemented optimal edge detector. *Int. J. of Computer Vision*, 1(2):167–187, Mayo 1987.

[Devernay y Faugeras, 1994] Frédéric Devernay y Olivier Faugeras. Computing differential properties of 3-D shapes from stereoscopic images without 3-D models. En *Proc. of the IEEE Conf. Computer Vision and Pattern Recognition*, páginas 208–213, Seattle, WA, Junio 1994. IEEE Computer Society Press. Also Research Report 2304, INRIA Sophia-Antipolis.

[Devernay y Faugeras, 1995] Frédéric Devernay y Olivier Faugeras. Automatic calibration and removal of distortion from scenes of structured environments. En Leonid I. Rudin y Simon K. Bramble, editores, *Investigative and Trial Image Processing*, volumen 2567 de *Proc. SPIE*, San Diego, CA, Julio 1995. SPIE.

[Devernay y Faugeras, 1996] Frédéric Devernay y Olivier Faugeras. From projective to euclidean reconstruction. En *Proc. of the IEEE Conf. Computer Vision and Pattern Recognition*, páginas 264–269, San Francisco, CA, Junio 1996. IEEE Computer Society Press.

[Devernay, 1996] Frédéric Devernay. *Vision stéréoscopique et propriétés différentielles des surfaces*. PhD thesis, École Polytechnique, Octubre 1996.

[Dhond y Aggarwal, 1989] Umesh R. Dhond y J.K. Aggarwal. Structure from stereo - a review. *IEEE Trans. on Systems, Man, and Cybernetics*, 19(6):1489–1510, 1989.

[Díaz Hernando, 1991] J. A. Díaz Hernando. *Cálculo Integral. Integrales y Series*. Editorial Tebar Flores, 1991.

[Eklundh, 1994] J-O. Eklundh, editor. volumen 800-801 de *Lecture Notes in Computer Science*, Stockholm, Sweden, Mayo 1994. Springer Verlag.

[El-Sonbaty y Ismail, 1998] Yasses El-Sonbaty y M. A. Ismail. A new algorithm for subgraph optimal isomorphism. *Pattern Recognition*, 31(2):205–218, Febrero 1998.

[Elsgoltz, 1977] L. Elsgoltz. *Ecuaciones diferenciales y cálculo variacional*. Editorial MIR, Moscu, 1977.

[Faugeras *et al.*, 1992] Olivier Faugeras, Pascal Fua, Bernard Hotz, Ruihua Ma, Luc Robert, Monique Thonnat, y Zhengyou Zhang. Quantitative and qualitative comparison of some area and feature-based stereo algorithms. En Wolfgang Förstner y Stephan Ruwiedel, editores, *Robust Computer Vision: Quality of Vision Algorithms*, páginas 1–26. Wichmann, Karlsruhe, Germany, 1992.

[Faugeras *et al.*, 1993] Olivier Faugeras, Bernard Hotz, Hervé Mathieu, Thierry Viéville, Zhengyou Zhang, Pascal Fua, Eric Théron, Laurent Moll, Gérard Berry, Jean Vuillemin, Patrice Bertin, y Catherine Proy. Real time correlation based stereo: algorithm implementations and applications. *Int. J. of Computer Vision*, 1993. also INRIA Technical Report 2013.

[Faugeras, 1993] Olivier Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press, 1993.

[Faugeras, 1995] Olivier Faugeras. Stratification of 3-D vision: projective, affine, and metric representations. *J. of the Optical Society of America*, 12(3):465–484, Marzo 1995.

[Froehlinghaus, 1997] Thorsten Froehlinghaus. Stereo images with ground truth disparity and occlusion. `http://www-dbv.cs.uni-bonn.de/~ft/stereo.html`, Agosto 1997.

[Fua, 1991] Pascal Fua. Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities. En *Int. Joint Conf. on Artificial Intelligence*, Sydney, Australia, Agosto 1991.

[Fua, 1993] Pascal Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*, 6(1), 1993. Available as INRIA research report 1369.

[Fusiello *et al.*, 1997a] A. Fusiello, V. Roberto, y E. Trucco. Efficient stereo with multiple windowing. En *Proc. of the IEEE Conf. Computer Vision and Pattern Recognition*, páginas 858–863, Puerto Rico, Junio 1997. IEEE Computer Society Press.

[Fusiello *et al.*, 1997b] A. Fusiello, V. Roberto, y E. Trucco. Experiments with a new area-based stereo algorithm. En del Bimbo [1997], páginas 669–676.

[Gamble y Poggio, 1987] E. Gamble y T. Poggio. Visual integration and detection of discontinuities: The key role of intensity edges. A.I.Memo 970, MIT Artificial Intelligence Laboratory, 1987.

[Geiger *et al.*, 1992] Davi Geiger, Bruce Ladendorf, y Alan Yuile. Occlusions and binocular stereo. En G. Sandini, editor, *Proc. of the 2nd European Conf. on Computer Vision*, volumen 588 de *Lecture Notes in Computer Science*, páginas 425–433, Santa Margherita Ligure, Italy, Mayo 1992. Springer Verlag.

[Geiger *et al.*, 1995] Davi Geiger, Bruce Ladendorf, y Alan Yuile. Occlusions and binocular stereo. *Int. J. of Computer Vision*, páginas 221–226, 1995.

[Gennery, 1980] D.B. Gennery. *Modelling the Environment of an Exploring Vehicle by means of Stereo Vision*. PhD thesis, Stanford University, Junio 1980.

[Grimson, 1981] W.E.L. Grimson. A computer implementation of a theory of human stereo vision. *Philosophical Trans. of the Royal Society of London, B.*, 292(1058):217–253, 1981.

[Grimson, 1985] W.E.L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7(1):17–34, 1985.

[Hartley *et al.*, 1992] Richard Hartley, Rajiv Gupta, y Tom Chang. Stereo from uncalibrated cameras. En CVPR'92 [1992], páginas 761–764.

[Herault *et al.*, 1990] Laurent Herault, Radu Horaud, Francoise Veillon, y Jean-Jacques Niez. Symbolic image matching by simulated annealing. En *Proc. of the 1st British Machine Vision Conference*, páginas 319–324, Oxford, England, 1990.

[Hollinghurst y Cipolla, 1994] Nicholas Hollinghurst y Roberto Cipolla. Uncalibrated stereo hand-eye coordination. *Image and Vision Computing*, 12(3):187–192, Abril 1994.

[Horaud y Skordas, 1989] R. Horaud y Th. Skordas. Stereo correspondence through feature grouping and maximal cliques. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(11):1168–1180, 1989.

[Intille y Bobick, 1994a] Stephen S. Intille y Aaron F. Bobick. Disparity-space images and large occlusion stereo. En Eklundh [1994], páginas 674–677. Extended version in M.I.T Media Lab Computing Group Technical Report No. 220.

[Intille y Bobick, 1994b] Stephen S. Intille y Aaron F. Bobick. Incorporating intensity edges in the recovery of occlusion regions. En *Int. Conf. on Pattern Recognition*, volumen A, páginas 674–677, Octubre 1994. Also M.I.T Media Lab Computing Group Technical Report No. 246.

[Jones, 1997] Graeme A. Jones. Constraint, optimization, and hierarchy: Reviewing stereoscopic correspondence of complex features. *Computer Vision and Image Understanding*, 65(1):57–78, Enero 1997.

[Kanade y Okutomi, 1990] Takeo Kanade y Masatoshi Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiments. En *Proc. of the Image Understanding Workshop*, páginas 383–398, PA, Septiembre 1990. DARPA, Science App. Int. Corp.

[Kanade y Okutomi, 1994] Takeo Kanade y Masatoshi Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(9):920–932, Septiembre 1994.

[Kitamura y Yachida, 1990] Yoshifumi Kitamura y Masahiko Yachida. Three-dimensional data acquisition by trinocular vision. *Advanced Robotics*, 4(1):29–42, 1990. Robotics Society of Japan.

[Lacey *et al.*, 1998] A. J. Lacey, N. A. Thacker, S. Crossley, y R. B. Yates. A multi-stage approach to the dense estimation of disparity from stereo sem images. *Image and Vision Computing*, 16:373–383, 1998.

[Lane y Thacker, 1994] R. A. Lane y N. A. Thacker. Stereo vision research: An algorithm survey. Technical Report 94/16, University of Sheffield, Electronic Systems Group, 1994.

[Lee *et al.*, 1993] C.-Y. Lee, D.B. Cooper, y D. Keren. Computing correspondence based on region and invariants without feature extraction and segmentation. En *Proc. of the IEEE Conf. Computer Vision and Pattern Recognition*, páginas 655–656. IEEE Computer Society Press, Junio 1993.

[Li, 1994] Ze-Nian Li. Stereo correspondence based on line matching in hough space using dynamic programming. *IEEE Trans. on Systems, Man, and Cybernetics*, 24(1):144–152, jan 1994.

[Lotti, 1996] Jean-Luc Lotti. *Mise en correspondance stéréo par fenêtres adaptatives en imagerie haute résolution.* PhD thesis, Universite de Nice-Sophia Antipolis, France, Febrero 1996.

[Luong y Faugeras, 1996] Quang-Tuan Luong y Olivier Faugeras. The fundamental matrix: theory, algorithms, and stability analysis. *Int. J. of Computer Vision*, 17(1):43–76, Enero 1996.

[Marapane y Trivedi, 1989] Suresh B. Marapane y Mohan M. Trivedi. Region-based stereo analysis for robotics applications. *IEEE Trans. on Systems, Man, and Cybernetics*, 19(6):1447–1464, Noviembre 1989.

[Marapane y Trivedi, 1992] Suresh B. Marapane y Mohan M. Trivedi. Multi-primitive hierarchical (mph) stereo analysis. En CVPR'92 [1992], páginas 499–505.

[Marapane y Trivedi, 1994] Suresh B. Marapane y Mohan M. Trivedi. Multi-primitive hierarchical (mph) stereo analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(3):227–240, 1994.

[Marr y Poggio, 1976] D. Marr y T. Poggio. Cooperative computation of stereo disparity. *Science*, 194:283–287, 1976. introduccion muy interesante.

[Marr y Poggio, 1977] D. Marr y T. Poggio. A theory of human stereo vision. A.I.Memo 451, MIT Artificial Intelligence Laboratory, 1977.

[Marr y Poggio, 1979] D. Marr y T. Poggio. A computational theory of human stereo vision. *Proc. R. Soc.*, B-204:301–328, 1979.

[Mayhew y Frisby, 1981] J.E.W. Mayhew y J.P. Frisby. Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence*, 17:349–386, 1981.

[Menard y Kropatsch, 1997] Christian Menard y Walter G. Kropatsch. Adaptive stereo matching in correlation scale-space. En del Bimbo [1997], páginas 677–684.

[Mohr *et al.*, 1993] Roger Mohr, Boubakeur Boufama, y Pascal Brand. Accurate projective reconstruction. En Joseph Mundy y Andrew Zisserman, editores, *Applications of Invariance in Computer Vision*, volumen 825 de *Lecture Notes in Computer Science*, páginas 257–276, Berlin, 1993. Springer Verlag.

[Moravec, 1977] H. P. Moravec. Towards automatic visual obstacle avoidance. En *Proc. of the 5th Int. Joint Conf. on Artificial Intelligence*, página 584, 1977.

[Mundy y Zisserman, 1992] Joseph L. Mundy y Andrew Zisserman, editores. *Geometric Invariance in Computer Vision.* MIT Press, 1992.

[Nasrabadi y Liu, 1989] N. M. Nasrabadi y Y. Liu. Stereo vision correspondence using a multichannel graph matching technique. *Image and Vision Computing*, 7(4):237–245, Noviembre 1989.

[Nasrabadi, 1992] N. M. Nasrabadi. A stereo vision technique using curve segments and relaxation matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(5):566–572, Mayo 1992.

[Nishihara, 1984a] Keith Nishihara. Practical real-time imaging stereo matcher. *Optical Engineering*, 23(5), 1984.

[Nishihara, 1984b] Keith Nishihara. PRISM: A practical real-time imaging stereo matcher. A.I.Memo 773, MIT, 1984.

[Nomura *et al.*, 1992] Yoshihiko Nomura, Michihiro Sagara, Hiroshi Naruse, y Atsushi Ide. Simple calibration algorithm for high-distortion-lens camera. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(11):1095–1099, Noviembre 1992.

[Ohta y Kanade, 1985] Y. Ohta y T. Kanade. Stereo by intra- and inter-scanline search. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 7:139–154, 1985.

[Okutomi y Kanade, 1992] Masatoshi Okutomi y Takeo Kanade. A locally adaptive window for signal matching. *Int. J. of Computer Vision*, 7(2):143–162, 1992.

[Okutomi y Kanade, 1993] Masathoshi Okutomi y Takeo Kanade. A multiple-baseline stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(4):353–363, Abril 1993.

[Oren y Nayar, 1994] Michael Oren y Shree K. Nayar. Generalization of lambert's reflectance model. En Andrew Glassner, editor, *Proc. of the SIGGRAPH*, páginas 239–246, Orlando, Florida, Julio 1994.

[Pajares *et al.*, 1998] G. Pajares, J. M. Cruz, y J. Aranda. Relaxation by hopfield network in stereo image matching. *Pattern Recognition*, 31(5):561–574, 1998.

[Papadimitriou y Dennis, 1995] D. V. Papadimitriou y T. J. Dennis. A stereo disparity algorithm for 3d model construction. En *Proc. of the Image Processing and its Applications*, páginas 178–182, Julio 1995.

[Pollard *et al.*, 1985] S.B. Pollard, J.E.W. Mayhew, y J.P. Frisby. PMF : a stereo correspondence algorithm using a disparity gradient constraint. *Perception*, 14:449–470, 1985.

[Prazdny, 1985] K. Prazdny. Detection of binocular disparities. *Biological Cybernetics*, 52:93–99, 1985.

[Quan, 1993] Long Quan. Affine stereo calibration for relative affine shape reconstruction. En *Proc. of the 4th British Machine Vision Conference*, páginas 659–668, Surrey, England, 1993.

[Randriamasy y Gagalowicz, 1991] S. Randriamasy y A. Gagalowicz. Region based stereo matching oriented image processing. En CVPR'91 [1991], páginas 736–737.

[Ranganath y Chipman, 1992] Heggere S. Ranganath y Laure J. Chipman. Fuzzy relaxation approach for inexact scene matching. *Image and Vision Computing*, 10(9):631–640, Noviembre 1992.

[Robert *et al.*, 1995] L. Robert, M. Buffa, y M. Hebert. Weakly-calibrated stereo perception for rover navigation. En *Proc. of the 5th Int. Conf. on Computer Vision*, páginas 46–51, Boston, MA, Junio 1995. IEEE Computer Society Press.

[Robert *et al.*, 1997] L. Robert, C. Zeller, O. Faugeras, y M. Hébert. Applications of non-metric vision to some visually-guided robotics tasks. En Y. Aloimonos, editor, *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles*, capítulo 5, páginas 89–134. Lawrence Erlbaum Associates, 1997. also INRIA Technical Report 2584.

[Robert y Deriche, 1996] L. Robert y R. Deriche. Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities. En Bernard Buxton, editor, *Proc. of the 4th European Conf. on Computer Vision*, Cambridge, UK, Abril 1996.

[Robert y Faugeras, 1991] Luc Robert y Olivier Faugeras. Curve-based stereo: Figural continuity and curvature. En CVPR'91 [1991], páginas 57–62.

[Robert y Faugeras, 1995] Luc Robert y Olivier Faugeras. Relative 3-D positioning and 3-D convex hull computation from a weakly calibrated stereo pair. *Image and Vision Computing*, 13(3):189–197, 1995. also INRIA Technical Report 2349.

[Robert, 1995] Luc Robert. Camera calibration without feature extraction. *Computer Vision and Image Understanding*, 63(2):314–325, Marzo 1995. also INRIA Technical Report 2204.

[Rojas *et al.*, 1997] Angela Rojas, Antonio Calvo, y Jose Mu noz. A dense disparity map of stereo images. *Pattern Recognition Letters*, 18:385–393, 1997.

[Rosenfeld y Kak, 1982] A. Rosenfeld y A.C. Kak. *Digital Picture Processing*, volumen 1. Academic Press, New York, 1982. Second Edition.

[Rothwell *et al.*, 1995] C. A. Rothwell, A. Zisserman, D. A. Forsyth, y J. L. Mundy. Planar object recognition using projective shape representation. *Int. J. of Computer Vision*, 16(1):57–99, Septiembre 1995.

[Sakamoto *et al.*, 1997] Shizuo Sakamoto, Ingemar J. Cox, y Johji Tajima. A multiple-baseline stereo for precise human face acquisition. *Pattern Recognition Letters*, 18(9):923–931, Septiembre 1997.

[Sander *et al.*, 1989] Peter T. Sander, Laurent Vinet, Laurent Cohen, y A. Gagalowicz. Hierarchical region based stereo matching. En *Proc. of the 6th Scandinavian Conf. on Image Analysis*, páginas 71–78, Oulu, Finland, Junio 1989.

[Satoh *et al.*, 1996] Kiyohide Satoh, Itaru Kitahara, y Yuichi Ohta. 3D image display with motion parallax by camera matrix stereo. 1996.

[Satoh y Ohta, 1994] Kiyohide Satoh y Yuichi Ohta. Passive depth adquisition for 3d image displays. *IEEE Trans. on Information and Systems*, (9):949–957, 1994.

[Satoh y Ohta, 1995] Kiyohide Satoh y Yuichi Ohta. Occlusion detectable stereo using a camera matrix. En *Proc. of the 2nd Asian Conf. on Computer Vision*, volumen 2, páginas 331–335, 1995.

[Satoh y Ohta, 1996] Kiyohide Satoh y Yuichi Ohta. Occlusion detectable stereo —systematic comparison of detection algorithms—. En *Proc. of the 13th Int. Conf. on Pattern Recognition*, Vienna, Austria, Agosto 1996.

[Sonka *et al.*, 1993] Milan Sonka, Vaclav Hlavac, y Roger Boyle, editores. *Image Processing, Analysis and Machine Vision.* Chapman & Hall, 1993.

[Suganthan *et al.*, 1998] P. N. Suganthan, H. Yan, E. K. Teoh, y D. P. Mital. Optimal encoding of graph homomorphism energy using fuzzy information aggregation operators. *Pattern Recognition*, 31(5):623–639, 1998.

[Super y Klarquist, 1997] Boaz J. Super y William N. Klarquist. Patch-based stereo in a general binocular geometry. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(3):247–253, 1997.

[Tarel y Vézien, 1995] Jean-Philippe Tarel y Jean-Marc Vézien. A generic approach for planar patches stereo reconstruction. En *Proc. of the 11th Scandinavian Conf. on Image Analysis*, páginas 1061–1070, Norway, Junio 1995.

[Tarel, 1996] Jean-Philippe Tarel. Reconstruction globale et robuste de facettes 3D. Rapport de recherche 2813, INRIA, Febrero 1996.

[Terzopoulos, 1986] Demetri Terzopoulos. Regularization of inverse visual problems involving discontinuities. páginas 183–194. 1986.

[Tsai, 1987] Roger Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE J. of Robotics and Automation*, 3(4):323–344, Agosto 1987.

[Vézien y Gagalowicz, 1991] Jean-Marc Vézien y André Gagalowicz. Reconstruction 3-D basée sur une analyse en régions d'une paire d'images stéréoscopique. En *Actes de AFCET-RFIA*, Lyon, France, 1991.

[Viola, 1997] Paul Viola. Alignment by maximization of mutual information. *Int. J. of Computer Vision*, 1997.

[Wildes, 1991] Richard P. Wildes. Direct recovery of three-dimensional scene geometry from binocular stereo disparity. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(8):761–774, Agosto 1991.

[Wilson y Hancock, 1997] Richard C. Wilson y Edwin R. Hancock. Structural matching by discrete relaxation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19:634–648, Junio 1997.

[Xu y Zhang, 1996] Gang Xu y Zhengyou Zhang. *Epipolar Geometry in Stereo, Motion and Object Recognition.* Kluwer Academic Publishers, 1996.

[Yang *et al.*, 1989] B. Yang, W. E. Snyder, y G. L. Bilbro. Matching oversegmented 3d images to models using association graphs. *Image and Vision Computing*, 7(2):140–149, Mayo 1989.

[Zhang *et al.*, 1995] Zhengyou Zhang, Rachid Deriche, Olivier Faugeras, y Quang-Tuan Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78(1-2):87–119, Octubre 1995. Also INRIA Research Report No.2273, May 1994.

[Zhang y Faugeras, 1992] Zhengyou Zhang y Olivier D. Faugeras. *3D Dynamic Scene Analysis: A Stereo Based Approach.* Springer, Berlin, Heidelberg, 1992.

[Zhang, 1993] Zhengyou Zhang. Le problème de la mise en correspondance: L'état de l'art. Technical Report RR 2146, INRIA, Diciembre 1993.

[Zhang, 1996a] Z. Zhang. A stereovision system for a planetary rover: Calibration, correlation, registration, and fusion. En *Proceedings IEEE Workshop on Planetary Rover Technology and Systems*, Minneapolis, Minnesota, USA, Abril 1996.

[Zhang, 1996b] Zhengyou Zhang. Determining the epipolar geometry and its uncertainty: A review. Technical Report 2927, INRIA Sophia-Antipolis, France, Julio 1996.

[Zhang, 1996c] Zhengyou Zhang. On the epipolar geometry between two images with lens distortion. En *Int. Conf. on Pattern Recognition*, volumen I, páginas 407–411, Vienna, Austria, Agosto 1996.