



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA

---

**RECONSTRUCCION CONFIABLE  
DE SUPERFICIES USANDO RANGO  
DE DISPARIDAD ADAPTIVO**

**MIGUEL ATTILIO TORRES TORRITI**

Tesis para optar al grado de  
Magister en Ciencias de la Ingeniería

Profesor Supervisor:  
**ANDRES GUESALAGA M.**

Santiago de Chile, 1998



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA  
Departamento de Ingeniería Eléctrica

---

# **RECONSTRUCCION CONFIABLE DE SUPERFICIES USANDO RANGO DE DISPARIDAD ADAPTIVO**

**MIGUEL ATTILIO TORRES TORRITI**

Tesis presentada a la Comisión integrada por los profesores:

**ANDRES GUESALAGA M.**

**PABLO IRARRAZAVAL M.**

**RENATO SALINAS S.**

**VLADIMIR MARIANOV K.**

Para completar las exigencias del grado  
de Magister en Ciencias de la Ingeniería

Santiago de Chile, 1998

Keywords: Stereo Vision, Stereo Matching, Adaptive Disparity Range, Stereoscopic Sequence Analysis, 3D Reconstruction, 3D Scene Analysis, Computer Vision, Machine Vision, Robot Vision.

A mi familia.

## **AGRADECIMIENTOS**

Quiero agradecer a todos aquellos que colaboraron en este trabajo, especialmente al Profesor Andrés Guesalaga por su constante apoyo y sus valiosos consejos, así como a los miembros de la comisión por sus apreciaciones y sugerencias. También debo expresar mis agradecimientos a Cristián Abaroa y a Juan Pablo Duclos, por su importante ayuda.

## INDICE GENERAL

	Pág.
KEYWORDS.....	ii
DEDICATORIA .....	iii
AGRADECIMIENTOS.....	iv
INDICE DE TABLAS.....	ix
INDICE DE FIGURAS .....	x
RESUMEN .....	xv
ABSTRACT .....	xvi
NOTACION.....	xvii
PREFACIO .....	xviii
Objetivos.....	xviii
Resumen de Contribuciones .....	xix
Organización de la Tesis .....	xxi
I. INTRODUCCION .....	1
1.1 Descripción del Problema y Terminología Básica .....	4
1.2 Estereoscopia vs. Otras Técnicas .....	8
1.2.1 Descripción de Métodos Alternativos.....	10
1.2.2 Comparación: Estereoscopia vs. Otras Técnicas .....	15
1.3 Trabajo Relacionado .....	18
II. ESTEREOSCOPIA .....	21
2.1 Adquisición de Imágenes.....	22
2.2 Modelación de las Cámaras.....	23
2.2.1 Coordenadas Tridimensionales.....	23
2.2.2 Proyección en Perspectiva .....	26

2.2.3	Modelación Estereoscópica .....	28
2.2.4	Geometría Epipolar.....	32
2.2.5	Calibración.....	35
2.2.6	Definición de Disparidad y Cálculo de las Coordenadas 3D.....	39
2.3	Elección de Características .....	45
2.4	Búsqueda de Correspondencias.....	46
2.4.1	Métodos de Correspondencia.....	48
2.5	Determinación de Profundidad y Reconstrucción 3D .....	48
2.6	Interpolación.....	51
<b>III.</b>	<b>TRABAJO DESARROLLADO.....</b>	<b>52</b>
3.1	Análisis del Error en la Estimación de la Profundidad.....	53
3.1.1	Efectos de los Errores de Calibración .....	55
3.1.2	Importancia Relativa de los Errores.....	65
3.1.3	Error en la Profundidad Debido a Calibración Incorrecta .....	66
3.1.4	Efectos de los Errores de Muestreo .....	70
3.1.5	Efectos del Movimiento de la Cámaras.....	76
3.1.6	Elección de Parámetros para el Sistema Estéreo .....	81
3.1.7	Ejemplo de Diseño del Sistema .....	89
3.2	Adquisición de las Imágenes .....	90
3.3	Procesamiento Secuencial y Análisis de las Imágenes .....	97
3.3.1	Calibración del Sistema y Rectificación de las Imágenes .....	100
3.3.2	Procesamiento del Par Estéreo: Disparidades y Confiabilidad.....	104
3.3.3	Interpolación .....	121
3.3.4	Procesamiento de la Secuencia de Pares Estereoscópicos.....	124
3.3.5	Ajuste del Rango de Disparidades .....	127
3.3.6	Mediciones y Reconstrucción Tridimensional .....	128
<b>IV.</b>	<b>RESULTADOS EXPERIMENTALES .....</b>	<b>129</b>
4.1	Evaluación de los Algoritmos Estereoscópicos.....	129
4.1.1	Imágenes Sintéticas .....	130
4.1.2	Imágenes de Prueba.....	140
4.2	Reconstrucción y Mediciones Tridimensionales .....	141
4.3	Implementación de los Algoritmos .....	147

V.	CONCLUSIONES Y DISCUSION.....	149
5.1	Resumen de Contribuciones.....	150
5.2	Limitaciones de las Técnicas Utilizadas.....	150
5.3	Desarrollos Futuros .....	151
	<b>BIBLIOGRAFÍA.....</b>	<b>155</b>
	Bibliografía Básica .....	155
I.	Estereoscopía .....	155
II.	Error en la Reconstrucción 3D usando Estereoscopía.....	157
III.	Reconstrucción 3D .....	158
IV.	Información en Internet.....	159
	Bibliografía Complementaria.....	162
I.	Algoritmos y Aplicaciones de Estereoscopía.....	162
II.	Calibración de las Cámaras .....	167
III.	Optic Flow - Shape from Motion - Motion Estimation.....	168
IV.	Aplicaciones basadas en Sonar, Range Finders, Dead Reckoning, GPS .....	168
V.	Gráficos en Computador.....	169
VI.	Procesamiento de Imágenes.....	170
VII.	Instrumentación y Control.....	171
VIII.	Procesamiento Digital de Señales:.....	171
IX.	Robótica .....	171
X.	Manuales .....	171
	<b>A N E X O S.....</b>	<b>173</b>
	<b>ANEXO A: METODOS DE BUSQUEDA DE CORRESPONDENCIAS .....</b>	<b>174</b>
A.1	Métodos Basados en Area .....	174
A.1.1	Correspondencia Basada en Correlación .....	174
A.1.2	Correspondencia a Múltiples Niveles de Resolución.....	175
A.1.3	Programación Dinámica .....	175
A.2	Métodos Basados en Características .....	176
A.3	Algoritmos Basados en Relajación.....	177
A.4	Algoritmos Basados en el Dominio de la Frecuencia .....	177
	<b>ANEXO B: APLICACIONES DE LA ESTEREOSCOPIA .....</b>	<b>178</b>



ANEXO C: ALGORITMOS IMPLEMENTADOS .....	182
C.1 Algoritmo de Correlación Normalizada Acelerada .....	182
C.2 Cálculo de la Confiabilidad.....	185
C.3 Cálculo de la Resolución Sub-Pixel mediante Interpolación Cuadrática.....	190
C.4 Cálculo del Rango de Búsqueda Ajustado Localmente.....	192
GLOSARIO .....	194

## INDICE DE TABLAS

	Pág.
Tabla 1.1: Ventajas y Desventajas de los Métodos para Reconstrucción 3D. ....	17
Tabla 1.2: Comparación de Costos: LRF vs. Sistema de Visión Estéreo. ....	18
Tabla 3.1: Características de las Cámaras. ....	91
Tabla 3.2: Resumen de Características de las Regiones de la figura 3.26.....	109
Tabla 4.1: Evaluación de los Algoritmos Estereoscópicos. ....	136
Tabla 4.2: Cuadro Comparativo de la Correspondencia SSD vs. NCC.....	142
Tabla 4.3: Evaluación de los Métodos de Procesamiento Secuencial. ....	143

## INDICE DE FIGURAS

	Pág.
Figura 1.1: Disparidad.....	4
Figura 1.2: Funciones de correlación. ....	7
Figura 1.3: Laser Measurement System LMS 200 Sick AG. ....	12
Figura 1.4: Pulsos transmitidos por el radar y ecos reflejados ( <i>backscatter</i> ). ....	13
Figura 1.5: Barrido de la antena. ....	14
Figura 2.1: Imagen del punto $P_i$ en la vista de cada cámara. ....	21
Figura 2.2: Vista de planta de una escena tridimensional. ....	24
Figura 2.3: Imagen de la escena de la figura 2.2 vista desde la cámara.....	25
Figura 2.4: Sistema de coordenadas tridimensionales. ....	26
Figura 2.5: Proyección en perspectiva.....	27
Figura 2.6: Ejemplo de par estéreo con desplazamiento lateral de la cámara. ....	28
Figura 2.7: Vista de planta de un sistema estéreo binocular. ....	29
Figura 2.8: Imágenes estereoscópicas de la escena representada en la figura 2.7. ....	30
Figura 2.9: Modelo estereoscópico simple de cámaras paralelas. ....	31
Figura 2.10: Rectas epipolares. ....	33
Figura 2.11: Segmento epipolar.....	34
Figura 2.12: Sistema binocular típico.....	39
Figura 2.13: Vista en perspectiva del sistema binocular. ....	40

Figura 2.14: Vista de planta del sistema binocular.....	41
Figura 2.15: Vista lateral de una cámara. ....	43
Figura 2.16: Mapa de disparidades para la vista de la cámara A de la figura 2.7.....	44
Figura 3.1: Representación geométrica del sistema binocular paralelo.....	57
Figura 3.2: Cambio en la proyección debido a un ángulo de elevación. ....	59
Figura 3.3: Cambio en la proyección debido a un ángulo de convergencia. ....	61
Figura 3.4: Desalineación del CCD.....	62
Figura 3.5: Distorsión de un lente con $f$ : 6.5 mm. ....	64
Figura 3.6a: Error relativo en función de la convergencia y la elevación. ....	69
Figura 3.6b: Error relativo en función de la rotación y la elevación. ....	69
Figura 3.7: Error absoluto debido a un error de $\pm 0.5$ pixels en la disparidad.....	72
Figura 3.8: Error absoluto en la estimación de la distancia.....	74
Figura 3.9: Familia de curvas de disparidad variando el <i>baseline</i> .....	75
Figura 3.10: Variación del rango de disparidades para un desplazamiento en $z$ .....	77
Figura 3.11a: Variación de la disparidad máxima y la disparidad mínima.....	79
Figura 3.11b: Crecimiento del rango <i>Disparidad Máxima - Disparidad Mínima</i> . ....	79
Figura 3.12: Area observable a través del sistema binocular.....	84
Figura 3.13: Regiones comunes en cada imagen del par estéreo. ....	87
Figura 3.14: Cámara Color Cohu 8380.....	91
Figura 3.14: Sistema de adquisición de imágenes. ....	92

Figura 3.15: Marca de alineación. ....	93
Figura 3.16a: Vista de elevación de la pirámide. ....	95
Figura 3.16b: Perspectiva de la pirámide. ....	95
Figura 3.17: Primer par de la secuencia Pira700 (Distancia a la base: 700mm). ....	96
Figura 3.18: Ultimo par de la secuencia Pira700 (Distancia a la base: 655mm). ....	96
Figura 3.19: Varianza de la imagen derecha de la figura 3.18. ....	97
Figura 3.20: Etapas del procesamiento de secuencias estereoscópicas. ....	99
Figura 3.21: Alineación de la secuencia Pira700. ....	102
Figura 3.22: Imagen $R_0-R'_{15}$ . ....	103
Figura 3.23: Señal de la línea horizontal 128 de las imágenes en la figura 3.21. ....	103
Figura 3.24: Cálculo de la correlación normalizada. ....	106
Figura 3.25: Resolución <i>sub-pixel</i> de la disparidad usando interpolación cuadrática ...	107
Figura 3.26: Texturas características en el primer par estéreo de Pira700. ....	108
Figura 3.27: Comparación de las curvas de correlación. ....	111
Figura 3.28: Ambigüedad en la búsqueda de correspondencias. ....	112
Figura 3.29: Correlación máxima en función del tamaño de ventana. ....	115
Figura 3.30: Disparidad en función del tamaño de ventana. ....	116
Figura 3.31: Confiabilidad en función del tamaño de ventana. ....	118
Figura 3.32: Validación y confiabilidad para distintos tamaños de ventana. ....	120
Figura 3.33: Eliminación de puntos no válidos usando filtro de mediana. ....	122

Figura 3.34: Interpolación bilineal en las zonas no correspondidas.....	123
Figura 3.35: Aplicación del método de interpolación bilineal.....	124
Figura 4.1: Vistas de elevación y planta del modelo artificial.....	131
Figura 4.2: Par estéreo sintético Sint1.....	132
Figura 4.3: Mapa de disparidad correspondiente al par Sint1.....	133
Figura 4.3: Modelo 3D del objeto en Sint1.....	133
Figura 4.5: Par estéreo sintético Sint2.....	134
Figura 4.6: Mapa de disparidad correspondiente a Sint2.....	135
Figura 4.8: Representación tridimensional del mapa de disparidades reales de Sint1... 138	138
Figura 4.9: Mapa de disparidades medidas en la prueba 10 a partir del par Sint1.....	138
Figura 4.10: Resultados de la búsqueda de correspondencias.....	139
Figura 4.11: Mapas de confiabilidad del par Sint2.....	140
Figura 4.12: Mapas de confiabilidad de Pira700.....	141
Figura 4.13: Representación con <i>texture mapping</i> de la pirámide de Pira700, $k=0$ .....	144
Figura 4.14: Representación con <i>texture mapping</i> de la pirámide de Pira700, $k=0-15$ . 144	144
Figura 4.15: Representación <i>malla de superficie</i> de la pirámide de Pira700, $k=0$ . ....	145
Figura 4.16: Representación <i>malla de superficie</i> de la pirámide de Pira700, $k=0-15$ ... 145	145
Figura 4.17: Reconstrucción 3D de la escena del muñeco.....	146
Figura 4.18: Malla 3D de la escena del muñeco.....	146
Figura 4.13: Consumo de recursos del sistema por el programa para estereoscopia.....	147

Figura 4.14: Interfaz del programa para análisis estereoscópico.....	148
Figura 5.1: La estereoscopía como componente esencial de un sistema robótico.....	153
Figura A.1: Mapa de costos para dos segmentos epipolares. ....	176
Figura B.1: Par estéreo: Who goes there! © 1878. Por J.P. Soule.....	178
Figura B.2.1: Módulo espacial con el Imager Stereo (IMP) y el Pathfinder.....	179
Figura B.2.2: Mars Pathfinder .....	180
Figura B.2.3: Imager Stereo del Mars Pathfinder (IMP).....	180
Figura B.3: Satélites SPOT: Système Probatoire d'Observation de la Terre.....	181
Figura C.1: Simplificación del cálculo de $C$ . ....	185
Figura C.3: Interpolación cuadrática .....	191
Figura C.4: Determinación del rango de búsqueda en zonas indeterminadas.....	193

## RESUMEN

Este trabajo tiene por objetivo el contribuir al mejoramiento de las técnicas basadas en estereoscopía para la reconstrucción tridimensional de escenas reales.

En la estereoscopía tradicional, las escenas son reconstruidas tridimensionalmente a partir de dos o más imágenes tomadas desde perspectivas distintas. El proceso de reconstrucción tridimensional se basa en el análisis del conjunto de imágenes para encontrar la posición de la proyección de un punto de la escena sobre cada plano óptico (plano sobre el cual se proyecta la imagen). Una vez determinadas las posiciones relativas de los puntos correspondientes (disparidades), mediante el proceso conocido como *búsqueda de correspondencias (matching)* y dada cierta configuración geométrica del sistema de cámaras, es posible determinar la posición del punto en el espacio tridimensional mediante triangulación.

El aspecto de mayor complejidad en el proceso de reconstrucción es el de búsqueda de correspondencias debido a que está sujeto a varias fuentes de error y ambigüedades, que a su vez originan estimaciones erróneas de la posición del punto en el espacio. Por esta razón es necesario desarrollar métodos que permitan reducir la ocurrencia de errores y un aumento de la exactitud y la confiabilidad de las mediciones de profundidad.

Se proponen dos aportes novedosos para contribuir al mejoramiento de los algoritmos estereoscópicos existentes:

- La incorporación de los factores de confiabilidad de las mediciones, obtenidos a partir del análisis de las curvas de correlación.
- La introducción de un método automático para el ajuste del rango de búsqueda de disparidades a partir de una estimación inicial de las disparidades de menor precisión, pero mayor confiabilidad.

También se analiza el uso de secuencias de pares de imágenes tomadas desde distintas posiciones como método para reducir el error de estimación de la profundidad.

El perfeccionar las técnicas de reconstrucción tridimensional es esencial para cualquier aplicación futura que pretenda confiar su automatismo a un sistema de visión artificial.



## ABSTRACT

The objective of this work is to contribute to the enhancement of stereo based techniques for 3D reconstruction of real scenes.

In traditional stereoscopy, a scene is tridimensionally reconstructed from two or more images taken from different perspectives. The process of 3D reconstruction is based upon the analysis of the set of images to find the position of a scene point's projection on each optical plane (plane on which the image is projected). Once the relative positions of the matching points (disparities) have been found, and given a certain geometric configuration of the cameras, by triangulation it is possible to determine the point's position in the tridimensional space.

The most complex aspect in the reconstruction process is the search of corresponding points due to the several sources of error and ambiguity, which in its turn cause erroneous estimations of the point's position in space. For this reason, it is necessary to develop new methods for the reduction of matching errors and to increase the accuracy and reliability of the depth measurements.




The main novel contributions to enhance the existing algorithms for 3D position estimation are the incorporation of:

- The measurements reliance scores, obtained from the analysis of the correlation curves.
- An automatic method for the adjustment of the disparity search range, based on a less precise, but more reliable initial estimation of the disparities.

The use of sequences of image pairs taken from different positions as a method to reduce the depth estimation error is also analysed.

The enhancement of the 3D reconstruction techniques is essential for any future application that would pretend to rely its automation on an artificial vision system.

## NOTACION

$u_i, v_i$	Coordenadas (horizontal, vertical) de la imagen origen $i$
$x_i, y_i$	Coordenadas (horizontal, vertical) de la imagen destino $i$
$b$	Baseline
$f$	Distancia Focal
$d$	Disparidad = $ x_l - x_r $ , donde l, r indican imagen izquierda y derecha, respectivamente
$\mathbf{a, b, q}$	Angulo de Elevación ( <i>Pitch</i> ), de Convergencia ( <i>Yaw</i> ) y de Rotación ( <i>Roll</i> ) entre cámaras
$d$	Umbral
$e$	Error Relativo al Rango
$m$	Media
$s$	Desviación Estándar
$\lambda$	Longitud de Onda [m]
$C(x, y)$	Correlación del punto (x, y)
$I(x, y)$	Elemento (pixel) de la imagen I en las coordenadas (x, y)
$W(x, y)$	Factor de Confiabilidad de la Disparidad del punto (x, y)
NCC	<i>Normalised Cross Correlation</i> : Correlación Normalizada
SSD	<i>Sum of Squared Differences</i> : Suma de Diferencias al Cuadrado
$\ m\ $	Distancia Euclideana
$\mathbf{M}$	Matrices en Mayúsculas
	Regiones No Válidas en Mapas de Disparidad (Rojo Oscuro)
	Disparidades o Confiabilidades Altas (Grisés Claros)
	Disparidades o Confiabilidades Bajas (Grisés Oscuros)

## **PREFACIO**

### **Objetivos**

El objetivo principal de esta investigación es mejorar la estimación de la posición, en el espacio tridimensional, de los puntos de una escena; empleando técnicas basadas en estereoscopía.

La estimación de la posición debería cumplir con ciertas características que se pueden resumir en:

- Exactitud y Precisión.
- Estimación de la posición tridimensional del mayor número de puntos posible. (Densidad).
- Aplicabilidad de los algoritmos estereoscópicos a distintos tipos de escenas.
- Tolerancia al ruido y desempeño competente que permita mediciones de distancias, volumen ocupado o la reconstrucción de la escena. (i.e. Adaptable a un rango amplio de aplicaciones).
- Velocidad de Procesamiento.

Para lograr este objetivo se propone el procesamiento de secuencias de pares estéreo, el cual permitiría reducir los errores que se generan principalmente en la etapa de búsqueda de correspondencias. La confiabilidad de los resultados obtenidos de la reconstrucción tridimensional o la medición de distancias, se estimaría utilizando un factor de confiabilidad obtenido mediante el análisis de las curvas de correlación.

Objetivos implícitos son: (1) Estudiar las fuentes de error en la estimación de la posición de un punto en el espacio tridimensional y establecer expresiones matemáticas sobre los límites superiores de la resolución del sistema. (2) Mejorar las actuales técnicas y herramientas de reconstrucción tridimensional utilizando cámaras CCD y computadores personales para futuras aplicaciones que

requieran de visión artificial, como por ejemplo, aplicaciones médicas, procesos industriales y robótica.

## **Resumen de Contribuciones**

A continuación se resumen los aportes realizados con este trabajo y se indica la sección donde pueden encontrarse los detalles respectivos:

1. Recopilación y clasificación de bibliografía (Bibliografía, pág. 155).
2. Determinación de expresiones matemáticas para el diseño de un sistema estereoscópico que cumpla una cierta tolerancia. (Sección 3.1, pág. 53)
3. Desarrollo de un algoritmo estereoscópico que permite una reconstrucción tridimensional con mayor precisión al incorporar:
  - A. Estimador de Confiabilidad (Sección 3.3.2, pág. 104 y Anexo C.2 pág. 185)
  - B. Rango de Búsqueda Adaptivo (Sección 3.3.5, pág. 127 y Anexo C.4 pág. 192)
4. Interpolación bilineal sólo en las áreas no correspondidas. (Sección 3.3.3, pág. 121)

Si bien la búsqueda bibliográfica es una parte necesaria en todo trabajo de investigación, en esta tesis se destaca la recopilación bibliográfica como un aporte debido a dos razones. En primer lugar se recopiló una cantidad notable de publicaciones en el tema de la estereoscopía y las técnicas de reconstrucción 3D, las cuales serán de gran utilidad en las futuras investigaciones. Por otro lado, la extensa recopilación de información se ordenó y clasificó según tema y utilidad. Sin lugar a duda, esto será apreciado por los futuros investigadores, quienes accederán más rápidamente a la información precisa.

Con respecto al segundo aporte, se empleo la geometría analítica para determinar los efectos que tienen las desalineaciones de las cámaras en un sistema

binocular paralelo sobre la medición de la distancia. Los resultados del análisis desarrollado por el autor en muchos casos coincidieron con aquellos realizados por otros autores. De este análisis se destacan tres aspectos: se consideraron una diversidad de fuentes de error y su influencia relativa sobre el error en la medición de las distancias. En segundo lugar se destaca el análisis sobre la variación del rango de disparidades en un sistema estereoscópico con desplazamiento; este aspecto puede ser muy importante en sistemas móviles como robots y rovers. Y por último, se destaca el enfoque del análisis del error desarrollado. Este enfoque pone énfasis en los aspectos prácticos del diseño de un sistema binocular que cumpla un cierta tolerancia en la medición. Las expresiones matemáticas derivadas de este análisis permiten establecer los parámetros apropiados del sistema binocular.

El aporte más importante, y que constituye el núcleo de esta tesis (como su título lo manifiesta) son las mejoras al algoritmo de búsqueda de correspondencias al introducir el concepto de *coeficiente* o *factor de confiabilidad*, y el *ajuste del rango de búsqueda de correspondencias*. Estos aportes originales logran mejorar los resultados de la reconstrucción tridimensional (i.e. aumentar la exactitud y precisión) como se demuestra en el capítulo IV. Los detalles de estos aspectos se discuten en las secciones indicadas. Para el lector familiarizado con la estereoscopía e interesado en conocer de manera rápida los detalles de la implementación se recomienda la lectura del Anexo C.

Por último, se destaca la interpolación bilineal de las áreas no correspondidas como un método para reducir aún más los errores en la reconstrucción 3D. Aunque la interpolación de superficies no es un tema novedoso, debemos hacer notar, primero, que sólo las zonas marcadas como *no válidas* son interpoladas, es decir el resto de la superficie no es alterada por la interpolación, la cual tiende a reducir la exactitud de la reconstrucción al efectuarse en forma global. Y en segundo lugar, debemos resaltar la simplicidad del método, el cual requiere de muy poco tiempo de procesamiento.

## **Organización de la Tesis**

En primer lugar se presenta el tema de la estimación de la profundidad y la terminología básica para la comprensión de las secciones posteriores (este punto puede omitirse si se está familiarizado con los conceptos del tema de la estereoscopía). En el primer capítulo también se realiza un análisis comparativo de los distintos métodos para la estimación de la profundidad.

En el segundo capítulo se presentan los fundamentos del proceso estereoscópico en forma detallada.

El trabajo desarrollado se encuentra descrito en el tercer capítulo, el cual se inicia con un análisis cuantitativo de las fuentes de error en el proceso estereoscópico, para luego presentar en forma detallada por etapas, el algoritmo de procesamiento de las secuencias estéreo, su implementación y los errores que pretende reducir.

Los resultados experimentales y su análisis se presentan en el capítulo cuatro.

El quinto y último capítulo discute los alcances del nuevo algoritmo y se proponen nuevas líneas de investigación a partir del trabajo realizado.

## I. INTRODUCCION

La estimación precisa de las coordenadas tridimensionales de un objeto en una escena es fundamental en muchas aplicaciones prácticas que pretendan determinar distancias de colisión, elevaciones o volúmenes mediante el procesamiento de imágenes en forma automática. Existe una gran variedad de aplicaciones, frecuentemente en las áreas de robótica y navegación autónoma [Krot94, Thor92], fotogrametría [Lotti94], automatización industrial, control de calidad, operación remota [Horn86] y de las aplicaciones médicas [Duc197]. La visión estereoscópica, basada en la triangulación entre un punto de la escena y al menos dos proyecciones de este punto sobre imágenes tomadas desde distintas perspectivas, es una de las técnicas más utilizadas para la reconstrucción tridimensional. Inicialmente los métodos de estereoscopía se inspiraron en los sistemas de visión biológicos como el humano [Marr76], por su eficacia y adaptabilidad. Hoy en día, bajo un mismo paradigma geométrico, se agrupan un gran número de enfoques menos biológicos al problema, pero que han cobrado fuerza gracias a los desarrollos tecnológicos como cámaras CCD de menor costo y computadores personales con mayor capacidad de procesamiento. Estos avances permiten la implementación práctica de los métodos estéreo.

Sin embargo, a pesar de que la estereoscopía es una técnica atractiva para la percepción de la profundidad porque reduce la ambigüedad de las mediciones monoculares y a diferencia de los métodos activos (radares y *laser range finders*) es utilizable en un rango mayor de aplicaciones, todavía posee ciertos inconvenientes. Algunos problemas inherentes de la búsqueda de correspondencias binocular incluyen [Coch90, Faug93, Stew89]: (1) la incapacidad de los algoritmos binoculares en la obtención de correspondencias para segmentos de bordes horizontales, (2) la incapacidad de los algoritmos binoculares en la obtención de correspondencias en regiones ocluidas, (3) correspondencias correctas/incorrectas que pueden ser rechazadas/aceptadas principalmente dependiendo del algoritmo de búsqueda de correspondencias (*matching*) utilizado, (4) la ambigüedad causada por patrones periódicos en las imágenes, y (5) la falta de información en escenas con escasa textura. Todas éstas influenciadas fuertemente por (6) la incertidumbre generada en la

discretización de las imágenes y la desalineación de los sensores en un sistema estéreo real.

El uso de múltiples pares tomados desde distintas posiciones [Ayac91], [Kana93] y la combinación de sus mapas de disparidad, reduce los problemas de *matching* de patrones en el sentido del *baseline* (recta que une los centros ópticos del par de cámaras) y los problemas de oclusión. Sin embargo, las texturas inadecuadas de la escena, los métodos de validación no exentos de error y la incertidumbre originada por la discretización, afectarán negativamente la calidad de los resultados. Con el fin de mejorar la resolución de la disparidad calculada, i.e. incrementar la precisión de los resultados mediante la reducción de la incertidumbre, se propone la introducción de un rango adaptivo de búsqueda de disparidades y el filtraje FIR (*Finite Impulse Response*) de secuencias estereoscópicas. Este enfoque sería especialmente útil para sistemas estéreo montados sobre robots en movimiento o vehículos. De una manera heurística, los supuestos en que se basa este método de mejoramiento de la estimación 3D son:

- Las disparidades calculadas a partir de un par de imágenes puede usarse para predecir las disparidades del par siguiente de acuerdo al movimiento del sistema, de manera que la búsqueda de disparidades puede ser restringida a un rango adaptado con la consiguiente reducción del error de *matching* el cual es mayor cuando la búsqueda se realiza en rangos de disparidades muy amplios.
- Los mapas de disparidad contienen información que puede propagarse para validar nuevas correspondencias o incrementar la densidad de los nuevos mapas de disparidad donde el algoritmo ha fallado en la detección de correspondencias válidas.

En este trabajo se establecen las relaciones entre los errores de medición (absolutos y relativos) y los efectos de la resolución espacial del sensor en el cómputo de la disparidad. A partir de estas expresiones matemáticas un rango de disparidades apropiado puede ser elegido dependiendo de la resolución del sensor, la tolerancia del error y el rango de distancias en el cual el sistema operará. Con las expresiones se demuestra que el rango de operación elegido impondrá límites superiores a la resolución del sistema estéreo. El análisis del error realizado permite



establecer las mismas conclusiones que plantean [McVe82] y [Zhao95] en sus trabajos.

Con respecto al segundo supuesto (propagación de la información de la disparidad), resultados experimentales demostraron que es posible adaptar las disparidades calculadas para una escena (mediante correlación normalizada), de acuerdo a los cambios de posición del sistema, y usar los resultados adaptados para mejorar las disparidades calculadas de la siguiente escena. Una vez que los parámetros iniciales del sistema han sido calculados, el máximo desplazamiento aceptable para la propagación de mapas de disparidad y el número de etapas del FIR son actualizadas de acuerdo al movimiento. Se compararon los filtros IIR (*Infinite Impulse Response*) y FIR y se prefirió este último para evitar la propagación de información excesivamente antigua, especialmente cuando el desplazamiento ha sido lo suficientemente grande como para producir un cambio total en la escena observada.

Para poder evaluar las ventajas del método propuesto es necesario establecer ciertas condiciones que aseguren un *ambiente controlado*, libre de perturbaciones. Las simplificaciones al problema y supuestos que se asumirán son:

- Los objetos en la escena son opacos.
- La oclusión es mínima y puede detectarse mediante la comparación de la disparidad calculada en ambos sentidos: primero con la imagen derecha como referencia y realizando la búsqueda en la izquierda, y luego con la imagen izquierda como referencia (R-L vs. L-R) [Fua91, Faug93i].
- La iluminación es constante en posición e intensidad. Las variaciones de la intensidad inducida sobre los objetos cuando las cámaras se desplazan son mínimas y no se consideran.
- El desplazamiento de las cámaras es conocido antes de realizar el análisis estéreo y éste es sólo de traslación.
- Las cámaras se han calibrado y su configuración es tal que, no se requiere una rectificación de las imágenes.

## 1.1 Descripción del Problema y Terminología Básica

### Disparidad ( $d$ )

En la figura 1.1 se muestra que un punto de un objeto en la escena posee un punto imagen distinto dependiendo de la perspectiva de la cámara con respecto al objeto. Dado un par de pixels correspondientes en cada imagen, es posible encontrar la distancia  $Z$  mediante triangulación si se conoce el desplazamiento u offset entre los pixels. Este offset se denomina *disparidad*. El problema central en estereoscopía es encontrar la disparidad (búsqueda de correspondencias o *matching*), la cual, vista de otro modo, expresa cuanto debe correrse el pixel en una imagen para estar alineado con el correspondiente en la otra imagen.

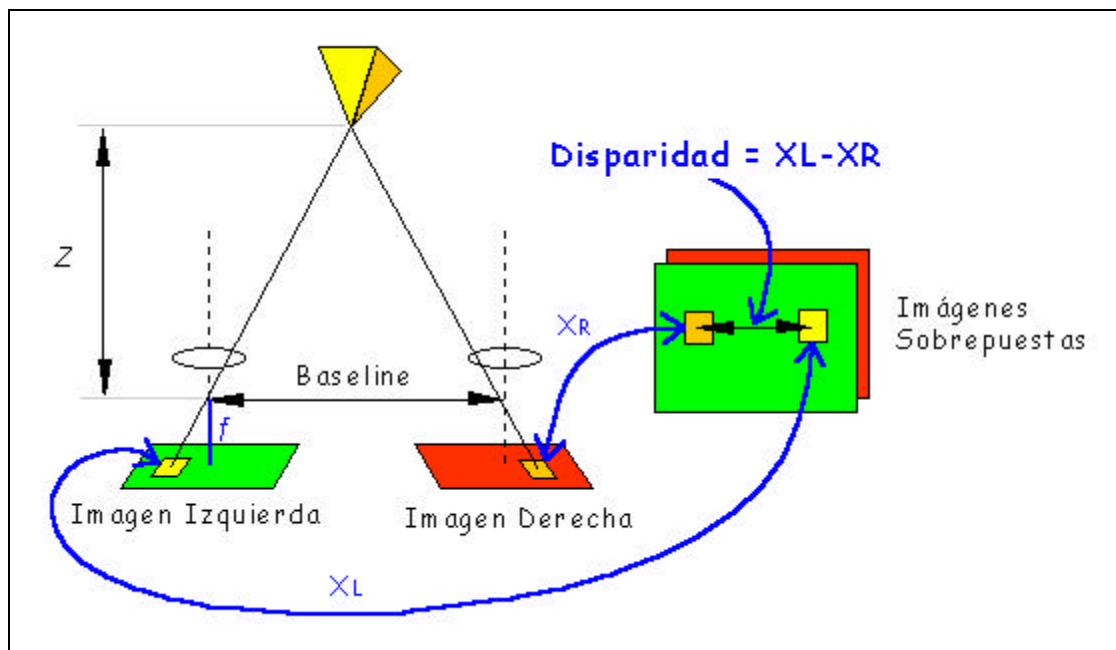


Figura 1.1: Disparidad.

**Baseline ( $b$ )**

El *baseline* es la distancia que existe entre los centros focales de cada cámara (ver figura 1.1). El centro focal (o centro óptico) de una cámara es el lugar geométrico dónde se intersectan todos los rayos correspondientes al haz entre un punto de la escena y su proyección sobre el plano de la imagen (también llamado plano retinal, plano óptico o plano de proyecciones). En el caso de las cámaras utilizadas el CCD corresponde al plano óptico.

**Distancia Focal ( $f$ )**

Corresponde a la distancia entre el *centro focal* y el plano de la imagen (ver figura 1.1).

**Mapa de Disparidad**

Es una imagen cuyos pixels no expresan la luminosidad de la escena sino más bien el grado de disparidad asignado al pixel en el par de imágenes correspondientes. En la práctica puede ocurrir que el rango de disparidades sea pequeño, razón por la cual en el trabajo desarrollado se muestran los resultados de disparidad escalados en intensidad, de modo de facilitar la visualización de los resultados. Disparidades pequeñas indican que los objetos están lejos de las cámaras por lo que se les asignan tonalidades grises oscuras. Los objetos cercanos tienen disparidades mayores y se les asignan en este trabajo tonalidades grises claras. Los puntos de color rojo oscuro corresponden a los pixels para los cuales no se encontró una correspondencia válida. En azul oscuro se muestran los pixels de disparidad menor a un cierto umbral prefijado (sólo utilizados en pruebas experimentales, normalmente no aparecen). Los pixels en verde son aquellos que corresponden a disparidades que sobrepasaron un umbral superior de disparidad prefijado.

Es importante tener presente que no se pueden comparar mapas de disparidades entre sí para distintas tomas. Sólo en los casos que se requiere se mantuvieron los mapas de disparidad con tonalidades absolutas (no re-escaladas al rango 0-255).

## Curva de Correlación

Las curvas de correlación son el resultado de la búsqueda del pixel de la *imagen de referencia* en la *imagen destino*. Los algoritmos implementados toman como imagen de referencia la imagen derecha.

Gráficamente, esta búsqueda de pixels correspondientes se muestra en la figura 1.2. El proceso de búsqueda de la pareja de puntos correspondientes a un punto de la pirámide se puede descomponer en los siguientes pasos:

1. Selección de un punto en la imagen derecha.
2. Comparación del punto de referencia con los puntos en un entorno de la imagen izquierda (línea azul), asignándole un valor a la similitud de acuerdo a un cierto criterio.

En la práctica son dos los criterios que más frecuentemente se utilizan [Faug93, Kana93]:

### Correlación Normalizada

La *correlación normalizada* (*Normalised Cross Correlation, NCC*), corresponde a la correlación clásica de dos variables aleatorias y puede tomar valores entre -1 y +1. Idealmente los puntos muy parecidos tendrán una NCC de +1. La NCC mide similitudes, por lo tanto se debe encontrar el punto de correlación máxima.

### Suma de Diferencias al Cuadrado

La *suma de diferencias al cuadrado* (*Sum of Squared Differences, SSD*), mide las diferencias entre los pixels. En este caso se debe minimizar la función, ya que el mínimo corresponde a la pareja de puntos con menor disimilitud. Idealmente la SSD debería ser cero en el óptimo.

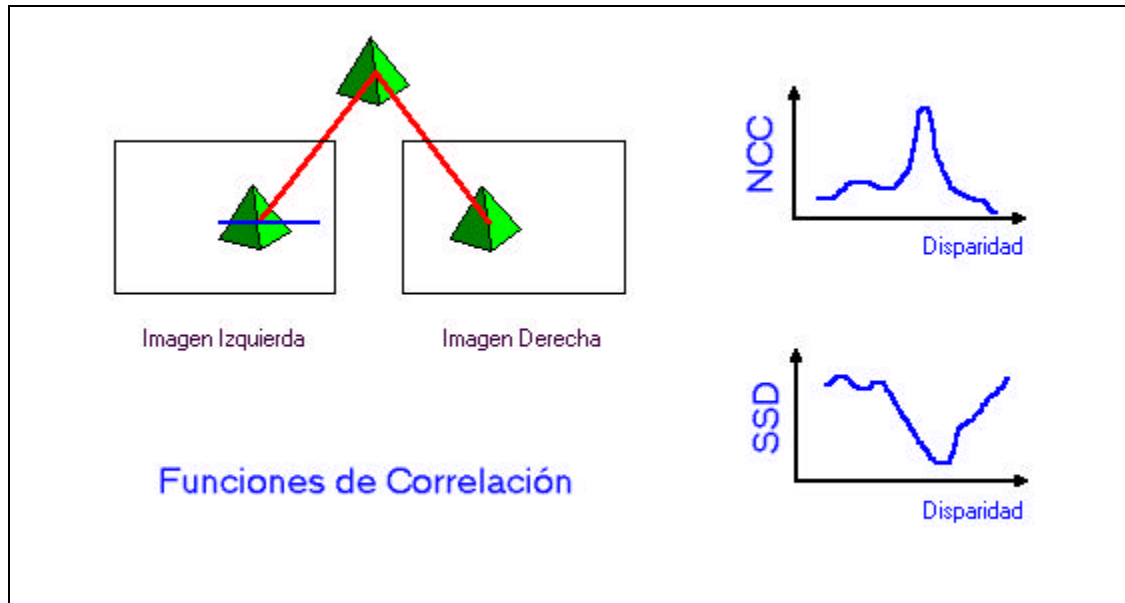


Figura 1.2: Funciones de correlación.

### Ventana o Bloque

Los principales cálculos que se realizan en los algoritmos de correlación tienen por finalidad la comparación de una área pequeña de una imagen con otra similar en la otra imagen. El tamaño del área o conjunto de pixels utilizados en la comparación, se denomina ventana, y puede incluso ser del tamaño de un pixel.

En el diseño de algoritmos basados en comparación de área, son varios los factores que motivan el uso de ventanas pequeñas: (1) Se reduce el costo computacional, (2) y se aumenta la exactitud al reducir el número de pixels afectados por el *desenfoque* que inevitablemente ocurre en las discontinuidades de la profundidad (e.g. bordes y vértices), ya que estas ventanas típicamente se desplazan sobre la imagen entera sin una segmentación inicial. Sin embargo, estos arreglos bidimensionales no pueden ser muy pequeños, ya que tratar de encontrar el pixel correspondiente a otro pixel (y no el promedio de la ventana a un promedio correspondiente similar) está notoriamente sujeto a errores debido a los efectos de la iluminación y cuantización. Además, aunque los pixels correspondientes entre sí

tengan intensidades similares, frecuentemente hay pixels de intensidad parecida en el entorno que introducen más ambigüedades a la etapa de *matching*.

## 1.2 Estereoscopía vs. Otras Técnicas

Existen varios tipos de técnicas de medición de distancias, las cuales se pueden clasificar en dos grandes grupos [Jarv83]:

**Directas y Activas:** Incluyen las basadas en el *tiempo de vuelo* de señales ultrasónicas y electromagnéticas. Requieren de una fuente que controle el rayo de energía y un detector de la energía reflejada.

**Pasivas:** Basadas en imágenes, éstas pueden ser de varios tipos. Las monoculares que realizan algún tipo de análisis como: análisis del gradiente de la textura, análisis fotométrico (las normales de la superficie a partir de la reflectancia), análisis de las oclusiones, análisis de tamaños de objetos, o análisis del enfoque (focus).

Otro tipo de técnicas pasivas incluye el análisis de más de una imagen. En técnicas como las basadas en medición de la disparidad estéreo, del flujo óptico (análisis de la variaciones de intensidad luminosa de la imágenes) y otras relacionadas al movimiento (*shape from motion*), múltiples imágenes son utilizadas para determinar las distancias y posiciones de los objetos en la escena, principalmente mediante triangulación.

En general, las técnicas pasivas tienen un rango de aplicación más amplio debido a que no se requiere una fuente artificial de energía y pueden ser muy apropiadas para aplicaciones tanto interiores como exteriores. Sin embargo, si las escenas son muy complejas se requiere de técnicas más intrusivas como ultrasónicas, radar o rayos laser.

Todas las técnicas tienen alguna desventaja que cae en a lo menos una de las siguientes categorías:

- Problemas de partes de la escena ausentes o no detectadas
- Complejidad computacional
- Costo de tiempo requerido en mejorar la calidad de la razón señal/ruido
- Limitados a aplicaciones en interiores
- Limitados a escenas con una textura determinada
- Limitados a una cierta orientación en las superficies
- Resolución espacial limitada

Los métodos pasivos monoculares son computacionalmente complejos y son muy susceptibles a cambios en las condiciones de la escena, sobre todo los basados en fotometría y los basados en reconstrucción a partir de las variaciones de intensidad luminosa en la escena (*shape from shading*); estos últimos también muy limitados a superficies suaves y continuas. Los métodos monoculares tienen razones señal/ruido más bajas.

Los trabajos basados en estereoscopía han demostrado que es un método confiable, y que dependiendo del algoritmo empleado en la búsqueda de correspondencias, permite obtener estimaciones densas de las distancias a los puntos en la escena. Por esta razón la estereoscopía es uno de las técnicas pasivas más empleadas.

El único método que compite con la estereoscopía en calidad de resultados es la estimación de distancias mediante técnicas activas. Estas se describen brevemente a continuación, para posteriormente mencionar las razones por las cuales se prefirió la estereoscopía.

### 1.2.1 Descripción de Métodos Alternativos

#### A. Sensores Ultrasónicos

Existen dos tipos de sensores basados en ultrasonido [PIIC93]: (1) de resonancia, los cuales producen ondas continuas de FM y (2) los de tipo *pulse-echo*, los cuales operan de manera similar a los sonares. Permiten determinar distancias midiendo el tiempo que requiere la señal en llegar al objeto y volver al transductor como un eco de la señal enviada (tiempo de vuelo).

La mayoría de los sensores industriales poseen compensación térmica de la medición por los efectos que tiene la temperatura sobre la velocidad de la onda ultrasónica. Por definición, cualquier sonido sobre los 20 kHz se denomina ultrasonido. Mientras más elevada sea la frecuencia de operación del sensor, más inmune al ruido del ambiente será, pero se reduce su distancia de “visión”. Los transductores ultrasónicos industriales típicamente operan a 215 kHz.

Su uso en robótica es muy frecuente, especialmente para la detección de obstáculos, y en menor grado para la estimación de distancias. Su difundido uso se debe a su costo reducido, sin embargo se ven muy afectados por las condiciones atmosféricas tales como turbulencias en el aire, humedad, temperatura, tipo de superficie y orientación (puede causar que el eco se refleje en una dirección tal que no sea recibido por el transductor) y por la distancia al blanco (debido a que la amplitud del eco disminuye en forma inversamente proporcional al cuadrado de la distancia). Los factores anteriores hacen que su exactitud se reduzca en forma notable.

El cross-talk (ecos percibidos por un receptor que no corresponde al transmisor asociado) limita el uso de muchos sensores (matrices más densas) por lo que no son buenos para generar estimaciones densas de puntos.

#### B. Laser Range Finders (LRF)

El principio de operación de los radares de longitud de onda óptica también llamados Lidar (*Light Detection And Ranging*) son similares a los de los sensores ultrasónicos, las consideraciones para su diseño se pueden encontrar en [Kelly95]. En este caso la onda corresponde a un haz de luz colimada, normalmente



en la parte infrarroja del espectro ( $\lambda = 750$  nm a 1000 nm), para reducir las interferencias del espectro visible. Estos sensores son más precisos que los ultrasónicos, pero su costo es mayor. Se ven menos afectados por las condiciones ambientales, pero las mediciones en este caso también dependen de la geometría del blanco y su posición con respecto a la fuente de luz, así como de las propiedades reflectivas del material.

Otra ventaja que ofrecen es que tienen un rango de operación mayor (el rango práctico puede alcanzar distancias de 30 m y errores menores a  $\pm 10$  mm) y pueden generar un muestreo denso de puntos mediante el raster-scan de dos espejos que direccionan el rayo en forma discreta. La desventaja de utilizar espejos movidos a través de dispositivos mecánicos es el mayor tiempo requerido para captar una escena.

Algunas de las aplicaciones de esta tecnología de medición son:

- Reconocimiento de la posición de objetos en movimiento.
- Monitoreo de llenado, monitoreo de volumen.
- Medición de objetos estáticos independientemente de su posición, forma o color, para reconstrucción tridimensional.
- Reconocimiento de posición y navegación.
- Determinación de la posición de containers y verificación de empaques o embalajes (ver figura 1.3).

Información sobre LRFs y sus aplicaciones se puede encontrar en: [DFR96, Kelly94b, TK92].

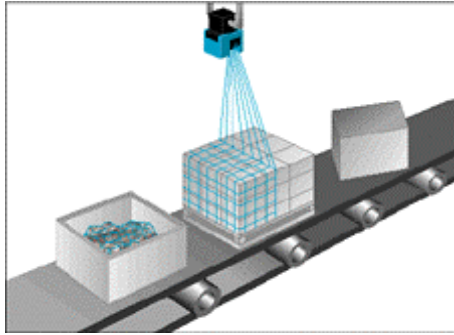


Figura 1.3: Laser Measurement System LMS 200 Sick AG [web1].

### C. Radares de Apertura Sintética (SAR)

Dentro de las distintas categorías de radares, se destacan en la reconstrucción tridimensional de superficies aquellos basados en técnicas de apertura sintética.

Los radares (*Radio Detection and Ranging*) miden la energía y el tiempo de vuelo de las señales de microondas emitidas por la antena del radar y que se reflejan en una superficie distante u objeto. La antena transmite y recibe en forma alternada pulsos a una longitud de microonda particular (en el rango de  $\lambda = 1 \text{ cm}$  a  $1 \text{ m}$ , que corresponden a frecuencias entre los 300 MHz y los 30 GHz) y con una polarización dada (ondas polarizadas en un solo plano vertical u horizontal). Para un sistema radar de imágenes (*imaging radar system*), aproximadamente 1500 pulsos por segundo de alto poder son emitidos al blanco o área de la imagen. Cada pulso con una duración de 10 a 50  $\mu\text{s}$  (ancho del pulso). El pulso normalmente cubre una pequeña banda de frecuencias, centradas en la frecuencia seleccionada para la operación del radar. Los anchos de banda típicos para radares de imágenes se encuentran entre 10 a 200 MHz. La energía en el pulso del radar es dispersada en todas direcciones al incidir sobre el objeto. Parte de la energía dispersada vuelve a la antena como un eco débil (*backscatter*) con una polarización (horizontal o vertical) no necesariamente igual a la del pulso transmitido. Los ecos son convertidos a datos digitales, los cuales son registrados para su posterior procesamiento y despliegue como una imagen. En la figura 1.4, se muestra una aplicación de SAR para topografía.

Dado que el pulso viaja a la velocidad de la luz, la medición de distancias a los objetos es bastante directa si se utiliza el tiempo que tardó el eco en llegar a la antena. El ancho de banda del pulso determina la resolución del rango de distancias medidas.

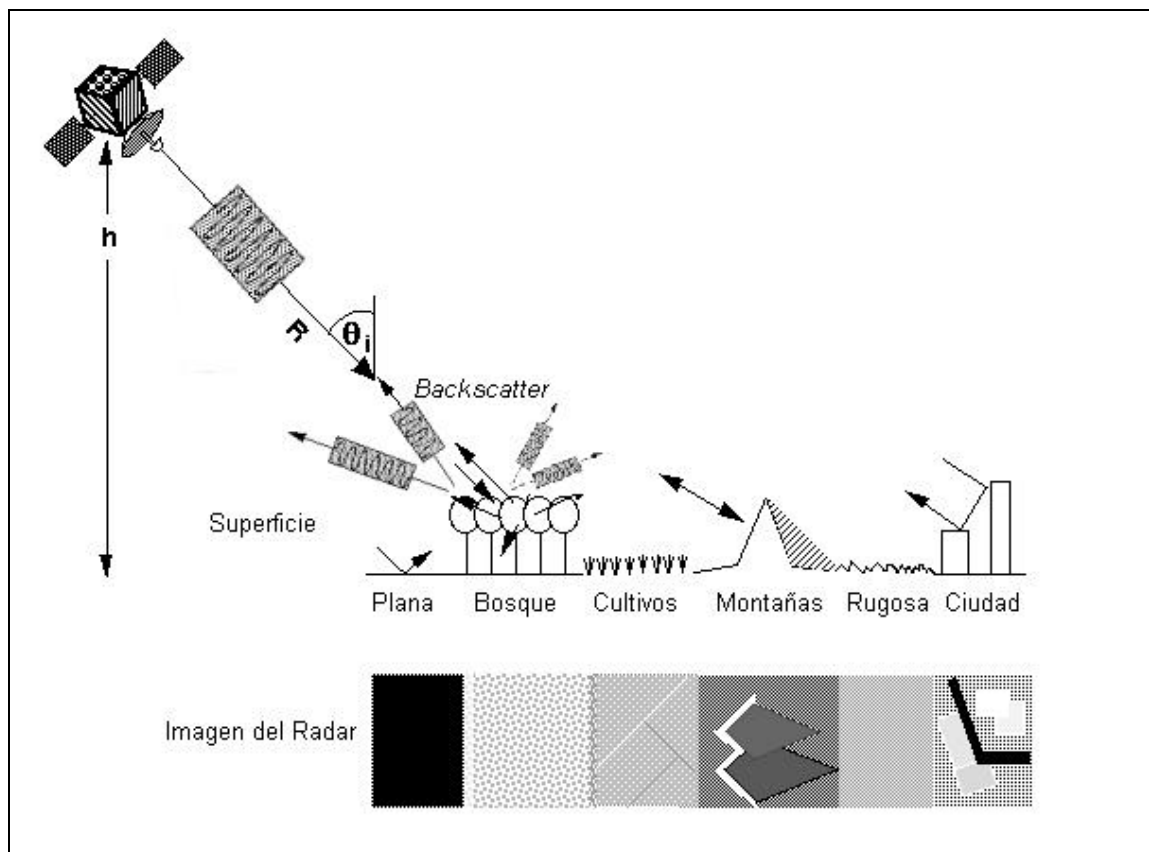


Figura 1.4: Pulsos transmitidos por el radar y ecos reflejados (*backscatter*).

En el caso de los radares de imágenes, el radar debe barrer un área para poder formar la imagen. Los SAR sobre satélites o naves realizan el barrido aprovechando su desplazamiento en torno a la Tierra (ver figura 1.5).

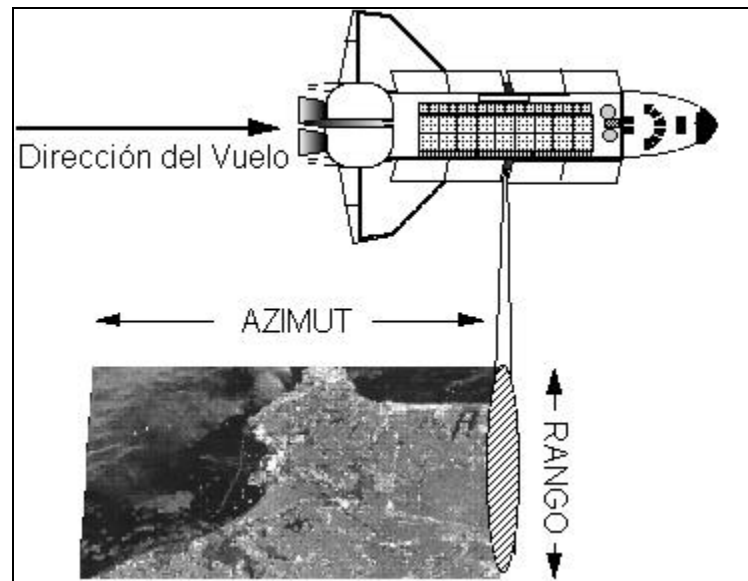


Figura 1.5: Barrido de la antena.

La longitud de la antena del radar determina la resolución en el sentido azimutal (ver figura 1.5). Mientras más larga es la antena, mayor es la resolución.

Los Radares de Apertura Sintética (SAR: *Synthetic Aperture Radar*) se refieren a una técnica para sintetizar una antena muy larga combinando los ecos recibidos a medida que se desplaza a lo largo de la dirección de vuelo o de barrido. La apertura se refiere a la distancia de barrido utilizada para recolectar la energía reflejada, que luego es usada para formar la imagen. Una apertura sintética se genera desplazando una apertura real o antena a lo largo de una serie de posiciones distintas en la dirección de barrido.

A medida que el radar barre un área, transmitiendo un pulso en cada posición, el eco de los pulsos es recibido por un receptor y luego almacenado. Como el radar se mueve con respecto al objeto (la Tierra, en el caso del satélite), los ecos están desplazados por el efecto Doppler. Comparando los desplazamientos de las frecuencias con una frecuencia de referencia, las señales recibidas pueden *enfocarse* sobre un área para aumentar la longitud efectiva de la antena que está creando una imagen de dicha área.

La operación de enfoque es comúnmente conocida como procesamiento SAR, y es realizada digitalmente en sistemas computacionales de alta velocidad de procesamiento. El aspecto clave del procesamiento SAR es encontrar en forma correcta la variación de la frecuencia debido al efecto Doppler, para cada punto de la imagen. Esto requiere de un conocimiento preciso del desplazamiento entre la plataforma del radar y el objeto.

Las técnicas empleadas en los radares de apertura sintética han logrado la madurez suficiente como para registrar imágenes con una resolución muy fina. Y tienen características que los hacen muy ventajosos frente a otras técnicas. Por ejemplo, no dependen de la iluminación externa, ya que proveen su propia iluminación de la escena a través de los pulsos del radar. También tienen mayores capacidades de visión frente a sistemas infrarrojos o de luz visible. Los sistemas SAR pueden *ver* a través de las nubes o el polvo, ya que utilizan longitudes de onda superiores a las de los otros sistemas.

Las imágenes de radar indican el *backscatter*. Si la dispersión de la onda es grande entonces la imagen muestra valores de intensidad luminosa bajos. Regiones de mayor intensidad luminosa indican que una mayor cantidad de energía fue reflejada en forma de eco. Las condiciones que afectan el eco o *backscatter* son: tamaño de los objetos en el área de la imagen, propiedades reflectivas de los materiales, contenido de humedad en el área, polarización de los pulsos, distancia al área barrida ( $h$ ), el ángulo de incidencia o de observación  $\theta_i$  y la frecuencia de la onda emitida.

Entre las principales aplicaciones de los SAR se encuentran: reconocimiento, monitoreo ambiental, verificación de tratados, navegación, detección de cambios, oceanografía, estudios de erosión, pesca, agricultura. Información adicional sobre la tecnología SAR se puede encontrar en [web8, web9, web10].

### **1.2.2 Comparación: Estereoscopía vs. Otras Técnicas**

En primer lugar debe definirse claramente el objetivo del sistema a implementar y los rendimientos o características que se esperan de éste. Una vez

realizado este paso, la elección del método, el enfoque y las tecnologías apropiadas que permitan lograr el objetivo será más eficaz.

Teniendo presente como objetivo el futuro desarrollo de sensores de visión artificial para aplicaciones de reconstrucción 3D (industriales y médicas), la estereoscopía es la técnica que ofrece mayores beneficios por las razones que se encuentran resumidas en la tabla 1.1 y que se discuten a continuación.

- a) **Amplio Rango de Operación:** Mediante ajustes de los parámetros del sistema como la elección del focus y de la longitud del *baseline*, es posible modificar el rango de operación en forma mucho más sencilla. Mientras que en los sistemas LRF, la calibración del sensor se hace una sola vez para un rango fijo, o de modificación bastante más compleja.
- b) **Existencia de Hardware que permite la implementación práctica de la Estereoscopía:** Con el avance de los sistemas digitales, hoy en día es posible contar con circuitos de mayor velocidad, reducido tamaño y gran capacidad de almacenamiento de información, lo cual hace factible la utilización de técnicas pasivas como la visión estéreo, que a pesar de las mejoras de los algoritmos, es computacionalmente intensiva. En la actualidad se han realizado implementaciones en hardware de algunos algoritmos para la detección estéreo, principalmente basados en NCC y SSD [Faug93i, Kana94], utilizan DSPs (*Digital Signal Processors*) y logran hasta 30 fps; incluso se ha llegado a desarrollar algunos ASICs (*Application Specific Integrated Circuits*) experimentales [Erte96]. El tiempo de procesamiento y la capacidad de memoria requerida ya no son una justificación para que el método sea prohibitivo en términos prácticos.
- c) **Inmunidad a las Interferencias:** Si bien los sonares son bastante utilizados en la evasión de obstáculos por su bajo costo, su aplicación se descarta porque su exactitud no es lo suficientemente alta como para ser aplicados en reconstrucción 3D ya que son muy susceptibles a interferencias y perturbaciones. En términos de *capacidad de visión* los sistemas basados en SAR son superiores a todos los demás, pero su aplicación práctica sólo ha alcanzado logros en el área de sensado remoto.

- d) **Cantidad de Información por Muestreo de la Escena:** Los LRF se descartan porque requieren de varios barridos con el haz. Aproximadamente entregan 4 veces menos información por barrido que una búsqueda de correspondencias mediante la correlación de imágenes de 256x256 [web2, web3].
- e) **Costo Reducido:** Actualmente un sistema estéreo puede tener costos mucho menores ( $\approx$ US\$1.000 [Hors93]) que un sistema LRF ( $\approx$ US\$15.000 [web4]). En la tabla 1.2 se comparan los costos de un sistema LRF vs. estéreo en el mercado.

Tabla 1.1: Ventajas y Desventajas de los Distintos Métodos para Estimar Distancias y Reconstrucción Tridimensional.

Método	Ventajas	Desventajas
Sonar	<ul style="list-style-type: none"> <li>- Económicos (US\$ 1.200).</li> </ul>	<ul style="list-style-type: none"> <li>- Baja resolución espacial.</li> <li>- Muy sensibles a las condiciones atmosféricas.</li> <li>- Dependientes de la geometría de los objetos.</li> <li>- Para mejorar la resolución espacial se requiere de muchos sensores y de repetidas mediciones, pero empiezan a aparecer problemas de <i>cross-talks</i>.</li> </ul>
Laser Range Finders	<ul style="list-style-type: none"> <li>- Alta exactitud y precisión.</li> <li>- Sujetos a menos perturbaciones.</li> <li>- Alta densidad de puntos estimados.</li> </ul>	<ul style="list-style-type: none"> <li>- Costo más elevado (US\$15.000).</li> <li>- Dependientes de la geometría de los objetos.</li> <li>- Rango de operación de complicada modificación.</li> <li>- Menor diversidad de posibles usos.</li> </ul>
Radares de Apertura Sintética	<ul style="list-style-type: none"> <li>- Alta exactitud y precisión.</li> <li>- Mayor capacidad de medición en condiciones de poca visibilidad para los demás sistemas (nubes, polvo y oscuridad).</li> <li>- Alta densidad de puntos estimados.</li> </ul>	<ul style="list-style-type: none"> <li>- Requieren de una estimación precisa del desplazamiento del radar con respecto al blanco.</li> <li>- Dependen del ángulo de incidencia.</li> <li>- Aunque muy utilizados en aplicaciones de sensado remoto, en otras áreas siguen en fase experimental.</li> </ul>
Estereoscopía usando Correlación	<ul style="list-style-type: none"> <li>- Sistemas de bajo costo [web7].</li> <li>- Alta densidad de puntos estimados.</li> <li>- Amplio rango de operación.</li> <li>- Mayor diversidad de posibles de usos.</li> </ul>	<ul style="list-style-type: none"> <li>- Menor exactitud.</li> <li>- Problemas con ciertas texturas.</li> </ul>

Tabla 1.2: Comparación de Costos: LRF vs. Sistema de Visión Estéreo.

LRF		Estéreo	
Componentes	Costo [US\$]	Componentes	Costo [US\$]
Sistema de Medición Laser: LightRanger HelpMate [web4]	15.000	Tarjeta Digitalizadora con DSPTMS320C40: DIPIX XPG-1000 [web5]	XPG-CORE-P 5.680 XPG-MOD-ACM-P 1.190 2xXPG-DMEM-4MB 1.480 XPG-DOC-KIT 450 Xtal 80
		Total Tarjeta	8.880
		2 Cámaras: CCD RGB	1.500
		Stereo Rig: Zebra Stereo Vergence Head	2.000
<b>Total</b>	<b>&gt;15.000</b>	<b>Total</b>	<b>&lt;12.380</b>

### 1.3 Trabajo Relacionado

La reconstrucción de imágenes 3D mediante el procesamiento de secuencias es un tópico bastante reciente. La mayoría de la investigación en visión artificial hasta la fecha ha concentrado sus esfuerzos en la recuperación de la información 3D a partir de varias fuentes [Klet95] como estereoscopia estática, a partir del movimiento (*shape from motion*), o a partir de las sombras sobre los objetos en la escena (*shape from shading*).

El mejorar los algoritmos para estereoscopia ha sido un aspecto de continuo trabajo e investigación en visión artificial. Varias investigaciones han tenido por objetivo mejorar la velocidad de procesamiento paralelizando los algoritmos [Faug93i, Kana94i, Kosc95]. Aunque los algoritmos desarrollados durante esta investigación podrían utilizar hardware especial, estos son particularmente apropiados para su implementación en computadores secuenciales, de bajo costo y fácil programación.

El enfoque utilizado en este trabajo está muy relacionado al algoritmo de correlación normalizada acelerada [Faug93i, Fua91]. A éste se le incorpora el



concepto de reducción del rango de búsqueda de disparidades mediante la combinación de dos técnicas:

- a) **Procesamiento Piramidal, Jerárquico o *coarse-to-fine*:** Consiste en encontrar disparidades iniciales para una resolución baja de la imagen y utilizar dichos resultados como referencia para la búsqueda de disparidades a resoluciones superiores [Falk97i]. Esto es equivalente a utilizar ventanas de correlación inicialmente de tamaños grandes y luego más reducidos. En ambos métodos se procesan distintas bandas del espectro de frecuencias de la imagen [Fua91], pero al trabajar con imágenes de menor resolución el procesamiento es más rápido.
- b) **Procesamiento de Secuencias de Pares Estéreo:** Al utilizar una secuencia de pares estereo tomados desde distintas posiciones, las disparidades calculadas para un par de imágenes pueden utilizarse para predecir las disparidades (o equivalentemente las distancias) del par siguiente de acuerdo al movimiento que realizó el sistema binocular en el espacio. De este modo, puede restringirse el rango de búsqueda de disparidades para el par siguiente a un rango adaptado de acuerdo al movimiento y a la información sobre las disparidades encontradas previamente. Al reducir el rango de búsqueda deberían reducirse los errores producto de las ambigüedades que ocurren en intervalos de búsqueda muy amplios. Además debería reducirse el tiempo de procesamiento [Tuca96].

Si bien los trabajos de [Yi97, Roy93, Koch95, Koch94, Koch94i] son similares en el sentido de que también introducen el uso de secuencias de pares estereo, éstos se diferencian del presente desarrollo en que Jae-Woong Yi utiliza como método de búsqueda de correspondencias la estimación del flujo óptico, la cual no es suficiente para generar mapas de la escena densos como se pretende en éste caso. Sebastián Roy [Roy93] utiliza un algoritmo basado en la minimización de una función de costo para la disparidad basado en [Cox94i], el cual permite resultados más densos, pero depende de la comparación directa de pixels y no de áreas. Por esta razón es muy susceptible al ruido de cuantización de las imágenes a pesar de las normalizaciones que emplea. El trabajo de Reinhard Koch *et al* es más similar, pero su interés está en la reconstrucción 3D para aplicaciones de visualización en arquitectura y en la reconstrucción 3D de personas para estudios de TV virtual y presencia virtual en *video-comunicaciones* [Koch96i].

La principal diferencia entre estos trabajos y los métodos que aquí se proponen está en el objetivo que persiguen. Los algoritmos desarrollados durante esta investigación pretenden: (1) Disminuir el error en la medición. (2) Evitar realizar interpolaciones que tiendan a reducir la precisión de los resultados.

Sólo hay un aspecto que no se ha considerado para la disminución del error. Este aspecto es el uso de múltiples *baselines*, que como se demuestra en [Kana93], permitiría reducir las correspondencias ambiguas. Al integrar los resultados de las etapas anteriores para distintas posiciones de las cámaras mediante el procesamiento secuencial de pares estéreoscópicos, el uso de *baselines* de distintas longitudes no sería realmente necesario, salvo el caso de *baselines* perpendiculares para compensar la incapacidad que tienen los algoritmos de correlación para detectar correspondencias en segmentos o patrones paralelos al *baseline*.

## II. ESTEREOSCOPIA

La estereoscopía es una técnica utilizada para la recuperación de las coordenadas tridimensionales de un número de puntos de una escena, a partir de al menos dos imágenes bidimensionales tomadas desde distintas perspectivas.

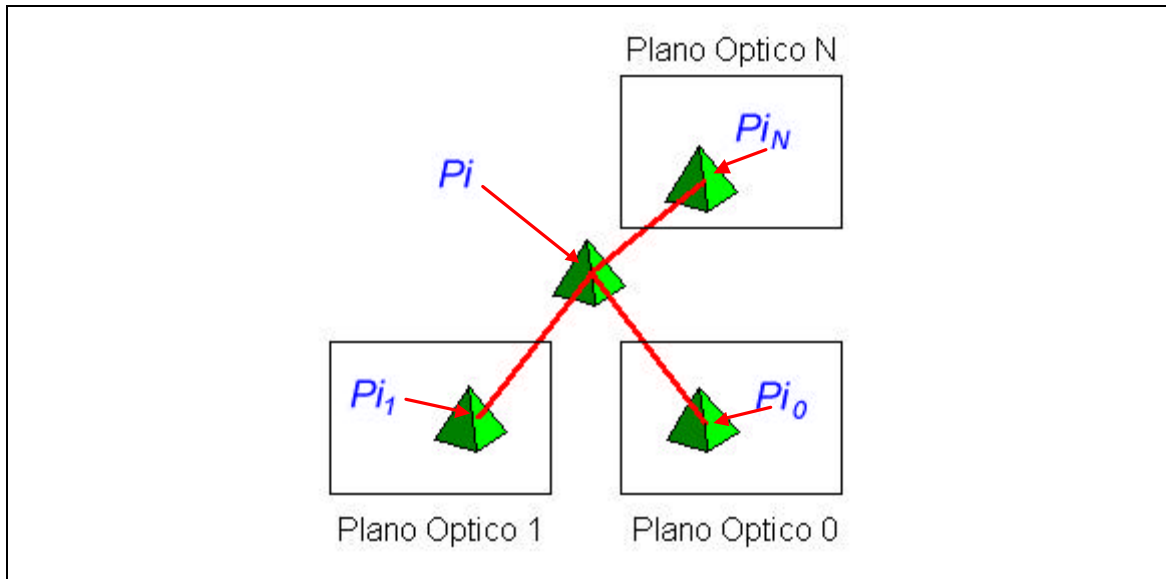


Figura 2.1: Imagen del punto  $P_i$  en la vista de cada cámara.

Dadas  $N$  imágenes (figura 2.1), el proceso se centra principalmente en la solución de dos problemas [Faug93]:

1. Encontrar conjuntos de puntos  $P_{ik}$  correspondientes a la proyección del punto  $P_i$  de la escena sobre el plano óptico de la cámara  $k$ . La búsqueda del punto  $P_i$  en cada una de las imágenes se conoce como el *problema de búsqueda de correspondencias* o *problema de matching*.
2. Conocidos los conjuntos  $P_{ik}$ , calcular las coordenadas tridimensionales de los puntos  $P_i$  de la escena. Este es el *problema de reconstrucción*.

Para la solución de estos problemas, el proceso estereoscópico involucra las siguientes etapas [Barn82]:

1. Adquisición de imágenes
2. Modelación de cámaras
3. Elección de características
4. Búsqueda de correspondencias
5. Determinación de la profundidad
6. Interpolación

A continuación se explican las etapas del proceso estereoscópico. El artículo de [Barn82] y más recientemente [Dhon89, Lane96] presentan revisiones detalladas sobre las distintas técnicas desarrolladas para resolver cada una de estas etapas; especialmente la cuarta de ellas.

## **2.1 Adquisición de Imágenes**

Las imágenes estereoscópicas se pueden adquirir de diversas formas. El número de cámaras, su resolución y su posición, son parámetros a considerar en el diseño de un sistema estereoscópico. Los trabajos de [Alve89] y [Chio95] analizan la geometría óptima para la posición de las cámaras en sistemas binoculares y trinoculares, respectivamente. Análisis sobre el uso de más de un par de cámaras se puede encontrar en [Kana93], donde se justifica el uso de múltiples *baselines* por su contribución a la reducción de correspondencias ambiguas. Otros estudios sobre sistemas trinoculares y sus ventajas, como confiabilidad y precisión, se presentan en [Ayac91, Blak93].

A continuación se estudiará el caso binocular, el cual corresponde a la configuración empleada en esta investigación; sin embargo, este análisis puede ser extendido a sistemas de más de dos cámaras. Con respecto a la resolución que deben poseer las cámaras, el análisis de los efectos de este parámetro sobre el error, que se presenta en el capítulo 3, permite definir su valor apropiado.

## 2.2 Modelación de las Cámaras

Un *modelo de las cámaras* es una representación de las características geométricas y físicas del sistema de adquisición estéreo. Este modelo tiene un componente relativo, que relaciona las coordenadas de ambas cámaras y que es independiente de la escena, y un componente absoluto que relaciona el sistema de coordenadas de una cámara con el sistema de coordenadas de la escena.

El proceso de obtención de estos componentes se denomina *calibración*. En la literatura se cita con frecuencia el método de calibración de Tsai [Tsai87]. Artículos más recientes [Luo93, Viev94, Zelle96] consideran métodos para la *autocalibración* del sistema.

La etapa de calibración es crucial en el proceso de visión estéreo ya que permite simplificar el *problema de correspondencias* y obtener una representación tridimensional precisa de los resultados.

### 2.2.1 Coordenadas Tridimensionales

En la Figura 2.2 se observa la vista de planta de una escena tridimensional creada con una aplicación de CAD que servirá de ejemplo para la presentación del sistema de coordenadas a utilizar. Las esferas  $S_1$  y  $S_2$  de radio 5 están centradas en  $(-10, -5, 60)$  y  $(0, -5, 90)$ , respectivamente. El cubo  $C_1$  de lados 10 está centrado en  $(15, -5, 45)$ . La malla está situada en  $y = 0$ . El centro óptico de la cámara está situado en  $(0, -5, 5)$ . La cámara posee un lente con distancia focal de 20 unidades. La imagen percibida por la cámara se ilustra en la Figura 2.3.

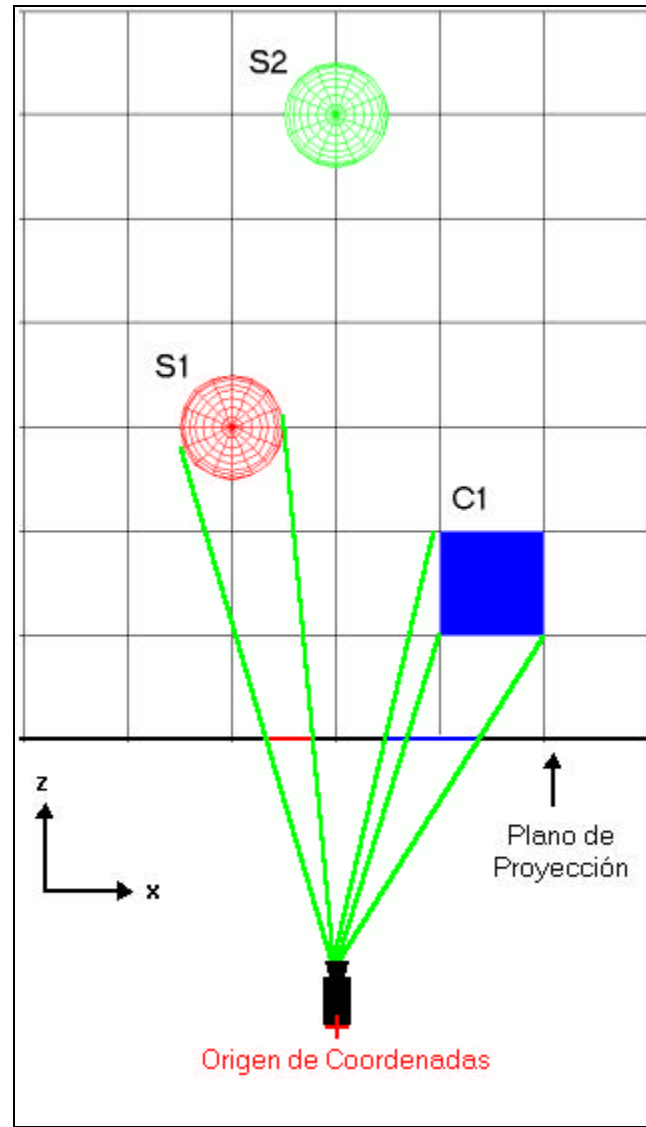


Figura 2.2: Vista de planta de una escena tridimensional.

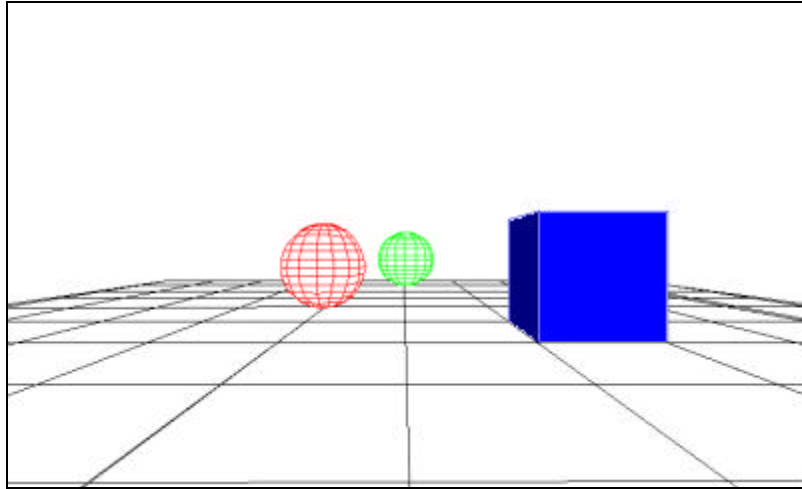


Figura 2.3: Imagen de la escena de la figura 2.2 vista desde la cámara.

Un punto tridimensional  $P_g$  está representado por el vector columna  $P_g = (x_g, y_g, z_g)^T$ . Para simplificar los cálculos, todos los puntos tridimensionales serán representados en coordenadas homogéneas, es decir  $P_g = (x_g, y_g, z_g, 1)^T$ . Una cámara se encuentra en el mismo sistema de referencia (que llamaremos *global*) que  $P_g$ . Si definimos  $P_l$  como las coordenadas de un punto  $P_g$  representado en el sistema de referencia de la cámara (que llamaremos *local*), cuyo origen se encuentra en el centro óptico de la misma) se cumple la relación:

$$P_l = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = T_{gl} \cdot P_g \quad (2.1)$$

donde  $T_{gl}$  es la matriz de transformación de coordenadas globales a coordenadas locales (ver Figura 2.4). Esta matriz puede estar compuesta de una rotación ( $R_{3 \times 3}$ ), una traslación ( $T_{3 \times 1}$ ), un cambio de escala y posiblemente una corrección de distorsiones introducidas por el lente de la cámara. Debido al modelo matricial utilizado, se asume que las distorsiones son lineales. Por lo tanto no es

posible representar con  $T_{gl}$  una deformación radial causada por el lente de la cámara. La matriz de transformación es invertible para permitir recuperar  $P_g$  a partir de  $P_l$ . Para el ejemplo de la Figura 2.2, la matriz de transformación corresponde a:

$$T_{gl} = \begin{bmatrix} R_{3 \times 3} & t_{3 \times 1} \\ 0_{1 \times 3} & w_{1 \times 1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.2)$$

donde el origen del sistema local está situado en las coordenadas  $(0, -5, 5)$  del sistema global.

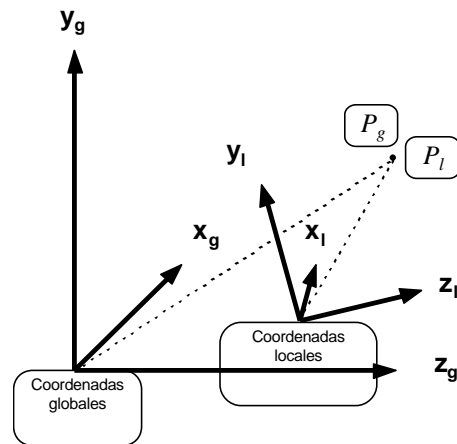


Figura 2.4: Sistema de coordenadas tridimensionales.

### 2.2.2 Proyección en Perspectiva

Como se muestra en la Figura 2.5, la proyección en perspectiva de un punto  $P_l$  sobre el plano de proyección de la cámara resulta en un punto  $P'_l = (x', y', f, 1)^T$  donde  $(x', y')$  son las coordenadas de la imagen en pixels y  $f$  la distancia focal de la cámara. Por lo tanto, la profundidad original  $z$  del punto tridimensional se pierde en la proyección, ya que todos los  $z'$  son iguales a  $f$ .



Se tiene la relación de proyección:

$$x' = \frac{x \cdot f}{r_x \cdot z}, \quad y' = \frac{y \cdot f}{r_y \cdot z} \quad (2.3)$$

donde  $r_x$  y  $r_y$  son factores de conversión de unidades en pixels a unidades métricas, que se pueden expresar en [mm/pixel], por ejemplo. Los factores  $r_x$  y  $r_y$  corresponden a la longitud horizontal y vertical de una celda del CCD, respectivamente. A veces  $r_x$  y  $r_y$  son llamados distancias interpixel, si se consideran como las distancias horizontal y vertical entre los centros de las celdas del CCD.

Hay que hacer notar que en una cámara real, el plano de proyección se sitúa detrás del centro óptico ( $f < 0$ ) y que las imágenes se invierten. Para fines de presentar el sistema de coordenadas simplificado, y sin pérdida de generalidad, se supondrá lo contrario [Coch90, Roy92, Maim96]. Se debe notar que las expresiones en (2.3) consideran una conversión de unidades (mm/pixel) dado que se supone que se utilizan cámaras CCD y  $x'$  e  $y'$  se expresan en pixels. En el caso de que  $x'$  e  $y'$  se expresasen en las mismas unidades que  $f$ ,  $x$ ,  $y$  y  $z$  entonces los factores  $r_x$  y  $r_y$  son adimensionales y de valor 1.

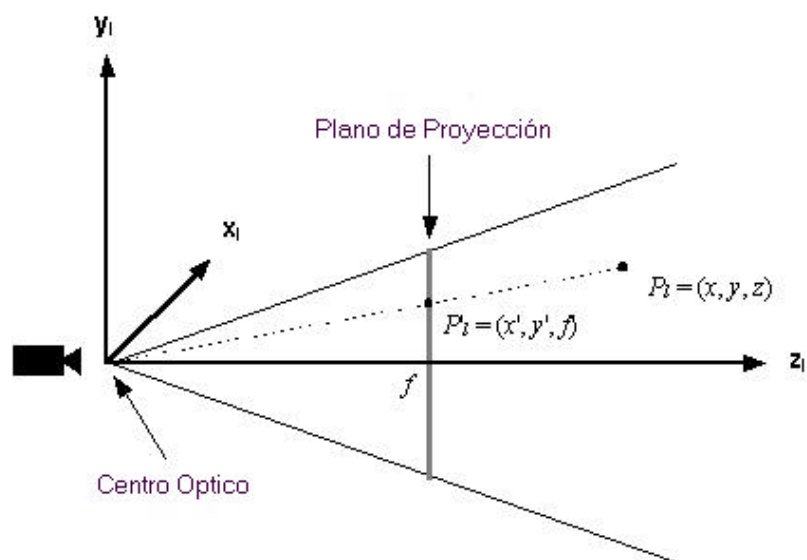


Figura 2.5: Proyección en perspectiva.

### 2.2.3 Modelación Estereoscópica

El *modelo estereoscópico* es el conjunto de expresiones que relacionan puntos en la escena, expresados en coordenadas globales, con los puntos proyectados expresados en las coordenadas locales de cada cámara para una cierta disposición geométrica de los sensores ópticos.

Al observar una escena desde dos puntos de vista diferentes, se obtienen imágenes en las cuales los objetos se ven menos o más desplazados según su profundidad y su posición en la escena. La Figura 2.6 es un ejemplo de un par de imágenes estereoscópicas tomadas con cámaras paralelas desplazadas lateralmente a lo largo del eje  $x$  en 50 mm y cuyos lentes tienen una distancia focal de 20 mm. La gran mayoría de los sistemas estereoscópicos utiliza la *configuración paralela*. El sistema visual humano tiene una configuración similar, pero además tiene la capacidad de ajustar el *ángulo de convergencia* (ángulo entre los ejes ópticos), que en el caso paralelo es fijo e igual a  $0^\circ$ . El ojo humano puede ajustar su distancia focal típicamente desde 14 mm hasta 17 mm [Gonz93], siendo el desplazamiento entre los ojos, conocido como *distancia intraocular*, de 65 mm [Wood93] típicamente.

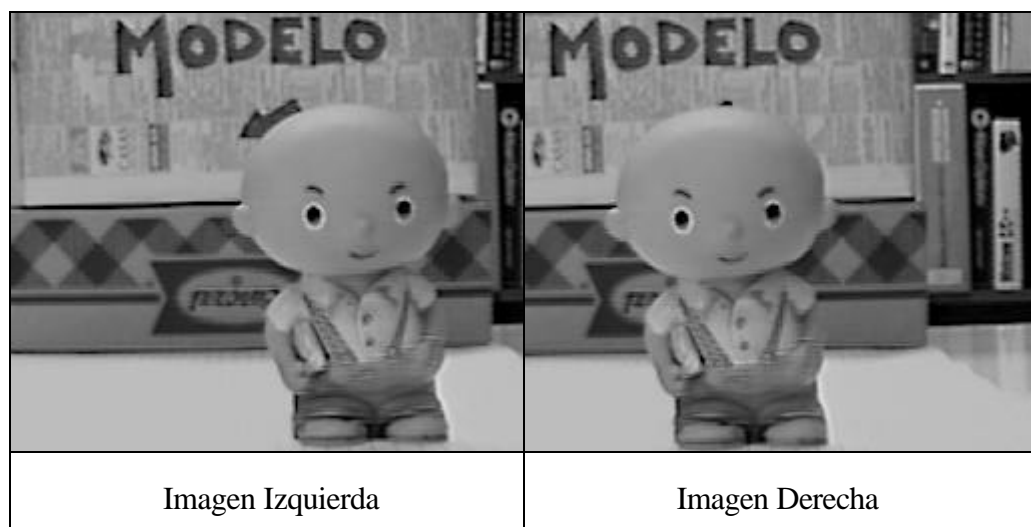


Figura 2.6: Ejemplo de par estéreo con desplazamiento lateral de la cámara.

Utilizando para la escena artificial la misma disposición paralela de cámaras, como se representa en la Figura 2.7, se obtienen las vistas de las cámaras A y B mostradas en la Figura 2.8. Los desplazamientos aparentes de los objetos que se pueden observar en las vistas de la Figura 2.8 permiten recuperar la información de profundidad.

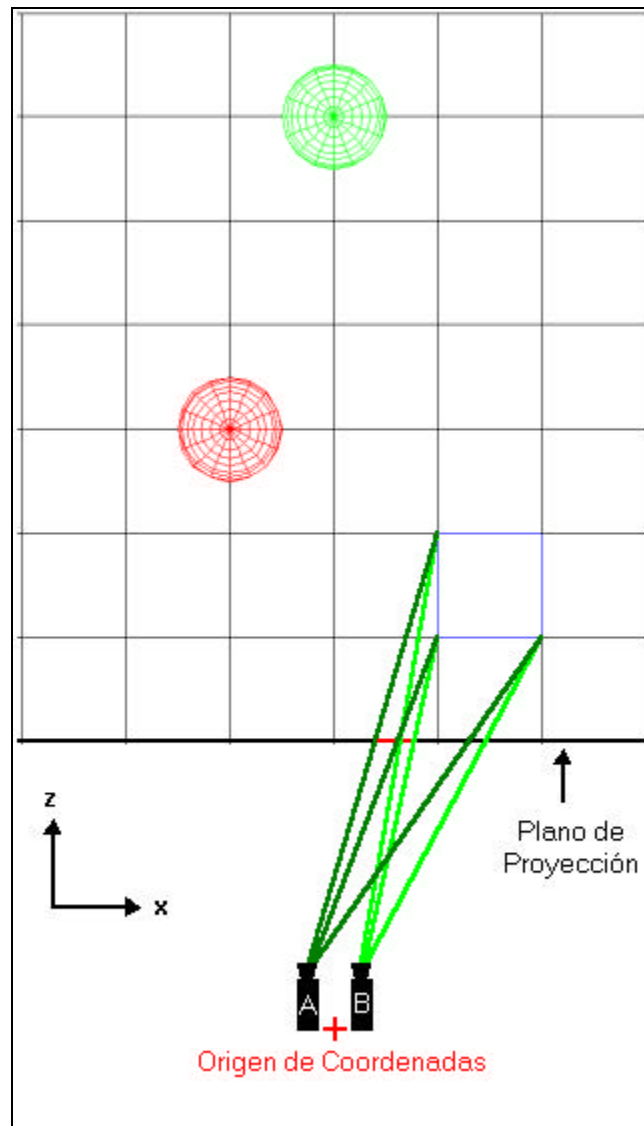


Figura 2.7: Vista de planta de un sistema estéreo con desplazamiento lateral de las cámaras a lo largo del eje  $x$ .

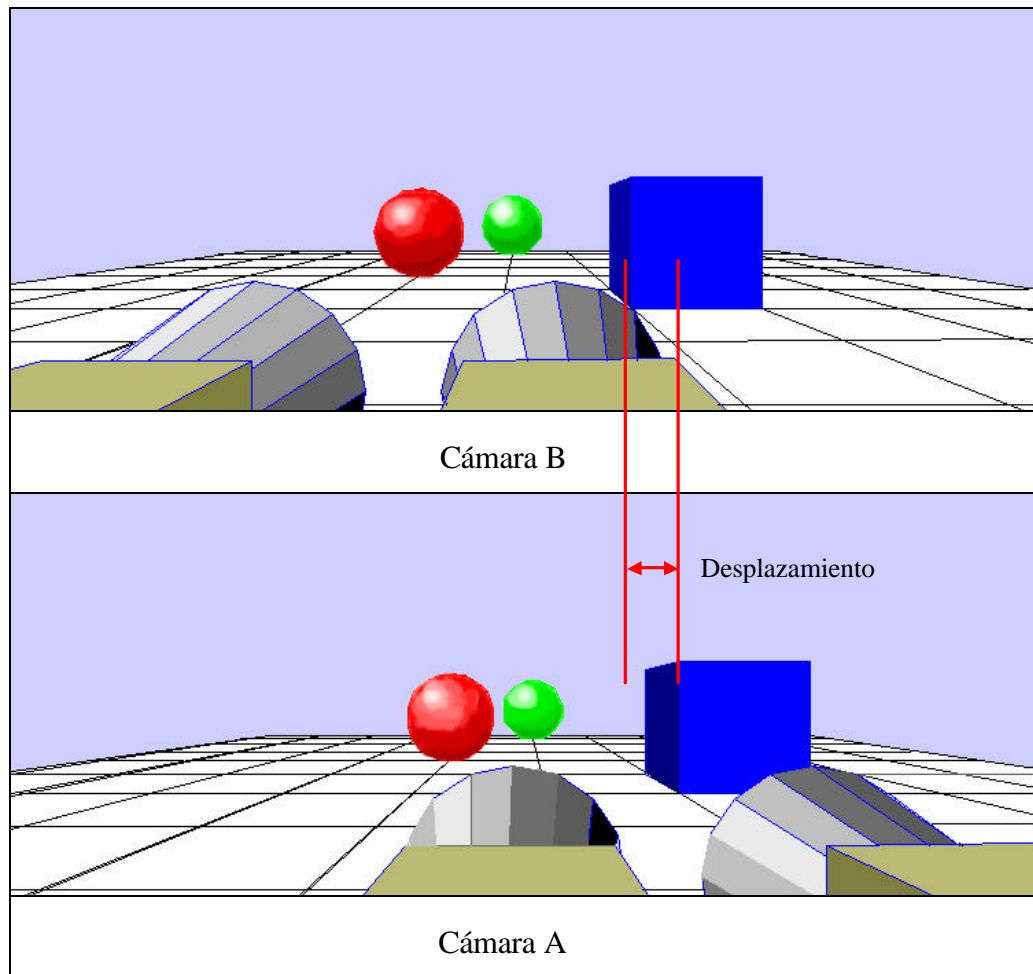


Figura 2.8: Imágenes estereoscópicas de la escena representada en la figura 2.7.

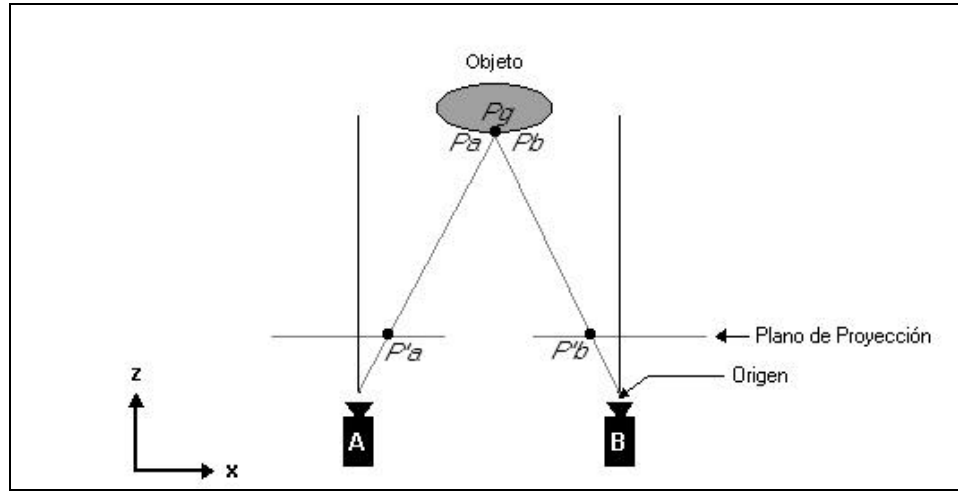


Figura 2.9: Modelo estereoscópico simple de cámaras paralelas.

La modelación estereoscópica que se utilizará se describe en la Figura 2.9. Cada punto de un objeto posee coordenadas en tres sistemas de referencia diferentes, que son el sistema global  $g$  y los sistemas locales de cada cámara (A y B). Los puntos proyectados son  $P'_a$  y  $P'_b$ .

$P_g$ : Punto tridimensional en el sistema de referencia global.

$P_a, P_b$ : Puntos correspondientes a  $P_g$  transformados al sistema de coordenadas de las cámaras A y B.

$P'_a, P'_b$ : Proyección en perspectiva de los puntos  $P_a$  y  $P_b$  sobre los planos de proyección de las cámaras A y B.

$T_{ga}, T_{gb}$ : Matrices de transformación de puntos del sistema global a los sistemas locales de las cámaras A y B.

$T_{ab}, T_{ba}$ : Matrices de transformación de puntos del sistema local de la cámara A hacia el sistema de la cámara B. Estas matrices son una combinación de una rotación y una traslación y cumplen que  $T_{ab} = T_{ba}^{-1}$ . Son de la forma:

$$T_{ab} = \begin{bmatrix} R_{3 \times 3} & t_{3 \times 1} \\ 0_{1 \times 3} & w_{1 \times 1} \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.4)$$

Se tienen las siguientes relaciones:

$$P_a = \begin{bmatrix} x_a \\ y_a \\ z_a \\ 1 \end{bmatrix} = T_{ga} \cdot P_g, \quad P_b = \begin{bmatrix} x_b \\ y_b \\ z_b \\ 1 \end{bmatrix} = T_{gb} \cdot P_g \quad (2.5)$$

$$P'_a = \begin{bmatrix} x'_a \\ y'_a \\ 1 \\ 1 \end{bmatrix} = H(T_{proy} \cdot P_a), \quad P'_b = \begin{bmatrix} x'_b \\ y'_b \\ 1 \\ 1 \end{bmatrix} = H(T_{proy} \cdot P_b) \quad (2.6)$$

Donde  $H$  es la función de homogeneización y  $T_{proy}$  es la matriz de proyección en perspectiva:

$$H \left( \begin{bmatrix} x \\ y \\ z \\ h \end{bmatrix} \right) = \begin{bmatrix} x/h \\ y/h \\ z/h \\ 1 \end{bmatrix}, \quad T_{proy} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & f & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (2.7)$$

$$\text{Luego, si } h = z, \quad H \left( T_{proy} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \right) = \begin{bmatrix} xf/z \\ yf/z \\ f \\ 1 \end{bmatrix} \quad (2.8)$$

#### 2.2.4 Geometría Epipolar

A partir de ciertas relaciones geométricas que se cumplen en el *modelo estereoscópico* se puede definir uno de los conceptos fundamentales del proceso de estereoscopia: la *geometría epipolar*.

En la proyección en perspectiva, un punto  $P'_a = (x'_a, y'_a, f)$  de la cámara A se puede asociar a la proyección en perspectiva de la línea formada por múltiples puntos tridimensionales  $P_a(z_a)$ . De la relación (2.6) se tiene que:

$$P_a = T_{proy} \cdot H^{-1}(P'_a) \quad (2.9)$$

Luego  $P_a(z_a)$  es de la forma:

$$P_a(z_a) = \begin{bmatrix} x'_a z_a / f \\ y'_a z_a / f \\ z_a \\ 1 \end{bmatrix} \quad (2.10)$$

La relación (2.10) expresa que para un mismo punto  $(x'_a, y'_a)$  en el plano óptico existen infinitos puntos en el espacio tridimensional a distintas distancias  $z_a$  con una misma proyección (ver Figura 2.10).

La proyección sobre la cámara B del conjunto de puntos  $P_a(z_a)$  asociado a un punto  $P'_a$  genera un conjunto de puntos  $P'_b(z_a) = (x'_b, y'_b, f)$  que forma la *recta epipolar* como se muestra en el plano de óptico B de la Figura 2.10.

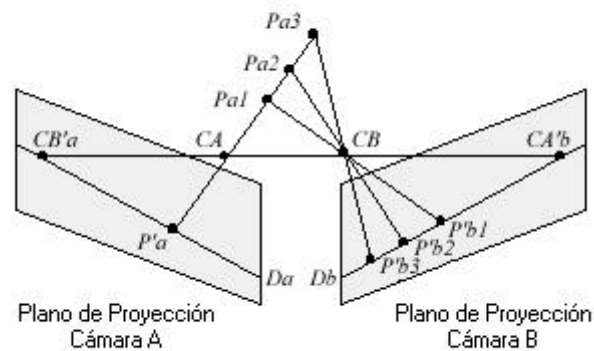


Figura 2.10: Rectas epipolares.

La recta epipolar asociada a un punto  $P'_a$  es  $D_b$ , y pasa por la proyección de todos los puntos  $P_a(z_a)$  a través del centro óptico  $CB$ . Todos los puntos  $(P_{a1}, P_{a2}, P_{a3})$  que se proyectan sobre  $P'_a$  se proyectan también sobre la recta epipolar  $(P'_{b1}, P'_{b2}, P'_{b3})$ . Hay que notar que en este caso los planos de proyección se

encuentran detrás de los centros ópticos para simplificar la figura. Las rectas epipolares también puede definirse como la intersección de un plano que contiene los centros ópticos y un punto en la escena con los planos de proyección.

Asumiendo que la profundidad  $z_a$  de cualquier punto  $P_a(z_a)$  está limitada a un cierto intervalo, es posible reducir la recta epipolar a un *segmento epipolar*, aún si el intervalo es muy grande (desde  $f$  hasta  $\infty$ ). El segmento epipolar representa todos los posibles puntos correspondientes en la imagen B a un punto  $P_a$  en la imagen A. Se establecerán  $z_{min_a}$  y  $z_{max_a}$  como los límites del intervalo de profundidades para la cámara A, asumiendo que son elegidos de tal forma que cualquier punto en este intervalo se proyecta en la región válida del plano óptico de la otra cámara.

Debido a que la proyección de una recta tridimensional produce una recta sobre el plano de proyección, se puede definir el segmento epipolar por los extremos del segmento de la recta formada por los puntos  $P_a(z_{min_a})$  y  $P_a(z_{max_a})$  donde  $z_a$  varía en el intervalo  $[z_{min_a}, z_{max_a}]$ , mientras que la proyección varía entre  $P'_b(z_{max_a})$  y  $P'_b(z_{min_a})$  (ver Figura 2.11).

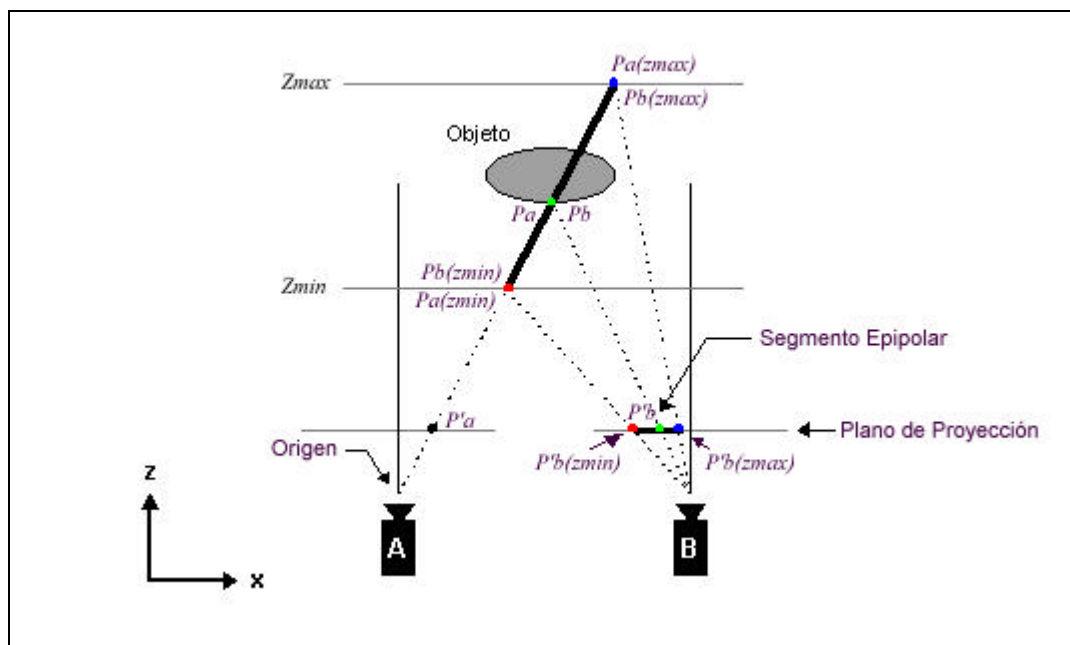


Figura 2.11: Segmento epipolar.



Cuando los planos de proyección son paralelos a la recta que une sus centros ópticos (*baseline*), los vectores epipolares en cada plano también lo serán. Esto simplifica enormemente el proceso de búsqueda estéreo, ya que la solución a este problema se transforma de un problema bidimensional a uno unidimensional. En la práctica es difícil lograr una alineación precisa, por ello se requieren etapas de calibración/rectificación que actualmente siguen siendo ampliamente estudiadas. El problema de calibración/rectificación es bastante extenso y está fuera de los objetivos de esta investigación, razón por la cual su descripción será breve.

Con la información de la geometría epipolar, es posible rectificar las imágenes (i.e. aplicarles una transformación geométrica, de modo que las rectas epipolares queden paralelas) antes de comenzar el proceso de búsqueda de correspondencias. Esto reduce drásticamente la demanda computacional de cualquier algoritmo de búsqueda de correspondencias, lo que permite que la estereoscopia pueda tener aplicaciones prácticas. Una búsqueda bidimensional para cada punto de una imagen requeriría de tiempos de procesamiento prohibitivos. Sin embargo, debemos recalcar que la rectificación sólo es posible si previamente se ha realizado una adecuada *calibración* del sistema de adquisición estéreo.

### 2.2.5 Calibración

Existen dos grupos de parámetros involucrados en todo modelo estereoscópico [Faug93]. El primer grupo describe la orientación y posición relativa de las cámaras mediante parámetros geométricos del sistema binocular, comúnmente llamados *parámetros extrínsecos*. Los parámetros extrínsecos están asociados a matrices de rotación y traslación entre los sistemas de coordenadas de cada cámara; siendo la separación entre las cámaras (*baseline*) y los ángulos de convergencia, los más relevantes en todo diseño. Los parámetros extrínsecos son suficientes para reconstruir tridimensionalmente una escena, pero a un cierto factor de escala con respecto a las dimensiones reales. Si se requiere una reconstrucción de la escena completa, es decir que incorpore las unidades métricas reales, un segundo grupo de parámetros debe conocerse. Estos parámetros se denominan *parámetros intrínsecos* porque no dependen de la posición y orientación espacial de las cámaras; y los principales son la distancia focal  $f$  y la resolución del sensor  $(\mathbf{r}_x, \mathbf{r}_y)$ .

En la práctica, la alineación perfecta es muy difícil de lograr, y los parámetros intrínsecos se ven afectados por la distorsión de los lentes o la desalineación de los sensores con respecto al plano óptico ideal. Por esta razón la calibración es un requisito en la mayoría de los sistemas de visión.

La calibración del sistema de adquisición estéreo tiene dos objetivos [Luo93]:

1. Corregir distorsiones y establecer la geometría epipolar del sistema para poder rectificar las imágenes, y así reducir la complejidad del proceso de búsqueda de correspondencias.
2. Permitir la reconstrucción tridimensional una vez que el proceso de correspondencias ha concluido.

Para la rectificación de las imágenes se requiere obtener los parámetros intrínsecos y extrínsecos de ambas cámaras. Esto significa encontrar la relación (proyección en perspectiva) entre los puntos tridimensionales de la escena y las imágenes registradas por las cámaras. Mediante la calibración se pueden estimar las distorsiones de las imágenes e incorporar esas deformaciones a la matriz  $T_{ga}$ . Además, se deben calibrar los desplazamientos y orientaciones relativas de las cámaras, modeladas por la matriz  $T_{ab}$  entre el sistema local de la cámara A y el de la cámara B.

Algunos modelos suponen algunos parámetros ya fijos. El proceso de calibración utiliza pares de puntos correspondientes conocidos que se determinan manualmente o automáticamente.

La calibración de la matriz  $T_{ga}$  se obtiene mediante la solución de un sistema de ecuaciones lineales. La matriz  $T_{ga}$  contiene doce parámetros a determinar, de los cuales seis están relacionados al desplazamiento y orientación, y los seis restantes corresponden a los parámetros intrínsecos. Los seis parámetros intrínsecos considerados son la densidad de pixels en el sentido vertical y horizontal ( $k_u$ ,  $k_v$  pixels/mm), la distancia focal  $f$ , las coordenadas  $(u_0, v_0)$  del centro de la imagen (resultantes de la intersección del *eje óptico* con en el plano de la imagen), y el ángulo

entre los ejes de los planos de la imagen (muy cercano a  $90^\circ$ , ya que las celdas del CCD son idealmente rectangulares).

Si la escala no se considera en unidades métricas, sino simplemente como factores  $\mathbf{a}_u = f \cdot k_u$ , y  $\mathbf{a}_v = f \cdot k_v$ , el número de ecuaciones se reduce a once [Luo93]. El sistema de once ecuaciones lineales se soluciona con un mínimo de seis pares de puntos correspondientes elegidos adecuadamente (i.e. no coplanares en el espacio tridimensional), si las posiciones exactas en el espacio tridimensional global (exterior a la cámara) son conocidas [Horn86].

Existen básicamente dos procedimientos generales para solucionar el problema de la calibración de la matriz  $T_{ab}$  [Horn86]:

#### **A. Calibración con Coordenadas Tridimensionales Conocidas**

Este método, también conocido como *calibración por orientación absoluta*, utiliza un objeto de referencia del cuál se conocen las coordenadas tridimensionales exactas dentro del sistema local de cada cámara. Así, para cada par de puntos correspondientes identificados, la posición tridimensional es conocida. En este caso, 4 puntos (no coplanares en el espacio tridimensional) son suficientes para producir un sistema de ecuaciones lineales cuya solución da directamente los coeficientes de la matriz  $T_{ab}$ .

#### **B. Calibración con Coordenadas Tridimensionales Desconocidas**

En este método, llamado también *calibración por orientación relativa*, sólo están disponibles los pares de puntos correspondientes. El sistema de ecuaciones es no lineal y al utilizar la restricción de ortonormalidad de la rotación y la eliminación del factor de escala, es posible obtener la matriz  $T_{ab}$  a partir de 5 puntos correspondientes elegidos apropiadamente (no coplanares en el espacio tridimensional).

En la práctica, el número de puntos correspondientes elegidos para resolver un sistema de ecuaciones es siempre superior a los requerimientos mínimos expuestos anteriormente. El uso de un mayor número de pares aumenta la precisión de calibración y permite validar los parámetros de calibración. Los sistemas de

ecuaciones de este tipo pueden ser solucionados usando métodos de aproximación de mínimos cuadrados.

Los métodos expuestos utilizan un patrón de calibración (como un objeto de referencia cuadrulado de dimensiones conocidas) del cual se extraen puntos específicos para obtener un número suficiente de combinaciones del tipo *{punto proyectado, punto tridimensional}* y así poder calcular las matrices de proyección en perspectiva. En [Tsai86] se describe un algoritmo frecuentemente citado en la literatura. Este se basa en un modelo de once parámetros. El código fuente del algoritmo y su documentación se encuentran disponibles en [web12].

Sin embargo, el uso de patrones de calibración puede ser problemático para algunas aplicaciones, especialmente las de operación remota y las que requieren de calibración en línea del sensor mientras se procesan las imágenes con otros fines. Por esta razón se ha intentado desarrollar técnicas de autocalibración para recuperar tanto los parámetros intrínsecos como extrínsecos [Luo93], o en el caso de la *calibración débil* solamente estos últimos bastarían para conocer la geometría epipolar. Algunos métodos de autocalibración utilizan información sobre el movimiento que han realizado las cámaras [Viev94], otros tratan de extraer esta información a partir de las variaciones en las imágenes mediante técnicas como el análisis del flujo óptico. En [Zhan93] se describe un caso particular de [Luo93] en que sólo se requiere un movimiento para recuperar la geometría epipolar. En [Zhan94] se describe un método incorporado al algoritmo de *mactching* estéreo para recuperar la geometría epipolar, que luego de una búsqueda inicial bidimensional permite obtener las ecuaciones de las rectas epipolares, para finalmente recalculer las disparidades mediante una búsqueda sobre las líneas epipolares predecidas.

Métodos sencillos para rectificar las imágenes, mediante una transformación geométrica (*warping*) de modo que las rectas epipolares queden paralelas se describen en [Papa96] y [Coch90]. El método de [Coch90] no es automático, pero es importante mencionarlo por su sencillez, ya que no requiere de la determinación de las ecuaciones de las rectas epipolares, y porque puede ser la base de un método de rectificación más sofisticado.

Todas estas técnicas sólo requieren conocer un número reducido de puntos correspondientes entre las imágenes, no siendo necesario el uso de un patrón de calibración. La desventaja de estos métodos es que no obtienen en forma automática el factor de escala que permita convertir las mediciones, en unidades de pixels, a unidades métricas; y en general no consideran las distorsiones producidas por los lentes. Para obtener el factor de escala se requieren de pasos adicionales que generalmente se basan en la comparación de objetos de tamaño conocido. En [Ste96] se describe un método para la calibración de la distorsión de los lentes usando sólo correspondencias de puntos en múltiples imágenes sin la necesidad de conocer su posición en el espacio o la posición de las cámaras.

### 2.2.6 Definición de Disparidad y Cálculo de las Coordenadas 3D

En la Figura 2.12 se muestra un sistema binocular típico. En éste los ángulos  $\theta$ ,  $\alpha$ ,  $\beta$  corresponden respectivamente a los ángulos de rotación, elevación y convergencia o azimut (también conocidos como ángulos RPY: roll, pitch, yaw).

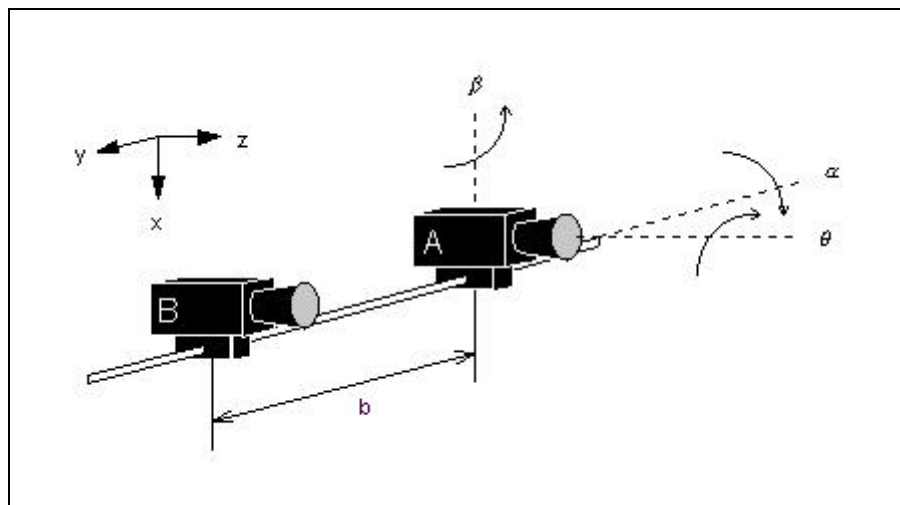


Figura 2.12: Sistema binocular típico.

Para simplificar el análisis se asume que los sensores están colocados adecuadamente, i.e. alineados tal que  $\alpha = 0^\circ$ ,  $\theta = 0^\circ$ . A partir del análisis geométrico del sistema mostrado en la figura 2.13, cuyas vistas de planta y lateral se muestran en las figuras 2.14 y 2.15, respectivamente, se obtienen las siguientes expresiones:

$$\frac{b}{z} = \left[ \frac{f \cdot \tan \mathbf{b} + u_b}{f - u_b \cdot \tan \mathbf{b}} + \frac{f \cdot \tan \mathbf{b} - u_a}{f - u_a \cdot \tan \mathbf{b}} \right] \quad (2.11)$$

Donde  $b$  es el *baseline* (también llamado distancia interfocal), y  $f$  es la distancia focal.

$$x = \frac{z}{2} [\tan(\mathbf{q} + \mathbf{g}_b) - \tan(\mathbf{q} - \mathbf{g}_a)] \quad (2.12)$$

$$\text{con} \quad \mathbf{g}_b = \text{atan}\left(\frac{u_b}{f}\right) \quad \mathbf{g}_a = \text{atan}\left(\frac{u_a}{f}\right) \quad (2.13)$$

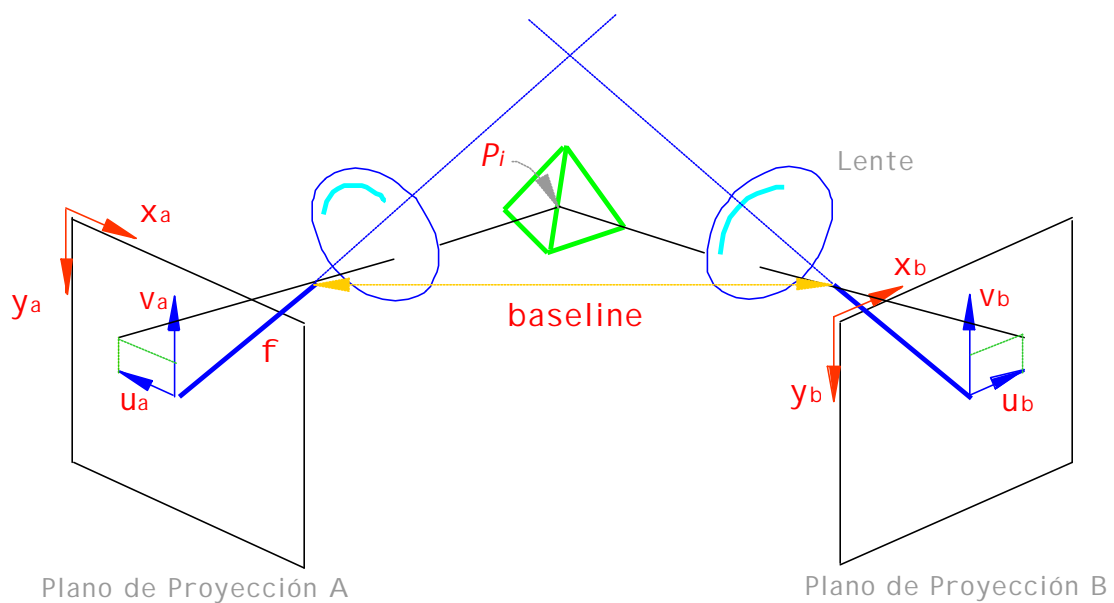


Figura 2.13: Vista en perspectiva del sistema binocular.

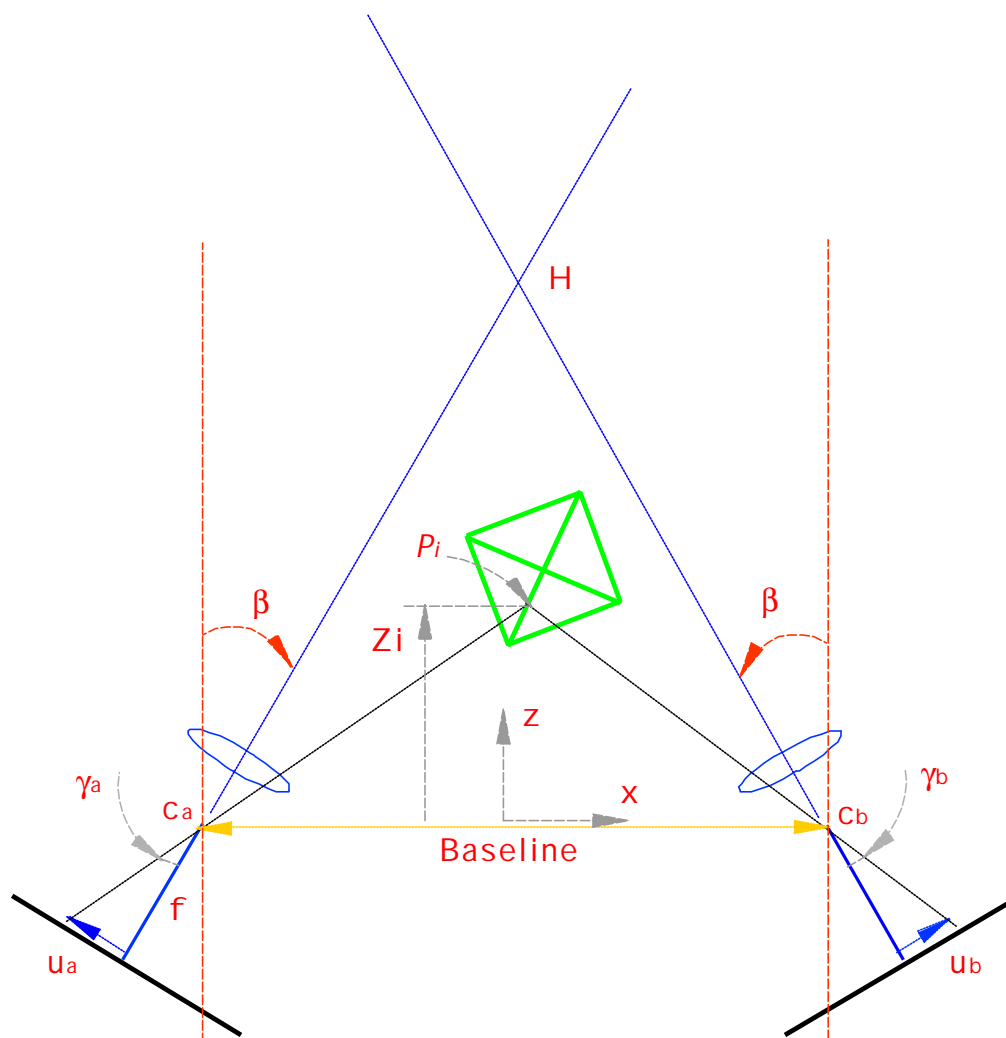


Figura 2.14: Vista de planta del sistema binocular.

La definición de disparidad resulta sencilla si las rectas epipolares son paralelas, es decir el ángulo de convergencia  $\mathbf{b}$  es nulo (en este caso los ejes ópticos  $c_aH$  y  $c_bH$  de Figura 2.14 serían paralelos). Dado un punto  $P_i(x,y,z)$  en el espacio

tridimensional, cuyas respectivas proyecciones sobre los planos de las imágenes son  $P'_a(u_a, v_a)$  y  $P'_b(u_b, v_b)$ , de (2.11) con  $\mathbf{b} = 0^\circ$  se obtiene  $z$  como:

$$z = \frac{b \cdot f}{u_b - u_a} = \frac{b \cdot f}{d} \quad (2.14)$$

En la práctica se tienen pixels en el sistema de coordenadas  $x$ - $y$  (ver Figura 2.13) que corresponde al estándar generalizado utilizado por los dispositivos gráficos de los equipos computacionales. En este caso las relaciones entre las coordenadas  $u$ - $v$  y las  $x'$ - $y'$  son:

$$x'_i = \frac{u_i}{\mathbf{r}_x} + \frac{w_x}{2} \quad \text{e} \quad y'_i = \frac{v_i}{\mathbf{r}_y} - \frac{v_i}{\mathbf{r}_y} \quad (2.15)$$

donde:  $i$  indica la cámara A o B.

$w_x, w_y$  corresponden a las longitudes horizontal y vertical del CCD, respectivamente.

$\mathbf{r}_x, \mathbf{r}_y$  son las longitudes horizontal y vertical de un pixel (mm/pixel).

Luego, de la relación (2.14),  $z$  se expresa como:

$$z = \frac{b \cdot f}{\mathbf{r}_x \cdot (x'_b - x'_a)} = \frac{b \cdot f}{\mathbf{r}_x \cdot d} \quad (2.16)$$

donde  $d$  es la *disparidad* (*paralaje* en fotogrametría) para el punto  $P'_a$  o  $P'_b$ , la cual se define como:

$$d = d(P'_a) = d(P'_b) = x'_b - x'_a \quad (2.17)$$

La coordenada  $x$  en el espacio tridimensional para el punto  $P_i$  estaría dada por las siguientes expresiones en el sistema paralelo:

$$x = \frac{b}{2d} \cdot (u_b - u_a) = \frac{b}{2d} \cdot \mathbf{r}_x \cdot (w_x - (x'_b + x'_a)) \quad (2.18)$$



Más sencilla aún es la obtención de la expresión para la coordenada tridimensional  $y$  de  $P_i$  a partir del análisis geométrico que se muestra en la Figura 2.15 para una vista lateral de una cámara. La posición y resulta ser:

$$y = z \cdot \frac{v_b}{f} = z \cdot \frac{v_a}{f} = \frac{b}{d} \cdot \mathbf{r}_y \cdot \left( y'_b - \frac{w_y}{2} \right) = \frac{b}{d} \cdot \mathbf{r}_y \cdot \left( y'_a - \frac{w_y}{2} \right) \quad (2.19)$$

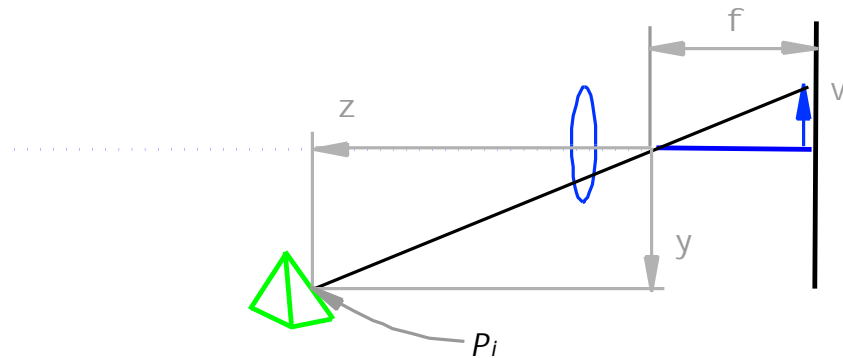


Figura 2.15: Vista lateral de una cámara.

La profundidad es inversamente proporcional a la disparidad y viceversa, como puede apreciarse en la relación (2.16). Otra característica que se desprende de esta relación es que para un punto de la escena a una distancia fija (i.e.  $z = z_0$ ) si se aumenta el *baseline* o la distancia focal, su disparidad aumenta. En otras palabras el *baseline* o la distancia focal actúan como coeficientes amplificadores de la disparidad. Este resultado es muy relevante en la determinación del error en la medición de la profundidad asociado a una cierta longitud de *baseline* como se verá más adelante.

Suponiendo que el sistema ha sido adecuadamente calibrado y alineado, y que las imágenes han sido rectificadas para corregir posibles distorsiones, el proceso de búsqueda de correspondencias entregará como resultado un *mapa de disparidades*. El mapa de disparidades se puede entender como una imagen cuyas intensidades no representan luminosidad en la escena, sino más bien la disparidad de cada punto en la escena. Cómo puede verse de la relación (2.16), los puntos de la escena próximos a

las cámaras, es decir aquellos puntos con  $z$  pequeño tendrán disparidades mayores que los puntos lejanos, con  $z$  más grande. Al ser las intensidades de la imagen o mapa de disparidades inversamente proporcionales a las de la imagen o mapa de profundidades  $z$ , visualmente la imagen negativa de un mapa de disparidades es muy similar a al mapa de profundidades.

Por esta razón la mayoría de los resultados se presentarán en mapas de disparidades, donde los puntos cercanos a las cámaras (puntos con disparidades altas) se representarán con tonalidades de grises claras y los lejanos con intensidades de grises oscuros. El blanco corresponde a la disparidad máxima del intervalo de búsqueda de disparidades y el negro corresponde al mínimo. El color rojo oscuro se utiliza para indicar los puntos que al someterse a validación no fueron aprobados, mientras que el verde oscuro indica aquellos puntos cuya disparidad resultó ser superior a un cierto umbral prefijado. La Figura 2.16 muestra el mapa de disparidades correspondiente a la escena de la Figura 2.7.

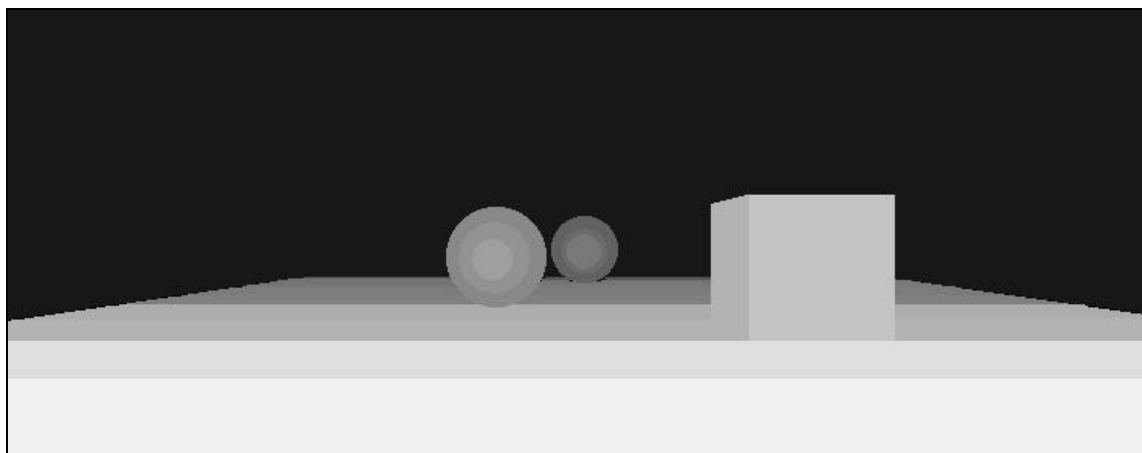


Figura 2.16: Mapa de disparidades para la vista de la cámara A de la figura 2.7.

### 2.3 Elección de Características

Las características usadas en estereoscopía deben ser extraíbles en forma confiable de las imágenes. Existen numerosos candidatos a ser utilizados [Faug93]:

- a) **Pixels:** La unidad elemental contiene información de intensidad lumínica de los puntos a comparar  $P'_a$  y  $P'_b$  de cada imagen. La luz reflejada por la superficie de un objeto es una función de la posición de las fuentes, la orientación de la superficie y el punto de vista [Horn86]. Este comportamiento se puede resumir en la función de la reflectancia. Para una superficie *lambertiana* (una superficie completamente mate), la luz reflejada es la misma en todas las direcciones [Irwi95]. Como resultado, las intensidades de los puntos correspondientes son las mismas. Sin embargo, muy pocas superficies son del tipo *lambertiana*, y el caso extremo es el de superficies especulares. En este tipo de superficies la intensidad de los dos puntos correspondientes es generalmente distinta. Entre estos dos extremos existen varias clases de superficies para las cuales la luz reflejada cambia lentamente con la dirección del punto de vista; en este caso los puntos correspondientes tienen intensidades parecidas si la distancia entre los centros ópticos de las imágenes es pequeña en comparación a la distancia de la escena. Muchas de las técnicas clásicas de correlación para resolver el problema de la correspondencia estéreo están basadas en la idea de correlacionar las intensidades de las imágenes estereoscópicas. Esto funciona bien para superficies cuya función de reflectancia es cercana a la de superficies *lambertianas*, pero no para otras [Faug93].
- b) **Bordes:** Pueden tener asignados atributos geométricos de forma (líneas y curvas), orientación e intensidad. Exceptuando el caso de bordes en discontinuidades de profundidad cuyas normales cambian suavemente desde el punto de vista del observador, la mayoría son fuentes de información confiables para guiar el proceso de búsqueda estéreo [Faug93]. Algunos bordes son un problema, ya que ambas cámaras no ven la misma parte del objeto. Sin embargo, no en todos los casos es posible extraer bordes en una escena. La calidad de los métodos basados en comparación de bordes y segmentos está muy limitada por la eficiencia del proceso de extracción de bordes.

- c) **Regiones de la Imagen:** Se pueden usar regiones de las imágenes como características a comparar. Dependiendo de como sean extraídas estas regiones, sus formas y atributos de intensidad, color y textura pueden o no ser invariantes ante la transformación en perspectiva. La dificultad de usar estas características se debe a la dificultad en segmentar la imagen y a la alta dependencia que tienen con el tipo de escena.

## 2.4 Búsqueda de Correspondencias

La tarea central y a la vez la más difícil del proceso de estereoscopía es la solución del *problema de búsqueda de correspondencias*. La búsqueda de los puntos correspondientes a la proyección de un punto de la escena sobre cada plano óptico para recuperar la profundidad es bastante compleja debido a los siguientes factores [Horn86, Coch92]:

- a) **Variación de la intensidad luminosa:** La intensidad luminosa observada de un punto en la escena varía al desplazar la cámara en función del ángulo entre la luz incidente sobre el punto y el eje óptico de la cámara.
- b) **Fenómeno de oclusión:** Un punto de una escena visible en una imagen no es necesariamente visible en la otra imagen.
- c) **Ruido y errores de muestreo:** La digitalización puede conllevar imprecisiones debido a la resolución finita de las imágenes obtenidas.
- d) **Textura de los objetos:** Los objetos observados pueden presentar texturas que aumentan la ambigüedad de la búsqueda de puntos correspondientes. Una textura de muy alta frecuencia (patrones periódicos) aumenta el número de puntos similares (la curva de correlación tiene varios máximos locales), mientras que una textura de muy baja frecuencia (objetos lisos o de escasa textura) hace la búsqueda del punto correspondiente muy difícil (todos los puntos en ambas imágenes tienden a parecerse, por lo que su correlación es máxima y la búsqueda del máximo global se dificulta).

Cuando la textura se repite, como por ejemplo en ladrillos de una pared, pueden existir múltiples correspondencias. Este problema produce correspondencias falsas.

La mayoría de los objetos presentes en escenarios naturales posee texturas. Aunque la textura no necesariamente ayuda a identificar características abstractas, como segmentos, forma la base para los algoritmos basados en área, debido a que presenta un patrón estadísticamente relevante. Sin embargo, otros escenarios, como los ambientes industriales, pueden poseer objetos sin texturas. En este caso no es posible obtener ninguna correspondencia. Una posible solución es proyectar un patrón de manchas aleatorias para aumentar de manera artificial la textura en la escena [Nish84].

- e) **Errores de calibración:** Los errores de calibración, alineación y rectificación no sólo dificultan la búsqueda de puntos correspondientes, sino que contribuyen a aumentar los errores en la reconstrucción tridimensional. Parte del trabajo desarrollado corresponde al análisis de los errores y su efecto general sobre el proceso estereoscópico. El análisis de los errores es fundamental por tres razones. En primer lugar, permite diseñar un sistema que reduzca los errores. Por otro lado, el análisis permite establecer los rangos de precisión esperada. Y por último, al conocer las fuentes de error es posible enfrentarlas con algoritmos de búsqueda de correspondencia dedicados.

Para aminorar los efectos mencionados, pueden aplicarse una serie de restricciones que suponen características específicas de las superficies que se encuentran en la escena, siendo las más conocidas y útiles [Faug93]:

- a) **Restricción epipolar:** Ya mencionada anteriormente, esta restricción permite convertir el problema de búsqueda de dos dimensiones a una dimensión, reduciendo considerablemente el tiempo de cálculo.
- b) **Restricción de continuidad:** En este caso se suponen superficies *suaves*. Esta restricción funciona muy bien al permitir descartar disparidades que no sean “apoyadas” por disparidades vecinas en superficies continuas, pero falla ante la presencia de discontinuidades.
- c) **Restricción de oclusión:** En el caso de una oclusión, un punto en la imagen A puede no ser perceptible en la imagen B. Mediante una simple comparación de los mapas de disparidades de A a B y B a A se pueden eliminar los puntos que no tienen posibilidad de ser correlacionados.

### 2.4.1 Métodos de Correspondencia

Los primeros trabajos en estereoscopía utilizando equipos computacionales datan de 1973, pero no ha sido sino hasta casi veinte años después que los primeros resultados prácticos han sido logrados. Durante los últimos años se han desarrollado diversos métodos para solucionar el problema de correspondencia. Revisiones de éstos métodos se presentan en: [Barn82], [Dhon89], [Lane96] y [Faug93]. Se pueden reconocer básicamente cuatro categorías importantes: métodos basados en área, basados en características, basados en relajación de restricciones y basados en el dominio de la frecuencia. Una descripción breve de estos métodos se encuentra en el Anexo A.

### 2.5 Determinación de Profundidad y Reconstrucción 3D

El objetivo del proceso estéreo es la *determinación de la profundidad*. La *determinación de la profundidad* para la reconstrucción 3D de la escena a partir de las disparidades debería ser la última etapa del proceso. Pero en la práctica, una etapa adicional de interpolación es muchas veces necesaria para completar los vacíos de información que no se pudo obtener en la etapa de *matching*.

El problema de *correspondencias* ha sido considerado como el más difícil y el más importante dentro del proceso estéreo, por esta razón la fase de determinación de distancias no ha sido muy estudiada como un problema específico. Como se explicó en la sección 2.2.6, la triangulación a través del sistema binocular de cámaras paralelas y el proceso de *matching*, permiten obtener las disparidades que son necesarias para calcular las coordenadas de los puntos en el espacio tridimensional. Sin embargo, esta etapa del proceso estéreo en que la disparidades deben convertirse a distancias presenta dificultades significativas; especialmente si las correspondencias son inexactas o poco confiables.

Como se verá en detalle en la sección 3.1 (sobre un aspecto esencial relacionado a la reconstrucción 3D: las causas de las alteraciones del proceso de reconstrucción), la determinación de la profundidad esta sujeta a errores: (1) que se originan en la *etapa de correspondencias*, como el error de posición en la correspondencia, el cual causa un error en la profundidad inversamente proporcional,

o (2) que provienen de errores en la calibración, por ejemplo errores en la medición de la longitud del *baseline* causan errores directamente proporcionales en la estimación de la profundidad.

Los errores en la etapa de *matching* se incrementan al aumentar la longitud del *baseline*, ya que el rango de búsqueda de disparidades aumenta. La ambigüedad en la determinación de correspondencias también aumenta al reducir el tamaño de la ventana de correlación. La elección de los parámetros apropiados del sistema debe basarse en el análisis del error que provocan en la estimación de  $z$ .

Se han desarrollado técnicas o enfoques distintos que buscan mejorar las mediciones de profundidad, éstos se pueden agrupar en las siguientes categorías:

1. Mejoras sobre los algoritmos de *matching*.
2. Estimación sub-píxel de la disparidad.
3. Uso de múltiples *baselines*.
4. Uso de secuencias o procesamiento estadístico de distintas vistas estereoscópicas.

El primer método en realidad es un enfoque indirecto al problema, y como se mencionó ha sido objeto de mucho estudio (existe una gran variedad de algoritmos de *matching*, basados en patrones distintos, algunos utilizando segmentación de la escena, otros utilizando múltiples resoluciones o ventanas de tamaño variable, etc.).

El segundo punto, la consideración de disparidades sub-píxel, normalmente forma parte del proceso de *matching*. En el caso en que se utiliza correlación de bloques, lo que interesa es encontrar la disparidad en el punto de correlación máxima, el cual puede determinarse por interpolación de la curva de correlación en la vecindad al máximo.

El uso de múltiples *baselines*, ya sea con sistemas trinoculares o de más cámaras, es beneficioso, ya que permite reducir las ambigüedades que se generan en los sistemas binoculares de *baseline* fijo. En [Kana93] se demuestran claramente las

contribuciones positivas de usar múltiples *baselines*. Recordando la figura 2.1, en la que se muestra la proyección de la escena sobre varias cámaras, y utilizando las relaciones derivadas en los puntos anteriores, es posible obtener para cada combinación de pares de cámaras una relación de la forma:

$$z_{ij} = \frac{b_{ij} \cdot f}{\mathbf{r}_x \cdot \mathbf{d}_{ij}} \quad (2.20)$$

Donde  $ij$  es el par formado por la cámara  $i$  y la  $j$  (siendo  $i \neq j$ ). Como las distancias a la escena son iguales para un conjunto de cámaras sobre un mismo plano se debe cumplir que  $z_{ij} = z_{mn}$  para todo par  $ij$  distinto del  $mn$ . Si todas las cámaras tienen el mismo tipo de lente (igual  $f$ ), de (2.20) se desprende que para un *baseline*  $b_{ij}$  grande deben tenerse disparidades  $d_{ij}$  grandes, mientras que para un *baseline*  $b_{ij}$  pequeño debe tenerse disparidades  $d_{ij}$  pequeñas para mantener  $z_{ij}$  constante. El usar *baselines* grandes permite una mayor precisión porque la relación disparidad a tamaño del pixel es mayor (i.e. aumenta la resolución). Sin embargo, al aumentar los rangos de disparidades, encontrar el pixel correspondiente se hace más complejo y sujeto a errores. Por eso *baselines* menores entregan disparidades más confiables pero menos precisas. Combinando los resultados para distintos *baselines* se pueden eliminar ambigüedades y aumentar la precisión de la reconstrucción 3D.

La cantidad de ambigüedades que se pueden eliminar si se usa un par de cámaras perpendicular a otro es incluso mayor, ya que los pixels de segmentos o patrones uniformes de la imagen, paralelos al *baseline* de un par, son difíciles de corresponder. En el otro par estos los pixels de segmentos serían fácilmente correspondidos porque serían perpendiculares al *baseline* de ese par.

Al incorporar múltiples *baselines* se usan múltiples vistas de una escena, todas tomadas desde una misma posición. En el último enfoque se usan también múltiples vistas, pero la principal diferencia está en que éstas habrían sido adquiridas desde distintas posiciones en distintos instantes de tiempo. El procesamiento de pares estereoscópicos tomados desde distintas posiciones también debería contribuir a reducir las ambigüedades e incorporar mayor información al proceso de reconstrucción tridimensional, por ejemplo aportando información sobre objetos que desde una cierta posición pueden aparecer ocluidos. Si bien con este enfoque se



podría lograr una reconstrucción tridimensional más acabada, su complejidad es mayor porque se requiere información adicional para referir los sistemas de coordenadas de cada par en cierta posición con respecto a un sistema de coordenadas global (i.e. se requiere información sobre la traslación y rotación de las cámaras al cambiar de posición para tomar un nuevo conjunto de pares estéreo).

El procesamiento de secuencias de imágenes ha sido regularmente estudiado con fines de obtener el desplazamiento de objetos dentro de una escena o un *feedback* sobre el desplazamiento de las cámaras en una escena de objetos estáticos. La mayoría de los desarrollos que utilizan el procesamiento de secuencias se basa en la obtención del flujo óptico monocular [Horn86, Roy92, Bab94]. En menor número se ha estudiado el flujo óptico obtenido a partir de secuencias estereoscópicas [Roy93, Yi96, Yi97] o el uso de secuencias de pares estéreo para reconstruir tridimensionalmente una escena [Koch93, Koch94, Falk95].

## 2.6 Interpolación

La interpolación es necesaria porque no siempre es posible tener una mapa de disparidades denso, debido a que para algunos puntos no se pudieron encontrar correspondencias o éstas fueron erróneas. Los escasos puntos resultantes de los métodos basados en bordes o los puntos menos confiables de los métodos basados en área, hacen que la interpolación sea necesaria para llenar vacíos o reemplazar errores.

La interpolación de superficies utilizando *splines* es frecuente; sin embargo, la interpolación no es la solución apropiada cuando las imágenes contienen oclusiones y existen fuertes discontinuidades en la escena [Moha89, Hoff89], pues reduce la precisión de los resultados.

### III. TRABAJO DESARROLLADO

#### *Objetivo Principal del Trabajo*

*Desarrollar nuevas técnicas y mejorar las existentes para obtener la mayor cantidad posible de información de distancias  $z$  a los puntos en la escena, con la mayor exactitud, precisión y confiabilidad posible; utilizando para ello el procesamiento de secuencias de imágenes estereoscópicas.*

Teniendo presente el *objetivo principal*, el trabajo desarrollado se inicia con un análisis teórico de los principales aspectos que influyen en la exactitud y precisión del sistema estéreo, es decir, se analizan las causas del error en la estimación tridimensional. El propósito de este análisis es establecer un cierto criterio para el diseño del sistema y la elección apropiada de sus parámetros. Como resultado de esta primera etapa, se establece un procedimiento sistemático para determinar las especificaciones del sistema como por ejemplo el *baseline*, de modo que las causas de error sean mínimas.

Una vez determinados los parámetros físicos óptimos del sistema, el análisis se centra en los algoritmos de búsqueda de correspondencias. La segunda etapa del trabajo desarrollado tiene por objetivo establecer cuales son las causas que introducen errores al procesar las señales (imágenes), y plantear técnicas de procesamiento que permitan obtener mediciones exactas y confiables.

Se propone el uso de secuencias de pares estereoscópicos como método para mejorar la densidad y precisión de la disparidades calculadas. Si bien el uso de secuencias estereoscópicas ha recibido atención reciente, con un enfoque orientado a la reconstrucción 3D del modelo [Koch93i], no se conocen trabajos que hagan un énfasis en el uso de la secuencia de pares con el fin de reducir las ambigüedades en la etapa de *matching*, y por consiguiente, aumentar la precisión de las estimaciones de distancia.

En forma novedosa se incorpora un índice o factor de confiabilidad de los resultados de búsqueda de correspondencias, a partir del análisis de las curvas de correlación.

El trabajo se desarrolló considerando varios supuestos que permitieron aislar las fuentes de error en la estimación de posiciones en el espacio tridimensional, y concentrarse en la implementación de algoritmos que entreguen mediciones confiables. Los supuestos más importantes son:

1. La secuencia de imágenes se tomó con cámaras paralelas calibradas, razón por la cual las imágenes no requieren de una rectificación adicional. Esto simplifica enormemente el proceso de búsqueda de correspondencias, restringiéndola a las líneas de barrido horizontal de las cámaras.
2. Se tiene un *feedback* preciso de los movimientos realizados por las cámaras.

La validez de estos supuestos se verificó en forma experimental.

En resumen, el trabajo realizado entrega herramientas que permiten fijar las bases del proceso de diseño y construcción de un sistema estereoscópico real.

### **3.1 Análisis del Error en la Estimación de la Profundidad**

El análisis de las causas de error en la estimación de la profundidad es fundamental para determinar los parámetros más críticos involucrados en el diseño de un sistema estereoscópico.

Las fuentes de error se pueden clasificar en tres grandes grupos:

1. Errores de *Matching*: Ocurren por las razones que se explican en el punto 2.4 (Búsqueda de Correspondencias).
2. Estructura Geométrica del Sistema Binocular: Parámetros fijos del diseño como el ángulo de convergencia, la distancia interpíxel, la distancia focal y el *baseline* influyen en la magnitud del error de reconstrucción tridimensional.

3. Exactitud con que se conocen los parámetros de la geometría del sistema, i.e. calibración y alineación de las cámaras.

El análisis del error en la reconstrucción 3D realizado por otros investigadores ha tenido uno de los cuatro enfoques:

1. **Análisis de los errores en la medición 3D debido a la cuantización de la imagen:** Al digitalizar la imagen, i.e. discretizar la realidad continua que se proyecta sobre el plano del CCD, la pérdida de continuidad se puede modelar como ruido [Oppe75]. Además puede existir interferencia electromagnética que cause errores en los datos ya digitalizados. Todo esto se traduce en que la posición aparente de un punto en la escena que se proyecta sobre el CCD es distinta a la real. En [Blos87, Chan94, Rodr90] se modela el ruido en forma estocástica, y se establecen expresiones para el error esperado en la medición de profundidad debido a errores en la cuantización. Mientras que en [McVe82] se realiza un análisis más sencillo, sin suponer una función de distribución de probabilidades para el ruido, pero que permite fijar los parámetros del sistema para el peor caso (error de cuantización máximo).
2. **Análisis de los Errores en la medición 3D debido a correspondencias falsas:** Son errores en la medición de la distancia a los puntos de la escena debido a las ambigüedades que se generan buscar el patrón o región correspondiente a una región de referencia en la otra imagen. Como ya se mencionó, las ambigüedades se deben principalmente a la repetitividad de patrones o a la escasa textura de una escena. En [Moha89] se analiza la detección de errores y su corrección en el caso de algoritmos de correspondencias basados en el *matching* de bordes.
3. **Análisis de los errores en el movimiento de las cámaras:** Especialmente los sistemas estereoscópicos sobre plataformas en movimiento, están sujetos a desplazamientos y vibraciones incontrolables. El análisis de [Keht94] demuestra que para reducir el error en la estimación de profundidad, se debe elegir una longitud de

*baseline* adecuado; sin embargo su análisis no considera errores en la medición del desplazamiento, que serían importantes para los sistemas que procesan secuencias de imágenes para obtener información de posición, distancias o movimiento.

4. **Errores de alineación de las cámaras:** La medición incorrecta de los parámetros geométricos del sistema, o las desviaciones que sufren las componentes del sistema con respecto a una posición ideal, se pueden clasificar como errores de alineación. En [Zhao95] se realiza un análisis para variaciones de algunos parámetros del sistema, como el ángulo de convergencia, pero carece de un análisis para variaciones en el *baseline* o la distancia focal; realiza muchas aproximaciones que el autor de esta investigación considera incorrectas cuando se requiere conocer con precisión los efectos de las desalineaciones sobre los parámetros del sistema.

El análisis del error que se presenta a continuación es parte del trabajo realizado por el autor, excepto donde se hace referencia a los resultados de otro autor con fines comparativos. En algunos casos las expresiones obtenidas coincidieron con las presentadas en [McVe82] y [Zhao95].

El análisis realizado está asociado a los enfoques 1, 3 y 4, tratando de presentarlos en una forma coherente; considera aspectos muy simples, pero que son esenciales para el diseño de un sistema estereoscópico.

Al igual que en [McVe82], [Keht94] y [Zhao95], el análisis que se presenta pone énfasis en el diseño del sistema mediante la elección de parámetros que permitan cumplir una cierta tolerancia.

### 3.1.1 Efectos de los Errores de Calibración

Considérese el sistema binocular típico de la figura 2.12. Idealmente ambas cámaras deberían tener sus ejes ópticos paralelos entre sí, y perpendiculares a la barra que soporta las cámaras. Para la discusión se analiza cada fuente de error

por separado, asumiéndose que todos los demás factores son *ideales*. Además se asume que no ocurren errores en la etapa de búsqueda de correspondencias.

Pueden distinguirse las siguientes fuentes de error:

**Errores del Sistema Binocular:**

- A. Error en la Medición del *Baseline* ( $\Delta b$ ).
- B. Error debido a Rotación entre las cámaras ( $\mathbf{q}$ , *roll*).
- C. Error debido a la Inclinación entre las cámaras ( $\mathbf{a}$ , *pitch*).
- D. Error debido a la Convergencia entre las cámaras ( $\mathbf{b}$ , *yaw*).

**Errores del Sistema Monocular:**

- E. Errores debido a los lentes (distancia focal) y CCDs no paralelos.
- F. Errores debido a la distorsión introducida por los lentes.

El análisis del error se realizará con la ayuda de la figura 3.1, que representa geoméricamente el sistema de la figura 2.12.

**A. Error en la Medición del *Baseline***

Si por errores de calibración se introduce un error en la medición del *baseline*  $\Delta b$ , se afectará la distancia estimada  $z$  en una cantidad dada por (de 2.16):

$$\Delta z_b = \frac{f \cdot \Delta b}{\mathbf{r}_x \cdot d} \quad (3.1)$$

Entonces, utilizando las expresiones (2.16) y (3.1) el error relativo sería:

$$\frac{\Delta z_b}{z} = \frac{\Delta b}{b} \quad (3.2)$$

La expresión (3.2) revela que conviene tener *baselines* grandes, porque de este modo el sistema se insensibiliza a errores cometidos al medir el *baseline*. Este resultado concuerda con los resultados obtenidos por [Rodr90] y [Keht94], quienes asumen funciones de distribución para el error  $\Delta b$ , y concluyen que la probabilidad de

error en la estimación disminuye a medida que el *baseline* o la distancia focal aumentan.

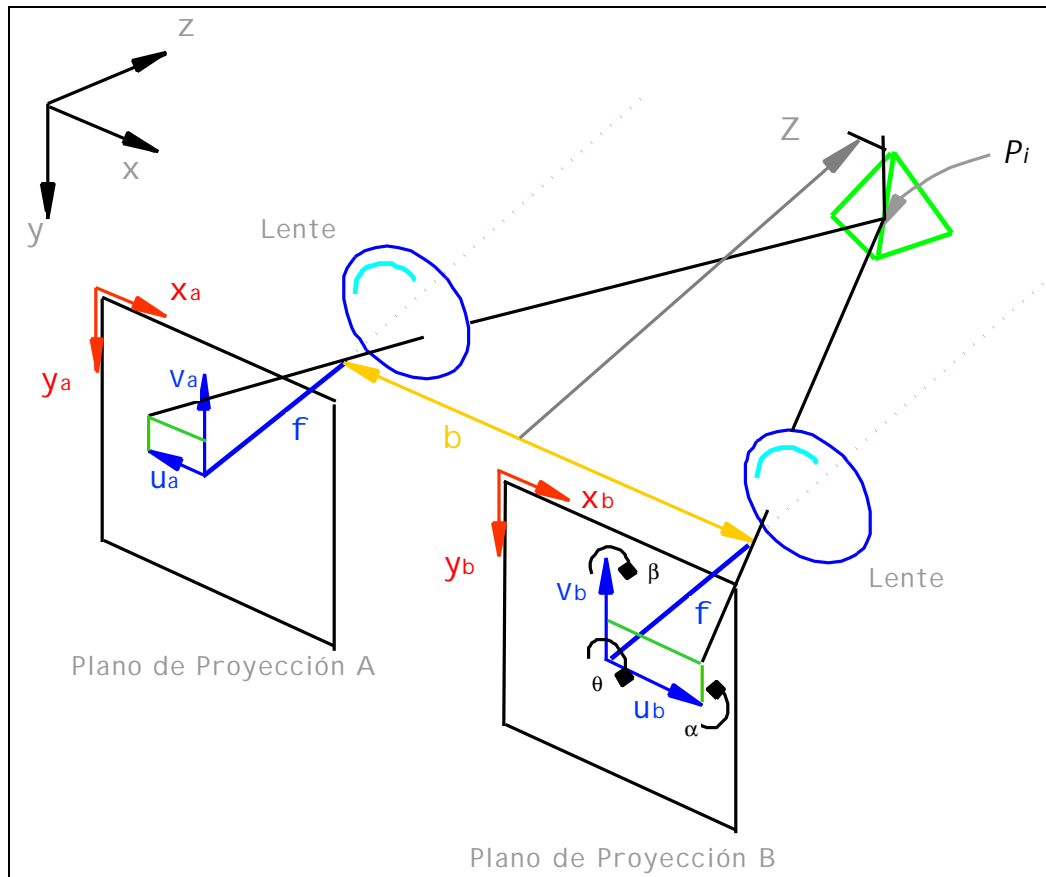


Figura 3.1: Representación geométrica del sistema binocular paralelo.

### B. Error debido a Rotación entre las cámaras ( $q$ )

Considérese la rotación de la cámara B (en torno a su eje óptico) con respecto a la cámara A de referencia en un ángulo  $\theta$ . Si la distancia real  $z_{real}$ , está dada por:

$$z_{real} = \frac{b \cdot f}{(u_b - u_a)} \quad (3.3)$$

Y la distancia medida (observada) se obtiene según:

$$z_{obs} = \frac{b \cdot f}{(\hat{u}_b - u_a)} \quad (3.4)$$

donde  $\hat{u}_b$  es la posición horizontal observada de la proyección del punto  $P_i(X,Y,Z)$  sobre la cámara B, que idealmente debería haber sido  $u_b$  en el sistema de coordenadas de la cámara.

Empleando las expresiones (3.3) y (3.4) el error absoluto y relativo están dados por las siguientes relaciones, respectivamente:

$$Eabs_z = \|z_{real} - z_{obs}\| = z_{real} \cdot \frac{(\hat{u}_b - u_b)}{(\hat{u}_b - u_a)} \quad (3.5)$$

$$Erel_z = \frac{\|z_{real} - z_{obs}\|}{z_{real}} = \frac{(\hat{u}_b - u_b)}{(\hat{u}_b - u_a)} \quad (3.6)$$

En éste caso, por la rotación en torno al eje óptico:  $\hat{u}_b = u_b \cdot \cos(\mathbf{q})$ .

Considerando las transformaciones de coordenadas del espacio 3D a las coordenadas del plano de óptico para cada cámara,

$$X_i = \frac{z}{f} \cdot u_i, \text{ con } i = a, b.$$

y recordando que la transformación entre sistemas de coordenadas es de translación en una distancia dada por el *baseline*, según:  $X_b = X_a - b$ , la ec. (3.5) sería:

$$\Delta z_q = z_{real} \cdot \frac{X_b \cdot (\cos(\mathbf{q}) - 1)}{(X_b + b) \cdot (\cos(\mathbf{q}) - 1) - b \cdot \cos(\mathbf{q})} \quad (3.7)$$



### C. Error debido a la Inclinación entre las cámaras (a)

Para el siguiente análisis se asumirá que la cámara B rota en torno a un eje paralelo a la barra que soporta las cámaras (paralela al *baseline*). El pivote se encuentra contenido en el plano formado por los ejes ópticos del sistema perfecto, como se muestra en la figura 3.2.

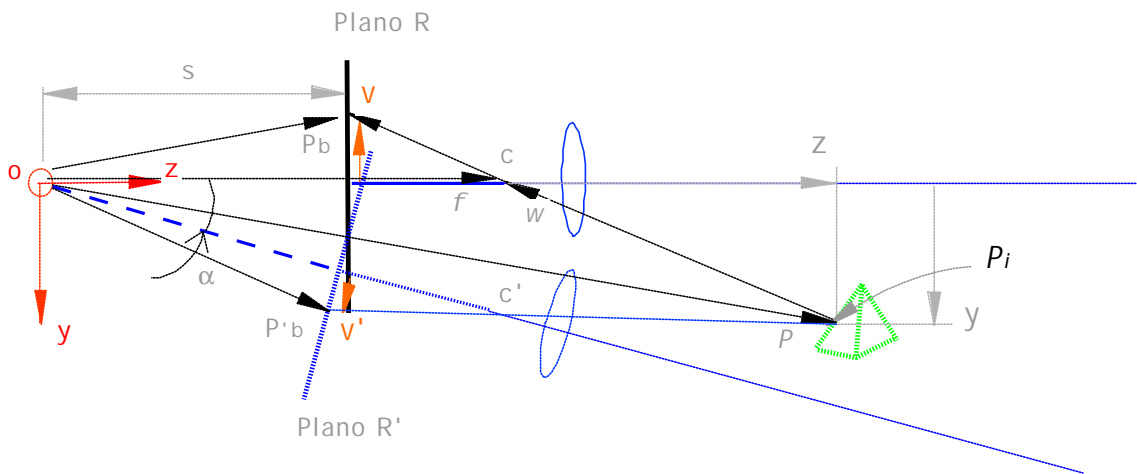


Figura 3.2: Cambio en la proyección de un punto en el espacio 3D por un movimiento de elevación (*pitch*) de la cámara.

De acuerdo al sistema de coordenadas con centro en O de la figura 3.2, la posición de la proyección del punto  $P_i(X,Y,Z)$  sobre el plano R está dada por un vector  $P_b$  de componentes:

$$P_b = (u_b, v_b, w_b)^T = \left( \frac{s \cdot X_b}{Z_b - (f + s)}, \frac{s \cdot Y_b}{Z_b - (f + s)}, s \right)^T \quad (3.8)$$

Mientras que la proyección del mismo punto  $P_i(X,Y,Z)$  sobre el plano R' está dada por un vector  $P'_b$  de componentes:

$$P'_b = (\hat{u}_b, \hat{v}_b, \hat{w}_b)^T = \begin{pmatrix} \frac{s \cdot X_b}{Y_b \sin(\mathbf{a}) + Z_b \cos(\mathbf{a}) - (f + s)} \\ \frac{s \cdot ((f + s) \sin(\mathbf{a}) - Y_b)}{Y_b \sin(\mathbf{a}) + Z_b \cos(\mathbf{a}) - (f + s)} \\ \frac{s \cdot ((f + s) \sin(\mathbf{a}) - Z_b)}{Y_b \sin(\mathbf{a}) + Z_b \cos(\mathbf{a}) - (f + s)} \end{pmatrix} \quad (3.9)$$

En este caso la expresión para el error  $\Delta z_a$  resulta bastante compleja. Es más práctico evaluar numéricamente las expresiones anteriores y reemplazar las coordenada real  $u_b$  y observada  $\hat{u}_b$  en (3.5) o (3.6), utilizando  $u_a$  a partir de (3.8). Al igual que en el caso anterior también deben considerarse las transformaciones de coordenadas:  $X_b = X_a + b$ ,  $Y_b = Y_a$  y  $Z_b = Z_a$ .

#### D. Error debido a la Convergencia entre las cámaras ( $b$ )

Considérese un sistema en que la cámara A está perfectamente alineada y la cámara B está calibrada correctamente excepto por una rotación en torno a un eje perpendicular al plano que contiene ambos ejes ópticos ( $l_a$ ,  $l_b$ ) y que pasa por el centro óptico  $c_b$ , como se muestra en la figura 3.3. De la misma manera que en los casos anteriores es posible encontrar la nueva proyección del punto  $P_i$  sobre la cámara B dado que giro en un ángulo  $\beta$ .

Del análisis geométrico se tienen las siguientes relaciones:

$$\mathbf{g} = \mathbf{f} - \mathbf{b} , \quad \hat{u}_b = f \cdot \tan(\mathbf{g}) \quad \text{y} \quad \tan(\mathbf{f}) = \frac{u_b}{f}$$

$$\text{Recordando que: } \tan(\mathbf{f} - \mathbf{b}) = \frac{\tan(\mathbf{f}) - \tan(\mathbf{b})}{1 + \tan(\mathbf{f}) \cdot \tan(\mathbf{b})}$$

se obtiene  $\hat{u}_b$  según:

$$\hat{u}_b = f \cdot \frac{u_b - f \cdot \tan(\mathbf{b})}{f + u_b \cdot \tan(\mathbf{b})} \quad (3.10)$$

La ec. (3.10) coincide con la obtenida por [Zhao95].

Con las transformaciones de coordenadas:

$$X_i = \frac{z}{f} \cdot u_i \quad \text{y} \quad X_a = X_b \cdot \cos(\mathbf{b}) - b$$

y reemplazando (3.10) en (3.5) se tiene que el error absoluto está dado por:

$$\Delta z_b = z_{real} \cdot \frac{(z_{real}^2 + X_b^2) \cdot \tan(\mathbf{b})}{[\cos(\mathbf{b}) - 1] \cdot X_b + b} \cdot z_{real} + [z_{real}^2 - X_b^2 \cdot \cos(\mathbf{b}) - X_b \cdot b] \cdot \tan(\mathbf{b}) \quad (3.11)$$

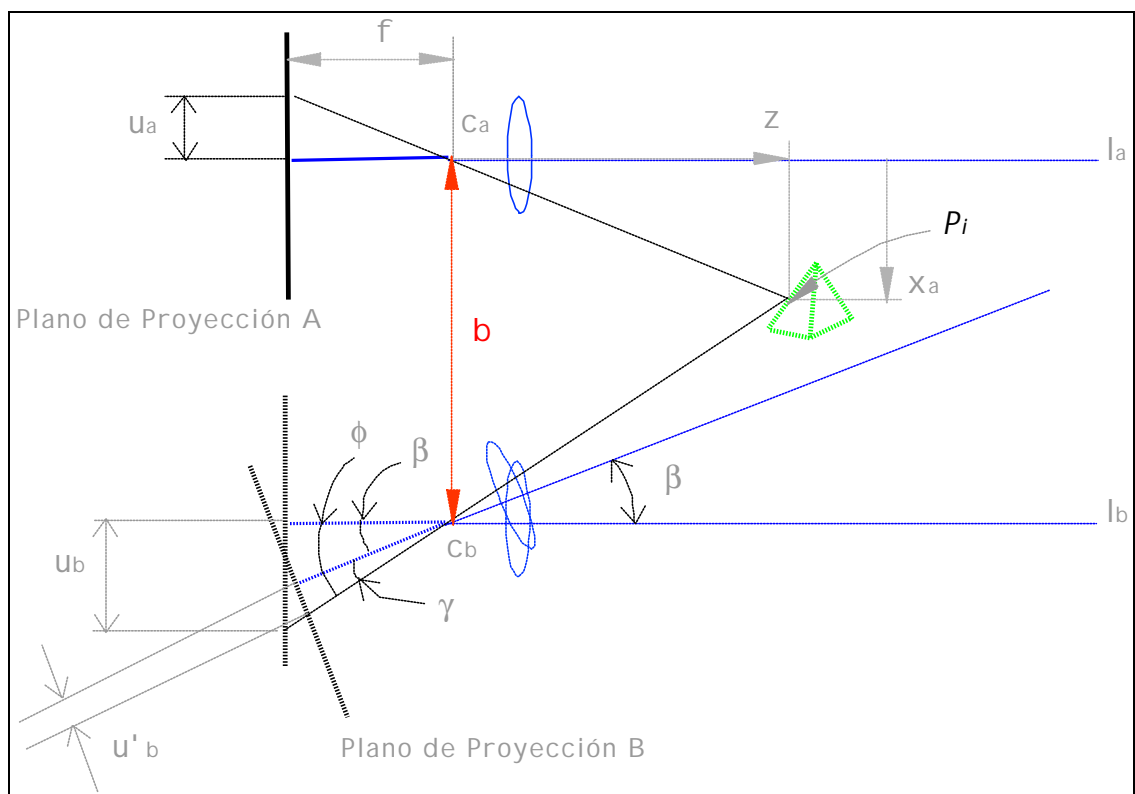


Figura 3.3: Cambio en la proyección de un punto en el espacio 3D debido a un ángulo de convergencia entre las cámaras.

### E. Errores debido a lentes (distancia focal) y CCDs no paralelos

Si el error se produce en la determinación de la distancia focal, la medición de la profundidad se verá afectada de manera similar que en el caso en que se produce un error en la medición del *baseline*. El error para esta situación sería:

$$\Delta z_f = \frac{b \cdot \Delta f}{\mathbf{r}_x \cdot \mathbf{d}} \quad (3.12)$$

Sin embargo, puede ocurrir que exista una desalineación del plano donde se forma la imagen. En este caso, considerando rotaciones del CCD como se muestra en la figura 3.4 y mediante un análisis similar al realizado en punto C, es posible encontrar la nueva posición de la proyección del punto  $P_i$  sobre el plano óptico.

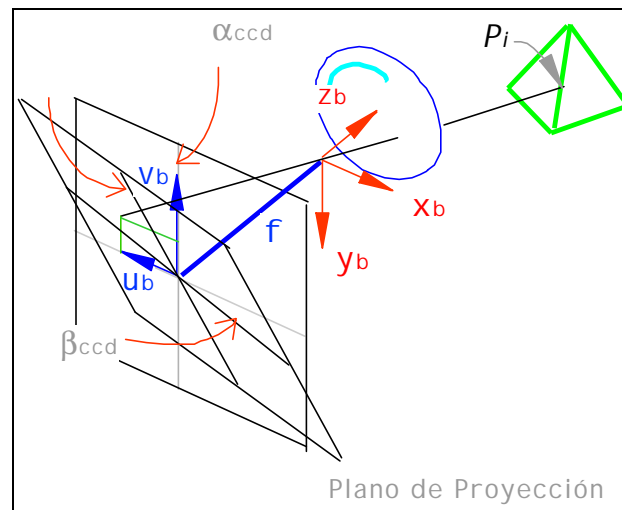


Figura 3.4: Desalineación del CCD.

Si el plano del CCD alineado tiene una normal dada por el vector  $(0,0,1)^T$  y el plano desalineado tiene una normal (debido a rotaciones a los ejes de  $\mathbf{a}$  y  $\mathbf{b}$ ) de la forma:

$$\left( \sin(\mathbf{b}_{CCD}), -\sin(\mathbf{a}_{CCD}), \cos(\mathbf{a}_{CCD}) \cdot \cos(\mathbf{b}_{CCD}) \right)^T$$

Suponiendo que ambos planos contienen el punto  $(0,0,-f)$ , entonces la coordenada horizontal observada será:

$$\hat{u}_b = \frac{X_b \cdot f}{Y_b \sin(\mathbf{a}_{CCD}) - X_b \sin(\mathbf{b}_{CCD}) + \cos(\mathbf{a}_{CCD}) \cdot \cos(\mathbf{b}_{CCD}) \cdot Z_{real}} \quad (3.13)$$

En este caso la expresión para  $\hat{u}_b$  también resulta extensa, y conviene evaluarla numéricamente utilizando la expresión (3.5) para encontrar el error  $\Delta z_{ab-CCD}$  en la medición de la profundidad.

Al igual que en el caso del cálculo del error para un ángulo de rotación ( $\mathbf{q}$ ) entre las cámaras,  $u_a$  y  $u_b$  se obtienen utilizando la relación:

$$X_i = \frac{z}{f} \cdot u_i, \text{ con } i = a, b; \text{ y considerando la transformación de coordenadas:}$$

$$X_b = X_a + b.$$

## F. Errores debido a la distorsión de los lentes

Determinar la distorsión introducida por los lentes es bastante complejo, especialmente porque los parámetros de los modelos existentes no son adquiribles de una forma simple. Para simplificar el análisis se utilizará un modelo de la distorsión como en [Zhao95] y [Stein96], quienes se basan en [Slam80]. El modelo de la distorsión radial con aproximación de segundo orden está dado por las siguientes relaciones:

$$u = \hat{u} + (\hat{u} - c_{xr}) \cdot (\mathbf{k}_1 \cdot \hat{r}^2) \quad (3.14a)$$

$$v = \hat{v} + (\hat{v} - c_{yr}) \cdot (\mathbf{k}_1 \cdot \hat{r}^2) \quad (3.14b)$$

$$\hat{r}^2 = (\hat{u} - c_{xr})^2 + (\hat{v} - c_{yr})^2 \quad (3.14c)$$

Las relaciones establecen una *mapeo* entre las coordenadas de los puntos observados (i.e. de los puntos con distorsión  $(\hat{u}, \hat{v})$ ), y las coordenadas  $(u, v)$  de los puntos de la imagen sin distorsión, los cuales no son físicamente medibles.

En el desarrollo de [Zhao95] se asume que el centro de distorsión radial ocurre en el centro de la imagen (i.e.  $(c_{xr}, c_{yr}) = (0,0)$ ). Este supuesto es válido, ya que es lo que normalmente ocurre en la práctica. Sin embargo, para simplificar la obtención del error en la medición de distancia, [Zhao95] asume que no hay distorsión vertical (i.e.  $v = \hat{v}$ ), lo cual puede ser una aproximación aceptable para lentes con una distancia focal grande (mayores que 80 mm), pero es totalmente incorrecta para lentes *gran angular*, con distancias focales en el rango medio o pequeño (5 mm - 60 mm).

Para analizar el error en la medición de la distancia introducido por la distorsión radial habría que utilizar las relaciones (3.14) en (3.5), pero por la forma no lineal de las relaciones (3.14) no es posible obtener una expresión cerrada. Una linealización de las ecuaciones (3.14) no tiene sentido para un análisis práctico ya que las coordenadas  $u$  y  $v$  toman valores en un amplio rango (típicamente entre 0-320 pixels).

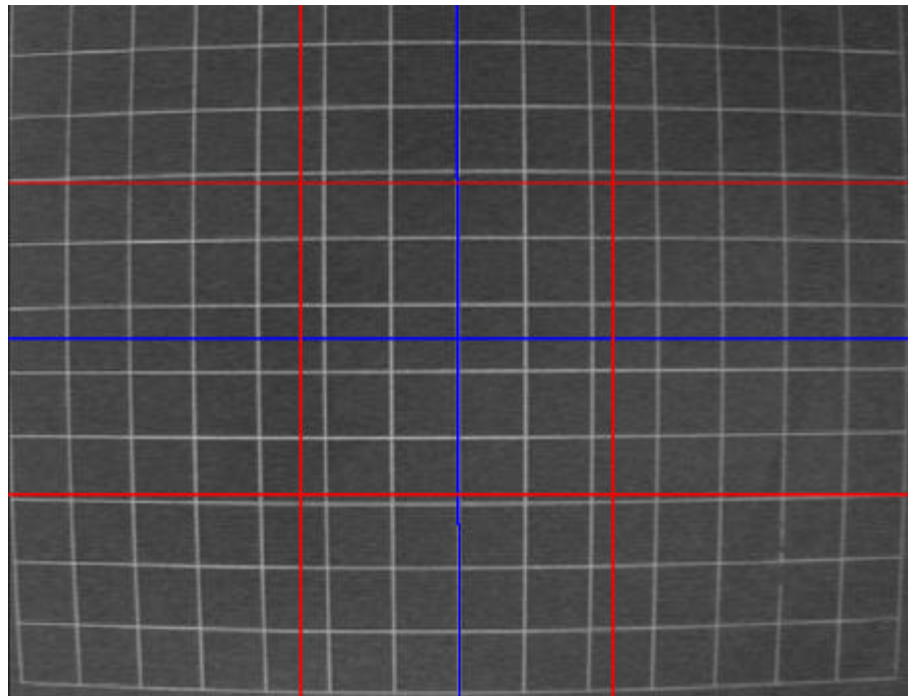


Figura 3.5: Distorsión de un lente con  $f$ : 6.5 mm.

Los efectos de la distorsión radial de un lente con distancia focal de 6.5 mm se pueden apreciar en la figura 3.5, la cual muestra la imagen de un patrón cuadriculado. Las líneas azules corresponden al centro de la imagen de 742x568 pixels. En el sector central de la imagen de 256x256 pixels, delimitado por las líneas rojas, la distorsión es casi imperceptible. El problema de la distorsión se puede corregir mediante una rectificación geométrica de la imagen llamada frecuentemente *warping* [Wolb92].

### 3.1.2 Importancia Relativa de los Errores

Basados en los resultados anteriores podemos determinar qué parámetros del sistema requieren de un especial cuidado al momento de su calibración. Una manera sencilla para determinar la influencia de cada parámetro sobre el error consiste en evaluar las expresiones derivadas dado un sistema binocular con una cierta geometría.

Considerando un sistema binocular paralelo con las siguientes características:

- **Lentes de distancia focal:** 16 mm.
- **Baseline:** 20 mm.
- **Rango de operación:** En torno a una distancia de 800 mm y desplazado 20 mm con respecto a la cámara B (figura 3.9).
- **Tolerancia:** Error absoluto no mayor a 8 mm (i.e. error relativo máximo del 1%).

Si se evalúan las expresiones (3.1-3.16), se obtiene que para cumplir con la tolerancia del 1% debe verificarse que:  $\Delta b < 2$  mm,  $\Delta q < 8.15^\circ$ ,  $\Delta a < 1.43^\circ$ ,  $\Delta b < 0.0145^\circ$ ,  $\Delta f < 1.6$  mm. Estos resultados indican que de los parámetros de alineación ( $q$ ,  $a$ ,  $b$ ), el más influyente en el error es el ángulo de convergencia de las cámaras  $b$ , seguido por el ángulo de inclinación de las cámaras  $a$ . El menos problemático debería ser la rotación entre las cámaras  $q$  (o entre los CCD, en el caso de cámaras fijas a una barra).

Estos resultados concuerdan con los que se hayan en la práctica. Un experimento simple consiste en cambiar ligeramente el ángulo de convergencia de una

de las cámaras y comparar la imagen con la que se tenía antes de mover la cámara. Los cambios son apreciables a simple vista.

En cuanto a la evaluación de los máximos errores tolerados en la desalineación del CCD con respecto al lente de la cámara ( $\Delta \mathbf{a}_{CCD}$ ,  $\Delta \mathbf{b}_{CCD}$ ), la evaluación de las expresiones indica que estos deben ser prácticamente nulos. El sistema es incluso más sensible a cambios en estos parámetros. Sin embargo, en la práctica estos parámetros son difícilmente modificables, ya que el sensor normalmente viene instalado en la cámara desde fábrica.

Dados el *baseline* y la distancia focal elegidos, el sistema es menos sensible a un error en la determinación del primero. El error en el *baseline* no debe superar los 2 mm para cumplir con el error máximo aceptado del 1%, mientras que la distancia focal debe cumplir con un margen de error no superior a los 1.6 mm. La comparación de las relaciones (3.1) y (3.12) indica que para una cierta tolerancia, existe una mayor libertad de error en el *baseline* o en la distancia focal dependiendo cual de los dos sea mayor.

### 3.1.3 Error en la Profundidad Debido a la Ocurrencia de Múltiples Errores de Calibración

En un sistema real de visión estereoscópica, es más frecuente que ocurran varios errores a la vez. Según [Zhao95] es posible calcular el efecto de la combinación de errores calculando la proyección observada  $\hat{u}_b$  y aplicando en forma secuencial las transformaciones para cada caso (rotación, elevación, convergencia). Recordando las expresiones utilizadas para la proyección observada en cada caso:

$$\text{Roll relativo:} \quad \hat{u}_b = u_b \cdot \cos(\mathbf{q}) \quad (3.15a)$$

$$\text{Pitch relativo:} \quad \hat{u}_b \approx \frac{u_b}{\cos(\mathbf{a}) - \frac{u_b}{f} \cdot \sin(\mathbf{a})} \quad (3.15b)$$

$$\text{Convergencia:} \quad \hat{u}_b = f \cdot \frac{u_b - f \cdot \tan(\mathbf{b})}{f + u_b \cdot \tan(\mathbf{b})} \quad (3.15c)$$



Nótese que (3.15b) es una aproximación de (3.9), bajo dos supuestos:

1. El eje de giro del ángulo  $\mathbf{a}$  pasa por el centro óptico ( $s = f$ ).
2. La escena observada es cuadrada ( $Y_b = X_b$ ), y por lo tanto  $v_b = u_b$ .

Reemplazando (3.15a) en (3.15b) y la relación resultante en (3.15c) se obtiene:

$$\hat{u}_b = f \cdot \frac{\left( \frac{u_b \cdot \cos(\mathbf{q}) \cdot f}{\cos(\mathbf{a}) \cdot f + u_b \cdot \cos(\mathbf{q}) \cdot \cos(\mathbf{a})} \right) - f \cdot \tan(\mathbf{b})}{f + \left( \frac{u_b \cdot \cos(\mathbf{q}) \cdot f}{\cos(\mathbf{a}) \cdot f + u_b \cdot \cos(\mathbf{q}) \cdot \cos(\mathbf{a})} \right) \cdot \tan(\mathbf{b})} \quad (3.16)$$

La expresión (3.16) debe evaluarse en (3.5) o (3.6) para obtener el error combinado en la estimación de la profundidad. Si  $u_b = \frac{f \cdot X_b}{z}$  y  $u_a = \frac{f \cdot (X_b + b)}{z}$ , entonces dado un error relativo  $e$ , debe resolverse la siguiente ecuación, para  $b$ ,  $Z$  y  $X_b$  dados:

$$f \cdot \frac{\left( \frac{\cos(\mathbf{q}) \cdot f \cdot X_b}{\cos(\mathbf{a}) \cdot Z + \cos(\mathbf{q}) \cdot \cos(\mathbf{a}) \cdot X_b} \right) - f \cdot \tan(\mathbf{b})}{f + \left( \frac{\cos(\mathbf{q}) \cdot f \cdot X_b}{\cos(\mathbf{a}) \cdot Z + \cos(\mathbf{q}) \cdot \cos(\mathbf{a}) \cdot X_b} \right) \cdot \tan(\mathbf{b})} = \frac{f \cdot X_b}{Z} - \frac{e}{1 - e} \cdot b \quad (3.17)$$

De (3.17) se pueden obtener varias combinaciones de  $\mathbf{a}$ ,  $\mathbf{b}$ , y  $\mathbf{q}$  que permitan reducir el error relativo máximo a la tolerancia especificada  $e$  para una cierta distancia  $Z$ . El problema no tiene solución única y es no lineal.

En las figuras 3.6a y 3.6b se muestran superficies que representan el error relativo, obtenido al estimar la posición de un punto que se encuentra a 800 mm de las cámaras, y desplazado 20 mm a la derecha de la cámara B. En la figura 3.6a se muestra el error relativo manteniendo la variable de rotación entre las cámaras constante e igual a  $0^\circ$ ; mientras que en la figura 3.6b, el ángulo de convergencia se mantiene fijo en  $0^\circ$ .

La figura 3.6a muestra claramente que el error relativo en la estimación de  $z$  es más sensible a los errores en la convergencia que a los errores en la elevación. La superficie es prácticamente plana en el sentido del parámetro de elevación (i.e. no hay variaciones apreciables del error frente a variaciones de  $\mathbf{a}$ ).

En cambio la figura 3.6b, permite comprobar que los efectos que tienen las desviaciones del ángulo de elevación relativa y de rotación relativa con respecto a la posición ideal ( $\mathbf{a}$  y  $\mathbf{q}$  iguales a cero), contribuyen aproximadamente del mismo modo al error en la medición de la distancia  $z$ .

Los resultados obtenidos por [Zhao95] son análogos, aunque sus expresiones difieren principalmente por las aproximaciones que realiza.

Hay que tener presente que se supuso que los algoritmos de búsqueda de correspondencias son capaces de encontrar todos los puntos correspondientes de ambas imágenes; sin embargo, la búsqueda de correspondencias normalmente se hace en un rango restringido, en el sentido de las líneas horizontales del CCD, para acelerar el proceso y lograr *tiempos prácticos*. El supuesto de que las parejas de puntos correspondientes se encuentran sobre una misma línea horizontal impone severas restricciones, especialmente sobre el grado de error tolerable en la alineación del ángulo de elevación  $\mathbf{a}$ .

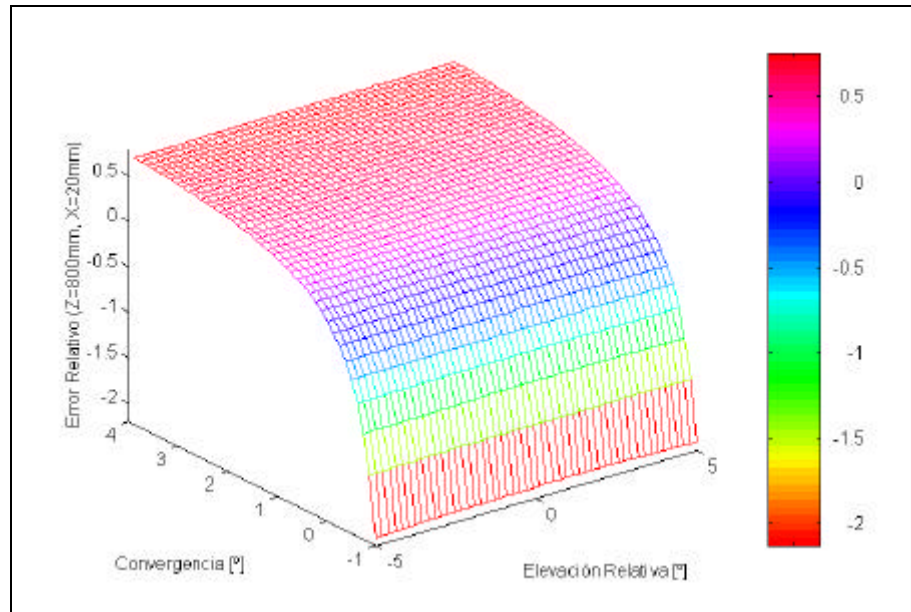


Figura 3.6a: Error relativo en función de la convergencia y la elevación. (Parámetros del Sistema:  $b= 20\text{mm}$ ,  $f= 16\text{mm}$ ,  $Z= 800\text{ mm}$ ,  $X= 20\text{mm}$ ,  $q= 0^\circ$ ).

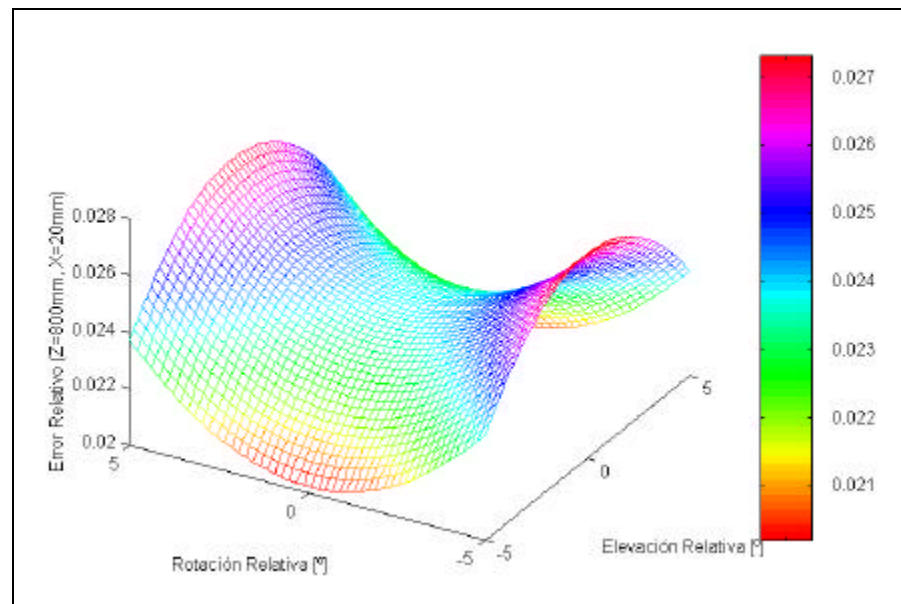


Figura 3.6b: Error relativo en función de la rotación y la elevación. (Parámetros del Sistema:  $b= 20\text{mm}$ ,  $f= 16\text{mm}$ ,  $Z= 800\text{ mm}$ ,  $X= 20\text{mm}$ ,  $b= 0^\circ$ ).

### 3.1.4 Efectos de los Errores de Muestreo

El análisis anterior consideraba que la imagen podía registrarse como una señal bidimensional continua, pero si se utilizan cámaras CCD, la realidad proyectada se discretiza *espacialmente* con un cierto número de niveles de cuantización, introduciendo errores en la determinación de la disparidad real. Algunos autores asocian el término error de cuantización con el concepto de error en las disparidades debido a la discretización. Sin embargo, es más apropiado considerar que el error de muestreo (*sampling*) produce errores en la determinación de la disparidad.

Recordando que el error absoluto y el relativo están dados por las siguientes expresiones:

$$Eabs_z = \|z_{real} - z_{obs}\| = \Delta z \qquad Erel_z = \frac{\|z_{real} - z_{obs}\|}{z_{real}} = \frac{\Delta z}{z}$$

Puede determinarse el error  $\Delta z_d$  que se produce al estimar  $z$  a partir de una medición incorrecta de la disparidad  $d$  como:

$$\Delta z_d = \left\| \frac{\mathbf{h}}{d} - \frac{\mathbf{h}}{d \pm \Delta d} \right\| = \mathbf{h} \cdot \left\| \frac{\pm \Delta d}{d^2 \pm \Delta d \cdot d} \right\| \quad (3.18)$$

$$\frac{\Delta z_d}{z} = \left\| \frac{\pm \Delta d}{d \pm \Delta d} \right\| \approx \frac{1}{d} \quad (3.19)$$

$$\text{Donde } \mathbf{h} = \frac{b \cdot f}{\mathbf{r}_x} \quad (3.20)$$

La relación (3.19) muestra que el error relativo en la estimación de profundidades es inversamente proporcional a la disparidad. Esta aproximación es válida con un error inferior al 5% para disparidades superiores a 5.0 unidades y con un error menor al 1% para disparidades superiores a 10.5 unidades, si el error en la disparidad  $\Delta d$  máximo es 1.

Si se asume que los errores en la búsqueda de correspondencias son sólo producto del muestreo ruidoso de la imagen, entonces  $\Delta d$  no debería sobrepasar 1 unidad, ya que el sensor puede registrar la imagen con una precisión de  $\pm 0.5$  pixels. En la práctica ocurren errores en la búsqueda de correspondencias que son superiores a 1 unidad, debido a que las curvas de correlación no tienen un máximo único cuando las escenas tienen cierta textura. En el presente análisis se supone que los errores de correlación son nulos y que sólo hay errores de muestreo, por lo tanto el supuesto de la aproximación es válido. Pero en un análisis más general no sería correcto utilizar la aproximación de (3.19).

Utilizando la aproximación del error relativo (3.19), éste puede expresarse en función de la distancia como:

$$\Delta z_d \% \approx \left(\frac{100}{\mathbf{h}}\right) \cdot \mathbf{h} \cdot \frac{1}{d} = \frac{100}{\mathbf{h}} \cdot z \quad (3.21)$$

La expresión (3.21) indica que el error relativo es aproximadamente proporcional a la distancia.

Otra conclusión importante que puede obtenerse de (3.21) es que la pendiente del error relativo puede controlarse mediante variaciones de  $\mathbf{h}$ , que depende de la longitud del *baseline*, de la distancia focal y del ancho horizontal de un pixel. Si el sistema es diseñado para cumplir una cierta especificación a la distancia máxima, entonces el error siempre será menor a otras distancias porque el error es máximo a la distancia máxima.

Utilizando (3.21) puede obtenerse una aproximación para el error absoluto en función de la distancia como:

$$\Delta z_d = \frac{\Delta z_d \%}{100} \cdot z \approx \frac{z^2}{\mathbf{h}} \quad (3.22)$$

Estos resultados pueden comprenderse mejor con ayuda de la figuras 3.7 y 3.8. En la figura 3.7 se muestra el error absoluto en la estimación de  $z$  ( $E_{abs.z}$ ) para dos niveles de profundidad. En ambos casos el error en la medición de la disparidad es de  $\pm 0.5$  pixels. Sin embargo, dada la forma de la curva

*Distancia z vs. Disparidad* (obtenida a partir de la relación (2.16), para cierto *baseline* y distancia focal), puede apreciarse que el error absoluto es mayor para distancias grandes (disparidades bajas:  $d_0$ ), y menor al estimar la distancia a la que se encuentran los objetos cercanos (disparidades altas:  $d_1$ ).

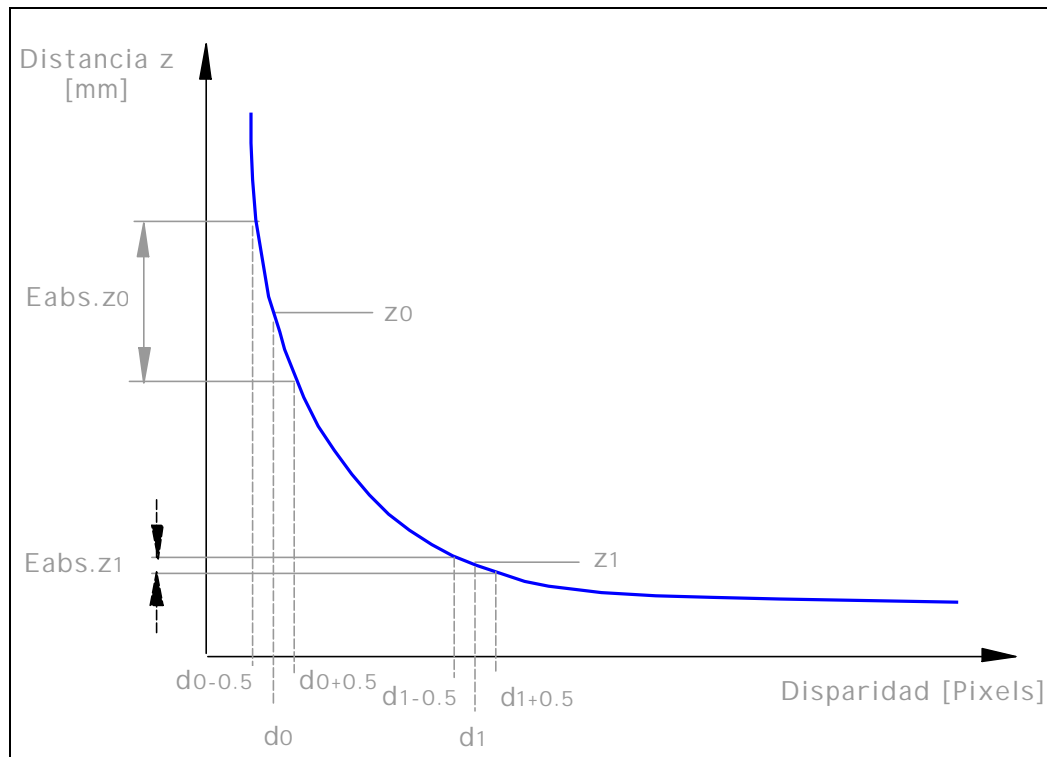


Figura 3.7: Error absoluto debido a un error de  $\pm 0.5$  pixels en la disparidad.

En la figura 3.8 se muestra una familia de curvas *Distancia z vs. Disparidad* obtenidas variando el *baseline*, para una distancia focal fija de 16 mm, y una longitud del pixel de 0.01 [mm/pixel]. La curva inferior corresponde a un *baseline* de 10 mm, mientras que para las curvas superiores el *baseline* se incrementa en pasos de 20 mm hasta un *baseline* de 150 mm, equivalente a un desplazamiento de 15 cm entre los centros ópticos de cada cámara.

Derivando la relación (2.16) se puede obtener la pendiente de las curvas como:

$$\frac{\mathcal{I}z}{\mathcal{I}d} = \frac{-\mathbf{h}}{d^2},$$

pero utilizando (2.16) la pendiente puede expresarse como:

$$\frac{\mathcal{I}z}{\mathcal{I}d} = \frac{-z^2}{\mathbf{h}} = \frac{-z}{d} \quad (3.23)$$

Ahora se puede expresar el error absoluto en función de la pendiente de la curva utilizando (3.22) y la relación anterior como:

$$\Delta z_d \approx -\frac{\mathcal{I}z}{\mathcal{I}d} \quad (3.24)$$

La expresión para el error absoluto en (3.24) también puede obtenerse aproximando directamente la relación (3.18) para  $\Delta d = 1[\text{pixel}] \ll d$ . Por (3.24) puede afirmarse que el error absoluto en la medición de profundidades es directamente proporcional a la magnitud de la pendiente de la curva *Distancia-Disparidad*. A mayor profundidad, mayor es la pendiente, y por consiguiente mayor es el error como era de esperarse por la aproximación de (3.22). En la figura 3.8 se muestra el crecimiento cuadrático del error absoluto en función de la distancia para un sistema binocular con parámetros  $b = 20$  mm,  $f = 16$  y  $r_x = 0.01$  [mm/pixel].

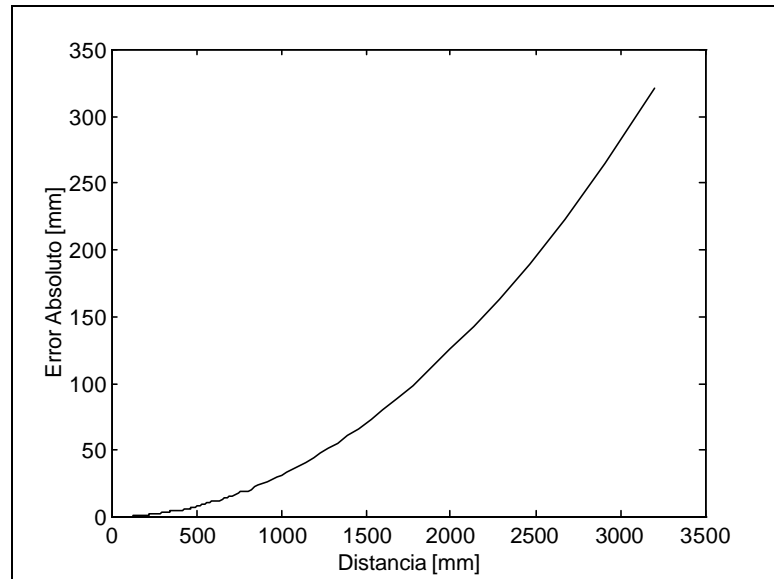


Figura 3.8: Error absoluto en la estimación de la distancia vs. distancia observada (Parámetros del Sistema:  $b = 20$  mm,  $f = 16$  y  $r_x = 0.01$  [mm/pixel]).

Utilizando el término de la derecha de (3.23), que expresa la pendiente de la curvas en función de la distancia  $z$  y de la disparidad  $d$ , se puede obtener el lugar geométrico de los puntos en la familia de curvas *Distancia-Disparidad* que tienen una misma pendiente (i.e. un mismo error absoluto).

El lugar geométrico que resulta al unir los puntos de igual pendiente (i.e. mismo error absoluto) de las curvas *Distancia-Disparidad* está representado por la familia de rectas  $z = k \cdot d$  (obtenidas de (3.23)) como se muestra en la figura 3.9 con líneas azules. Mientras mayor es la pendiente de estas rectas, mayor es el error absoluto asociado a la medición en el punto donde la recta interseca la curva *Distancia-Disparidad*.

En términos prácticos, el gráfico 3.9 establece las relaciones entre el nivel de error máximo aceptado (definido por alguna de las rectas azules), el *baseline* y las disparidades mínimas que deberían utilizarse para operar el sistema binocular a una distancia dada, cumpliendo la tolerancia exigida.



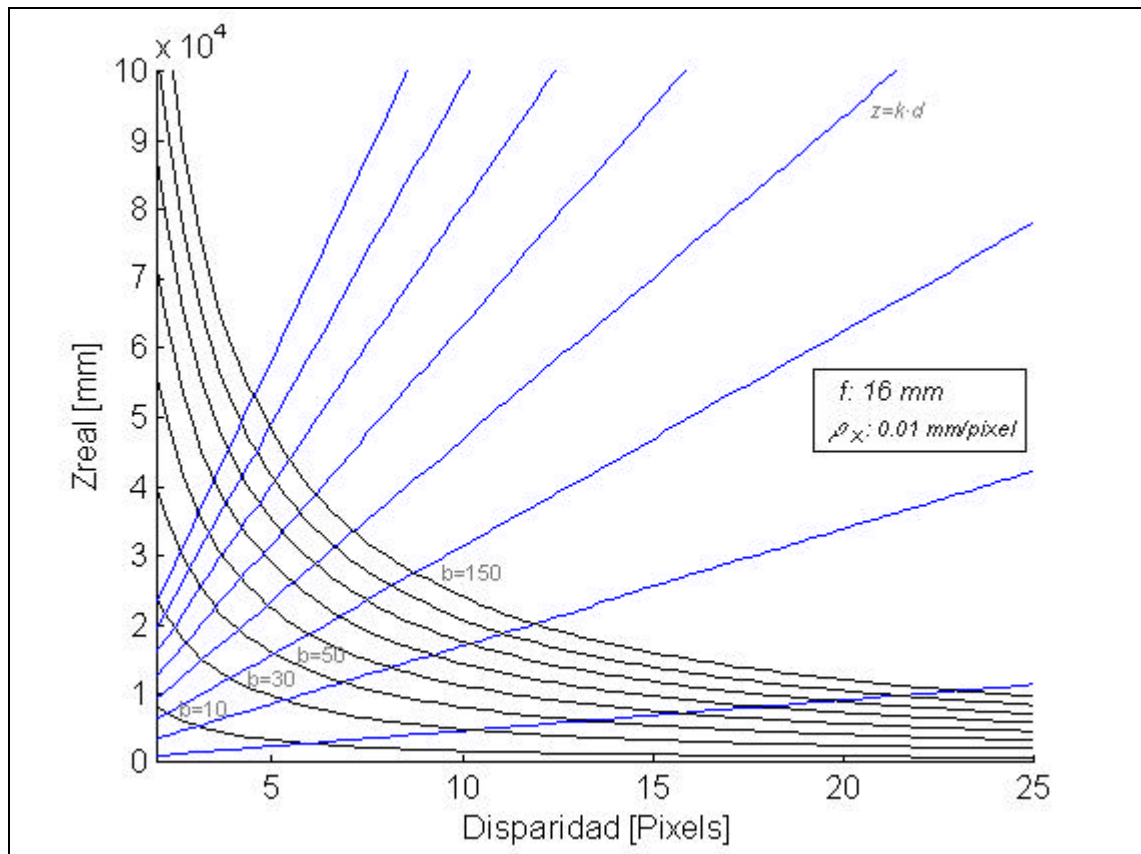


Figura 3.9: Familia de curvas de disparidad variando el *baseline*.

Una familia similar de curvas puede obtenerse variando el tamaño horizontal de la celda del CCD ( $r_x$ ), o la distancia focal. Sin embargo, en un sistema real,  $r_x$  depende de las especificaciones del fabricante de las cámaras, por lo que no tiene mayor sentido el análisis para variaciones de este parámetro. En la práctica tampoco es muy conveniente variar la distancia focal. Aunque se usen lentes con *zoom* controlable eléctricamente, el rango dinámico es más reducido (i.e. las familia de curvas que se puede generar no es tan grande).

Lo anterior concuerda con lo apreciado en las pruebas realizadas, en las que la determinación de los puntos del espacio muy lejanos del observador (i.e. puntos con muy baja disparidad) son medidos con un menor grado de certeza.

### 3.1.5 Efectos del Movimiento de la Cámaras

En esta sección se analiza el efecto que tiene un movimiento del sistema binocular en el sentido de los ejes ópticos (i.e. movimiento en el eje perpendicular al plano de la escena). Se asume que el sistema ha sido configurado para observar un objeto de tamaño  $s$  (rango de profundidades de la escena), a una distancia máxima inicial  $z_{max0}$ , y que el sistema se acerca al objeto en una cantidad  $movz$ . Debido al desplazamiento,  $z_{max0}$  se acorta a  $z_{max1}$ .

El análisis que se presenta en esta sección es importante para el procesamiento de secuencias estereoscópicas que utilicen los resultados del *matching* en forma recursiva con el fin de reajustar el rango de búsqueda de disparidades. No se ha encontrado en la literatura sobre estereoscopía un análisis similar.

En la sección 3.1.4 se demostró que el error en la estimación de la profundidad disminuye al restringir la distancia máxima de operación del sistema a valores menores. Sin embargo, como se muestra en la figura 3.10, al acercar el sistema binocular a la escena que se desea reconstruir tridimensionalmente, el rango de disparidades crece, pudiendo tener efectos negativos sobre el proceso de búsqueda de correspondencias, ya que éste se hace más inexacto por la mayor cantidad de ambigüedades que pueden generarse.

Por otro lado, el sistema no puede acercarse demasiado al área de interés en la escena, sino la disparidad máxima aumentará a valores que estarán fuera del área definida por las dimensiones del CCD. Como es lógico, al acercarse a la escena habrán objetos que desaparecerán del área de visión que inicialmente tenía el sistema binocular.

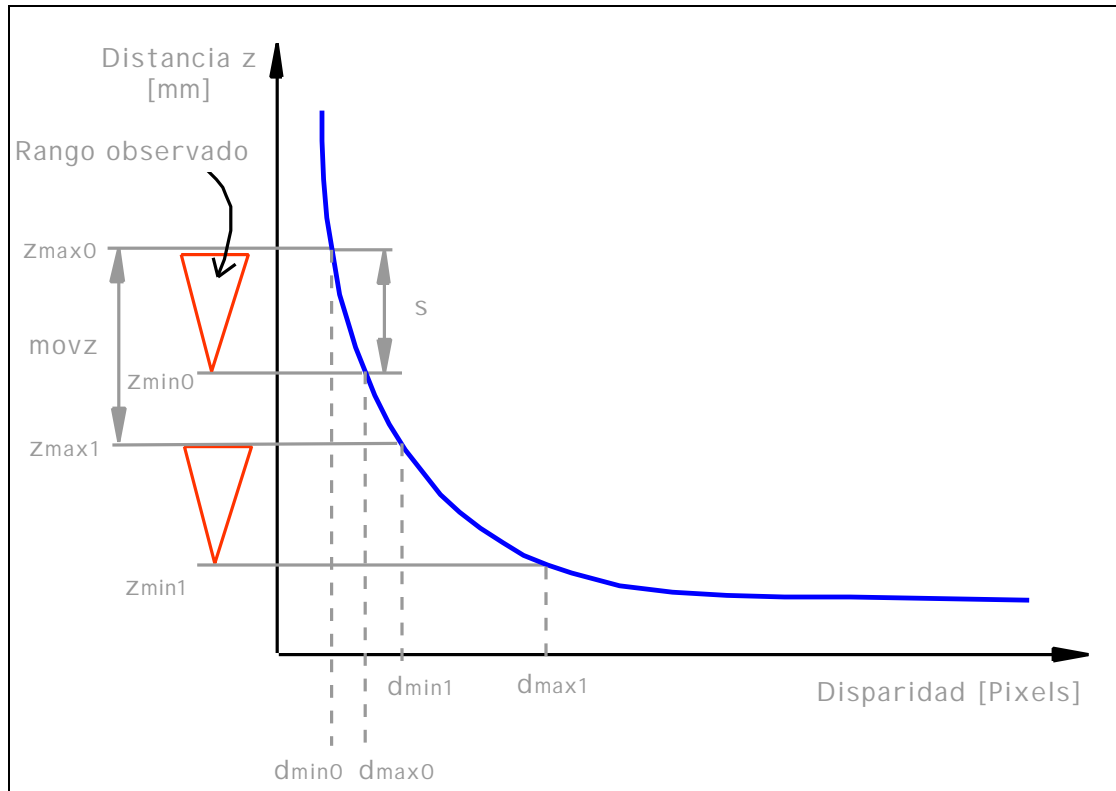


Figura 3.10: Variación del rango de disparidades de la escena para un desplazamiento en  $z$ .

Dados  $z_{max0}$ ,  $s$ ,  $movz$ , y considerando las siguientes relaciones (ver figura 3.10):

$$z_{min0} = z_{max0} - s \quad z_{min1} = z_{max1} - s \quad z_{max1} = z_{max0} - movz \quad (3.25)$$

Puede obtenerse el rango de disparidades inicial como:

$$rd_0 = d_{max0} - d_{min0} = h \cdot \left( \frac{1}{z_{max0} - s} - \frac{1}{z_{max0}} \right) = \frac{h \cdot s}{(z_{max0} - s) \cdot z_{max0}} \quad (3.26a)$$

y el rango de disparidades después del desplazamiento:

$$\begin{aligned}
 rd_1 &= d_{max1} - d_{min1} = \mathbf{h} \cdot \left( \frac{1}{z_{max1} - s} - \frac{1}{z_{max1}} \right) \\
 rd_1 &= \mathbf{h} \cdot \left( \frac{1}{z_{max0} - movz - s} - \frac{1}{z_{max0} - movz} \right) \\
 rd_1 &= \frac{\mathbf{h} \cdot s}{(z_{max0} - movz - s) \cdot (z_{max0} - movz)}
 \end{aligned} \tag{3.26b}$$

La variación del rango de disparidades para un desplazamiento  $movz$  del sistema será entonces:

$$rd_1 - rd_0 = \frac{\mathbf{h} \cdot s \cdot ((2 \cdot z_{max0} - s - movz) \cdot movz)}{(z_{max0} - movz - s) \cdot (z_{max0} - movz) \cdot (z_{max0} - s) \cdot z_{max0}} \tag{3.27}$$

Las figuras 3.11a y 3.11b ilustran las expresiones desarrolladas en esta sección. La figura 3.11a muestra la variación de la disparidad mínima (asociada a los puntos de los objetos más cercanos en la escena) y la variación de la disparidad máxima, con las curvas roja y azul, respectivamente. Estas curvas ilustran la variación del rango de disparidades y fueron obtenidas suponiendo que el sistema binocular (de parámetros  $b = 20$  mm,  $f = 16$  y  $r_x = 0.01$  [mm/pixel]) se acerca a una escena que tiene una diferencia entre la profundidad máxima y la mínima de 150 mm, partiendo de una distancia máxima inicial de 700 mm y finalizando el movimiento a una distancia de 300 mm del punto más lejano en la escena.

La diferencia entre la curva de disparidad máxima y la de disparidad mínima para cada posición en  $z$  del sistema binocular refleja la variación del rango de disparidades requeridos para estimar la geometría tridimensional, por ejemplo, de un objeto (de no más de 150 mm en el sentido de  $z$ ) que se acerca a las cámaras desde una posición inicial de su parte más alejada de las cámaras de 700 mm a una posición de 300 mm.

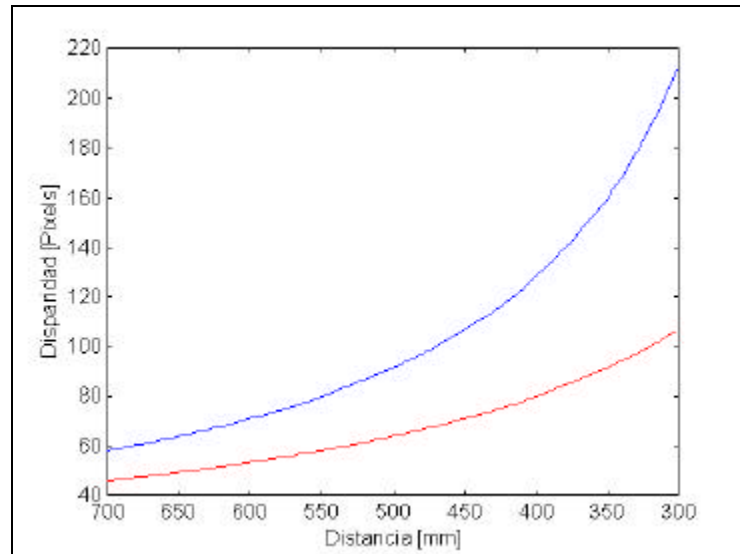


Figura 3.11a: Variación de la disparidad máxima y la disparidad mínima al acercar el sistema binocular a la escena. (Parámetros del Sistema:  $b=20\text{mm}$ ,  $f=16\text{mm}$  y  $r_x = 0.01[\text{mm/pixel}]$ ,  $z_{max}-z_{min}=150\text{mm}$ ).

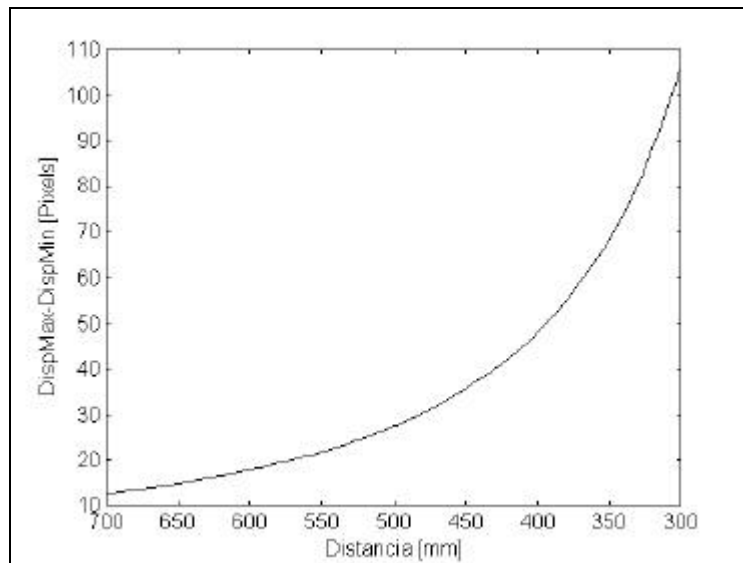


Figura 3.11b: Crecimiento del rango *Disparidad Máxima - Disparidad Mínima* al acercar el sistema binocular a la escena. (Parámetros del Sistema:  $b=20\text{mm}$ ,  $f=16\text{mm}$  y  $r_x = 0.01[\text{mm/pixel}]$ ,  $z_{max}-z_{min}=150\text{mm}$ ).

Existe un tipo de error en la medición del desplazamiento que no afecta el proceso de reconstrucción 3D y de medición directamente, sino el proceso inverso, de obtener mapas de disparidades a partir de la distancia  $z$ . Este proceso es menos frecuente, pero de todos modos es importante analizarlo, ya que puede darse el caso en que se requiera comparar mapas de disparidades para una misma escena registrada por el sistema estéreo desde posiciones  $z$  distintas.

A partir de un mapa de disparidades  $d(x,y)$  obtenido a una distancia  $z_0$  de referencia pueden conocerse las distancias a los puntos  $(x,y)$  como:

$$z_0(x, y) = \frac{\mathbf{h}}{d_0(x, y)} \quad (3.28)$$

Si luego se desplaza el sistema binocular en una cantidad igual a  $movz$ , en el sentido del eje  $z$ , puede pronosticarse el nuevo mapa de disparidades dado que conocíamos las distancias  $z_0(x,y)$  desde la posición anterior, utilizando las siguientes expresiones:

$$\hat{z}_1(x, y) = z_0(x, y) + movz \quad (3.29)$$

$$\hat{d}_1(x, y) = \frac{\mathbf{h}}{z_0(x, y) + movz} \quad (3.30)$$

Sin embargo, si se comete un error  $\Delta movz$  al estimar el desplazamiento, la estimación tendrá un error dado por:

$$\begin{aligned} \Delta \hat{d}_1(x, y) &= \frac{\mathbf{h}}{z_1(x, y)} - \frac{\mathbf{h}}{z_1(x, y) + \Delta movz} \\ \Delta \hat{d}_1(x, y) &= \frac{\mathbf{h} \cdot \Delta movz}{z_1(x, y)^2 + z_1(x, y) \cdot \Delta movz} \end{aligned} \quad (3.31)$$

Por otro lado, si el desplazamiento fue estimado correctamente, pero no el *baseline*, el error en el pronóstico de la disparidad será:

$$\Delta \hat{d}_{1B}(x, y) = \frac{\mathbf{h} \cdot \frac{\Delta b}{b}}{z_1(x, y)} \quad (3.32)$$

Dividiendo (3.32) por (3.31) puede establecerse una relación entre los errores en el pronóstico de la disparidad debido a errores la medición del *baseline* o del desplazamiento en  $z$  según:

$$\frac{\Delta \hat{d}_{1B}(x, y)}{\Delta \hat{d}_{1movz}(x, y)} = \frac{\Delta b}{b} \cdot \frac{(z_1(x, y) + \Delta movz)}{\Delta movz} \quad (3.33)$$

Si el error en medir el *baseline*  $\Delta b$  es igual al error  $\Delta movz$  de medición del desplazamiento, entonces (3.30) se puede expresar como:

$$\begin{aligned} \Delta \hat{d}_{1B}(x, y) &= \frac{(z_1(x, y) + \Delta movz)}{b} \cdot \Delta \hat{d}_{1movz}(x, y) \\ \Delta \hat{d}_{1B}(x, y) &\approx \frac{z_1(x, y)}{b} \cdot \Delta \hat{d}_{1movz}(x, y) \end{aligned} \quad (3.34)$$

En la práctica como  $z_1 \gg b$  se puede afirmar por (3.34) que el error en el pronóstico del nuevo mapa de disparidades introducido por un *error de medición del baseline* es varias veces el error en el mapa de disparidades introducido por un *error en la medición del desplazamiento* equivalente al *error de medición del baseline*. En otras palabras, un error en la medición del *baseline* tiene más efectos negativos que un error en la medición del desplazamiento sobre el pronóstico de las nuevas disparidades.

### 3.1.6 Elección de Parámetros para el Sistema Estéreo

En [McVe82] se establece un procedimiento de diseño para el sistema estéreo considerando:

1. Máximo error aceptable para la medición de distancia  $\Delta z_d$ .
2. Rango de distancias observadas  $z_{min} - z_{max}$ .
3. Máximo Alto/Ancho de la escena observada a la distancia mínima.

Tomando como referencia la figura 3.12, la cual muestra el área común a los dos cámaras (área encerrada por el triángulo azul), se pueden determinar los parámetros del sistema mediante los siguientes pasos [McVe82]:

**Paso 1: Establecer las especificaciones del sistema.**

Dentro de este paso se pueden diferenciar tres etapas:

- A. Establecer el rango  $[Z_{min}, Z_{max}]$  de distancias en que operará el sistema.
- B. Establecer el error relativo máximo  $(\Delta z_d \%_{max})$  aceptable a la distancia máxima:

Por (3.21) sabemos que el error  $\Delta z_d \%$  será menor a distancias menores. Sin embargo, un mejor índice para evaluar la exactitud del sistema estéreo es el *error relativo al rango* [Rodr90] definido como:

$$e = \frac{\Delta z}{z_{max} - z_{min}} \quad (3.35)$$

La expresión (3.35) se convierte en (3.19) si  $z_{min}$  es cero, pero en la práctica las distancias observadas corresponden a un rango reducido, en que  $z_{max} \gg z_{max} - z_{min}$ . Con (3.35) se describe mejor la resolución del rango que con  $\frac{\Delta z}{z}$ .

Por ejemplo, considérese un sistema de aerofotogrametría, posicionado a 1000 m sobre una ciudad bastante plana. Supóngase que las edificaciones más altas tienen 100 m en el sentido vertical, luego  $z_{min}$  será 900 m (los puntos más cercanos a las cámaras estarán a 900 m de éstas) y las calles estarán a  $z_{max} = 1000$  m. En estas circunstancias un error absoluto de 10 m, correspondería a un error relativo de 1%, pero a un error relativo al rango  $e$  del 10%. Para la mayoría de las aplicaciones de reconstrucción 3D el error relativo al rango es una cantidad más descriptiva porque los objetos usualmente ocupan una pequeña fracción del rango total. En otras aplicaciones, como las de navegación autónoma,



donde sólo se requiere la estimación de distancias pero no es necesaria una reconstrucción tridimensional, usar el error relativo al rango no tiene ventajas notables sobre el error relativo simple de (3.19). En los casos en que lo que interesa es determinar distancias de colisión, normalmente  $z_{min} \approx 0$ , por lo que  $z_{max} \approx z_{max} - z_{min}$ .

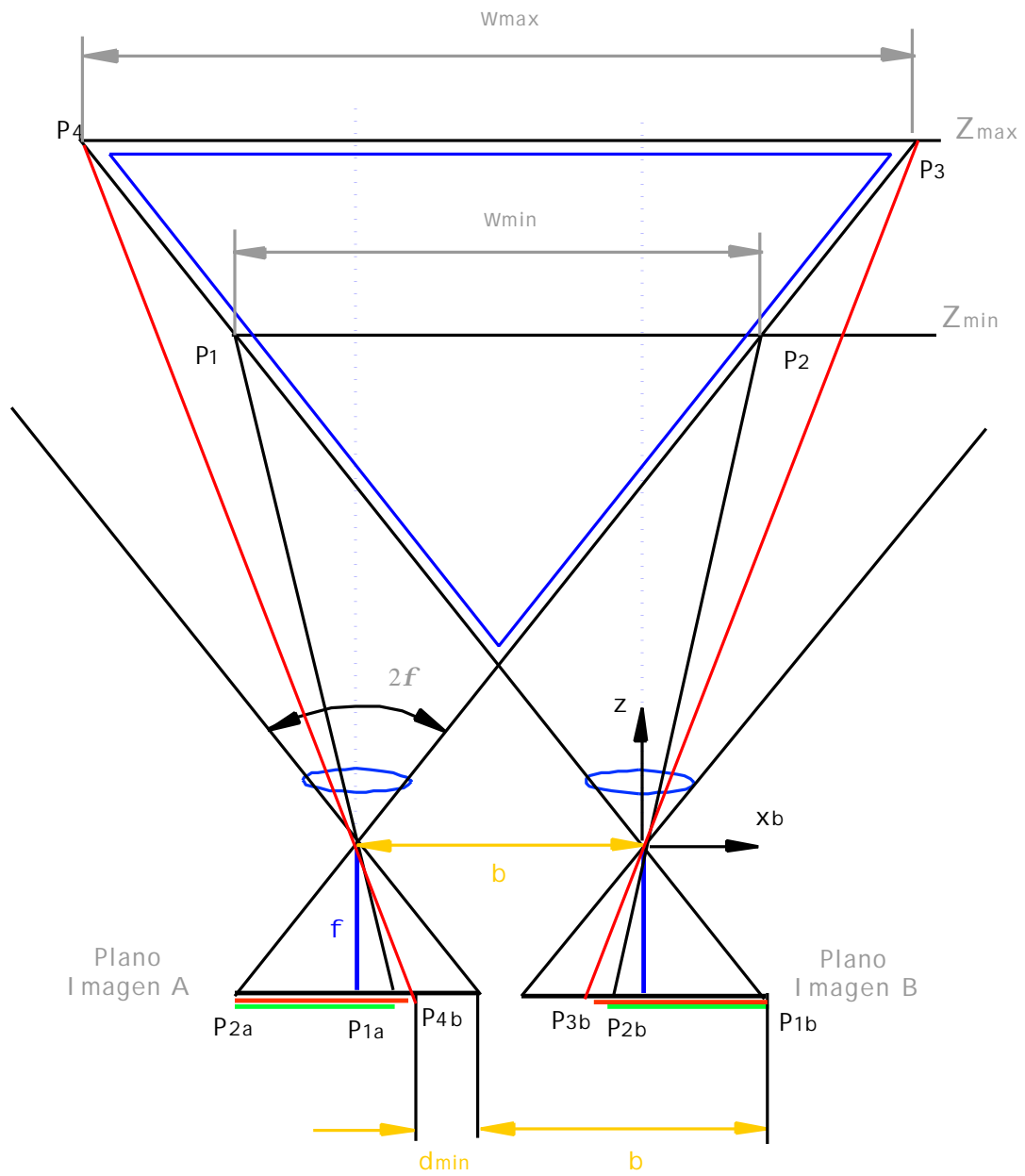


Figura 3.12: Area observable a través del sistema binocular.

De acuerdo al análisis estocástico del error debido al muestreo por [Rodr90], en el que se asume que  $z$  distribuye uniforme  $U \sim [z_{min}, z_{max}]$ , una expresión aproximada para el valor esperado de  $\mathbf{e}$ , sería:

$$E[\mathbf{e}] \approx \frac{E[\Delta z]}{z_{max} - z_{min}} = \frac{\mathbf{r}_x \cdot (z_{min}^2 + z_{min}z_{max} + z_{max}^2)}{9b \cdot f \cdot (z_{max} - z_{min})} \quad (3.36)$$

El *error relativo al rango esperado* es proporcional a la distancia interpixel (distancia entre los centros de los pixels), e inversamente proporcional al *baseline* y a la distancia focal. Este resultado es similar al derivado en (3.21), recordando la relación (3.20).

Habiendo elegido un rango de operación  $[z_{min}, z_{max}]$  y considerando una cierta tolerancia  $E[\mathbf{e}]$  según (3.36), es posible determinar por ejemplo  $b$ , dados  $f$  y  $\mathbf{r}_x$  fijos.

C. Establecer el Ancho y la Altura máxima de la escena más cercana a las cámaras que se quiere observar.

Estableciendo  $w_{min}$  y  $h_{min}$  a  $z_{min}$  (ver figura 3.12) se imponen más restricciones sobre la elección de parámetros como  $b$  o  $f$ .

### **Paso 2: Calcular las dimensiones de la imagen sobre el sensor.**

Dada la geometría del sistema, se pueden calcular las posiciones de las proyecciones de los puntos  $P_1(x_1, z_{min})$  y  $P_2(x_2, z_{min})$ , los cuales limitan el área mínima observable (ver figura 3.12). Los puntos  $P_1$  y  $P_2$  tienen coordenadas en el espacio 3D referidas a la cámara B:  $(x_{2b}-w_{min}, z_{min})$ ,  $(x_{2b}, z_{min})$ , respectivamente. Sus proyecciones sobre el sensor de la cámara B están dadas por las siguientes relaciones:

$$P_{2b} = f \cdot \frac{x_{2b}}{z_{min}} \quad (3.37a)$$

$$P_{1b} = f \cdot \frac{x_{2b} - w_{min}}{z_{min}} \quad (3.37b)$$

Con las expresiones anteriores es posible encontrar el área común mínima de las imágenes A y B (segmentos  $\overline{P_{2a}P_{1a}}$  y  $\overline{P_{2b}P_{1b}}$  de la figura 3.12 en color verde). La longitud horizontal del área común utilizando (3.37) sería:

$$ACmin_x = P_{2b} - P_{1b} = f \cdot \frac{w_{min}}{z_{min}} \quad (3.38a)$$

De la misma manera se puede obtener la máxima área observable, que corresponde a la región limitada por  $P_3(x_3, z_{max})$  y  $P_4(x_4, z_{max})$ . Esta región establece la longitud mínima de la región del CCD a utilizar para que todos los objetos de la escena que se encuentren a una distancia  $z_{max}$  sean visibles en ambas cámaras.

$$ACmax_x = P_{3b} - P_{4b} = f \cdot \frac{w_{max}}{z_{max}} \quad (3.38b)$$

Si el ángulo de visión de las cámaras es  $2\mathbf{f}$ , entonces el ancho de la escena observable en función de la distancia está dado por:

$$w(z) = 2 \cdot \tan(\mathbf{f}) \cdot z - b$$

Siendo  $\tan(\mathbf{f}) = \frac{\left(\frac{w_{CCD}}{2}\right)}{f}$ , donde  $w_{CCD}$  es el ancho del sensor.

Luego, el ancho de la escena en función de la distancia sería:

$$w(z) = \left(\frac{w_{CCD}}{f}\right) \cdot z - b \quad (3.39)$$

Evaluando (3.39) en  $z_{min}$  y  $z_{max}$ , y reemplazando en las relaciones (3.38a) y (3.38b) se tiene:

$$ACmin_x = w_{CCD} - \frac{b \cdot f}{z_{min}} = w_{CCD} - \mathbf{r}_x \cdot d_{max} \quad (3.40a)$$

$$ACmax_x = w_{CCD} - \frac{b \cdot f}{z_{max}} = w_{CCD} - \mathbf{r}_x \cdot d_{min} \quad (3.40b)$$

Como se muestra en la figura 3.12, el desplazamiento (*offset*) entre las regiones utilizadas de cada sensor está dado por la suma del *baseline* y la disparidad mínima (expresada en las mismas unidades que el *baseline*). El *offset* entre las áreas utilizadas de ambos sensores sería entonces:  $b + \frac{b \cdot f}{z_{min}}$ .

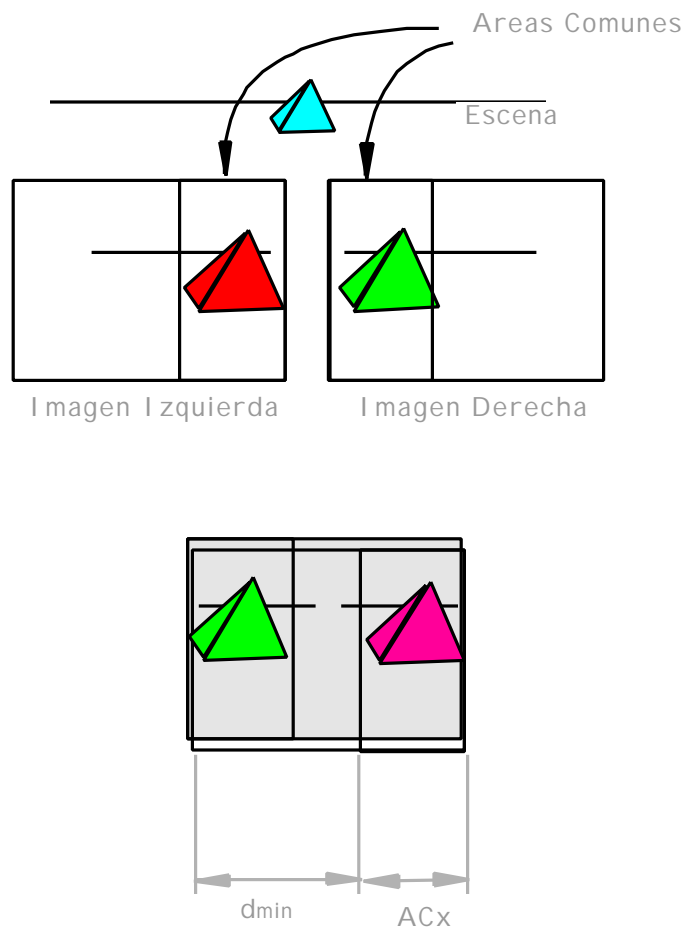


Figura 3.13: Regiones comunes en cada imagen del par estéreo.

Los resultados obtenidos se representan gráficamente en la figura 3.13, la cual muestra que sólo un sector de cada imagen es visible en ambas imágenes. El

tamaño del sector visible depende de la distancia a la cual se encuentran los elementos observados, el *baseline* y la distancia focal, como lo muestran las relaciones (3.40a) y (3.40b). La obtención de las áreas visibles en ambas imágenes es importante para limitar el proceso de búsqueda de correspondencias.

**Paso 3: Calcular el número de pixels requeridos para el rango de disparidades de acuerdo a la tolerancia.**

Usando (3.19) y la tolerancia  $\Delta z_d \%_{max}$  de la especificación, obtener la disparidad  $d$  [pixels] como:

$$d = \frac{1}{\Delta z_d \%_{max}} \cdot 100 \quad (3.41)$$

**Paso 4: Si se requiere especificar la densidad de pixels en pixels/mm.**

Primero debe calcularse el ancho de un pixel necesario para el peor caso (esto ocurre para la distancia máxima, porque el error en la estimación de  $z$  es máximo):

$$\mathbf{r}_x = \frac{1}{d} \cdot \frac{b \cdot f}{z_{max}} \quad (3.42)$$

La densidad de pixels es el inverso de  $\mathbf{r}_x$ . Este valor sirve para calcular la cantidad mínima de pixels que requiere el sensor y puede calcularse a partir de (3.40a) y (3.42) como:

$$Pixels_{CCD} = \frac{ACmax_x}{\mathbf{r}_x} = d \cdot \frac{w_{max}}{b} \quad (3.43)$$

### 3.1.7 Ejemplo de Diseño del Sistema

*Descripción de la Aplicación:*

Considere un sistema estereoscópico cuyo propósito es reconstruir tridimensionalmente la superficie de objetos cuya longitud máxima en el sentido  $z$  (sentido del eje óptico de las cámaras) es de 200 mm, con un error relativo al tamaño del objeto del 2.5% (i.e. un error absoluto de  $\pm 5$  mm). Las cámaras cuentan con lentes de  $f = 16$  mm, y CCDs de 6.3 mm x 4.8 mm. La longitud horizontal de cada pixel es  $r_x = 0.01$  [mm/pixel]. Se requieren mapas de disparidad con al menos 128 x 256 puntos.

#### 1. Determinación de la Disparidades

Usando (3.41) definir una aproximación para las disparidades requeridas para cumplir la tolerancia.

$$d \geq \frac{1}{\Delta z_d \%_{max}} \cdot 100 = \frac{100}{2.5} = 40 \text{ pixels}$$

#### 2. Definición del Rango

Si la parte más alejada del objeto se encuentra a 800 mm de las cámaras:

$$s = 200 \text{ mm}, z_{max} = 800 \text{ mm} \rightarrow z_{min} = 600 \text{ mm}$$

#### 3. Determinación del *Baseline*

Dado que se requiere reconstruir el objeto con un error relativo a su tamaño no superior al 2.5%, la relación entre los parámetros del sistema y el error relativo más apropiada es la establecida en (3.36).

$$\text{Se tiene entonces que } E[e] \approx \frac{0.5139}{b} = 0.025 \rightarrow b = 20.56 \text{ mm}$$

#### 4. Verificación del área del CCD requerida

Utilizando (3.40a) y los resultados anteriores puede calcularse el tamaño horizontal útil de los pares estéreo, este sería:

$$ACmin_x = w_{CCD} - \frac{b \cdot f}{z_{min}} = 6.3 - \frac{20 \cdot 16}{600} = 5.7 \text{ mm}$$

Usando el ancho por pixel, los 5.7 mm corresponden a 577 pixels horizontales, que son más que los 128 pixels requeridos.

Puede modificarse (3.40b) para calcular el área útil a partir de un tamaño de imagen capturada. Por ejemplo si las imágenes son de 256 x 256, entonces:

$$ACmin_x = w_x [pixels] - \frac{b \cdot f}{r_x \cdot z_{min}} = 256 - \frac{20 \cdot 16}{0.01 \cdot 600} = 202 \text{ pixels}$$

El resultado anterior confirma que es posible cumplir con el mínimo de 128 x 256 pixels, si se capturan imágenes de tan sólo 256 x 256 pixels, las cuales requieren menor cantidad de memoria de almacenamiento, que el tamaño completo de 577 x 256.

En caso de que el tamaño requerido no se hubiese cumplido, habría sido necesario repetir los pasos 3 y 4, utilizando, por ejemplo, lentes con una distancia focal menor.

### 3.2 Adquisición de las Imágenes

La obtención de pares estéreo con distintas características fue la parte inicial del proceso de experimentación y evaluación de las técnicas estereoscópicas más apropiadas para la reconstrucción tridimensional densa (estereoscopía mediante correlación y mediante programación dinámica).

Las imágenes se obtuvieron utilizando dos tipos de cámaras cuyas características se muestran en la tabla 3.1. Se utilizó principalmente la cámara Sony



XC-77. La digitalización de la señal CCIR se realizó mediante la tarjeta de adquisición de imágenes (*Frame Grabber*) Dipix XPG-1000 [Dipix96].



Figura 3.14: Cámara Color Cohu 8380.

Tabla 3.1: Características de las Cámaras.

Característica	Sony XC-77	Cohu 8380 (PAL)
Tipo	B/W - PAL	Color - PAL
Señal	CCIR (Interlaced)	CCIR (Interlaced)
CCD	2/3" Interline Transfer	1/2" Interline Transfer
Elementos Activos	768 x 581	752 x 582
Area Activa	8.8mm x 6.6mm	6.4mm x 4.8mm
Tamaño Celda	0.011mm x 0.011mm	0.0086mm x 0.0083mm
Iluminación Mínima	0.5 lux @ F1.4/AGC On	0.5 lux @ F7/AGC On
Consumo	2.2W @ 12V	4.5W @ 12V
Peso	190 g	49 g
Lente	Sony, $f = 16mm$	Tamron, $f = 6.5mm$

Para obtener imágenes estéreo se utilizó la configuración *binocular paralela*, capturando primero una imagen, y luego desplazando la cámara lateralmente sobre un riel graduado para adquirir el par correspondiente. El soporte de la cámara, mostrado en la figura 3.14, permite desplazamientos horizontales de 500 mm y verticales en el rango 200 mm - 1000 mm con respecto al plano horizontal.

Para la alineación de la cámara se utilizó un nivel y marcas de referencia sobre el monitor. La marcas sobre el monitor fueron de gran ayuda para alinear con

precisión la cámara. El procedimiento bastante simple consistía en hacer coincidir las marcas de referencia en la escena que se veían en la imagen, con las marcas hechas sobre la pantalla del monitor. La alineación se consideraba correcta si al desplazar verticalmente la cámara la marca central de la imagen coincidía siempre con la marca central del monitor.

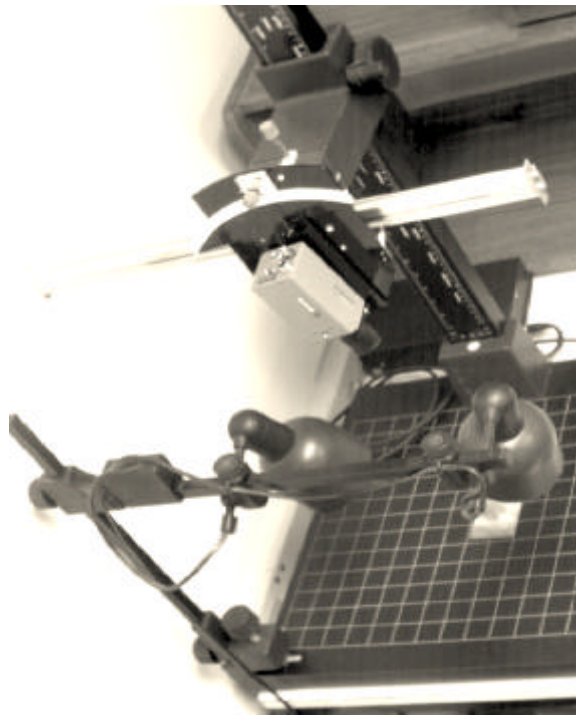


Figura 3.14: Sistema de adquisición de imágenes.

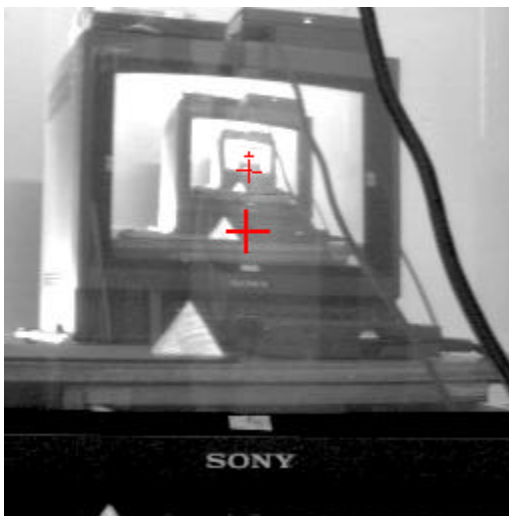


Figura 3.15: Marca de alineación.

El método de alineación se puede explicar con la figura 3.15, la cual muestra la marca de alineación sobre la pantalla del monitor (resaltada en rojo). La imagen en el monitor corresponde al mismo monitor con su marca de alineación (mostrado *ad infinitum*). Si la cámara se colocase perpendicular a la pantalla y centrada con respecto a ésta, entonces sólo se vería una marca.

Las imágenes adquiridas para experimentar las técnicas estereoscópicas son de 256x256 pixels cuantizados en 8 bits. Los pares estéreo se muestran horizontalmente, y pueden ser vistos utilizando lentes estéreo estándar. Algunas personas pueden verlos fusionando los pares en uno solo, sin ayuda de lentes.

Existen varias razones para la elección de las dimensiones de las imágenes. En primer lugar, el espacio de almacenamiento y el tiempo de procesamiento es menor al utilizar estas dimensiones. El tamaño 256x256 resulta apropiado para las pruebas en interiores, y permite tiempos de proceso del par estéreo razonables (típicamente entre 1 a 10 segundos).

Evitar la distorsión que introducen los lentes es la segunda razón que motiva el uso de la región central de 256x256 de la imagen original (776x568). En la adquisición de las imágenes se utilizó un lente con una distancia focal de 16 mm, que tiene una distorsión radial inferior a la que causa el lente de 6.5 mm (ver figura 3.5).

Al utilizar la región de 256x256 no se requiere de una rectificación de la imagen. Sin embargo, hay que tener presente que para aprovechar una mayor área del sensor, como podría ser necesario para aplicaciones en exteriores, la rectificación de la imagen es imprescindible si se utilizan lentes gran angular. Además los algoritmos comunes de correlación fallarían en la búsqueda de correspondencias para puntos en los bordes y especialmente en las esquinas de las imágenes si estas no son corregidas.

El principal objeto de prueba para la reconstrucción tridimensional es una pirámide que tiene una base cuadrada de 55 mm y una altura de 48 mm. La figura 3.16a muestra la pirámide relativa a un diskette de 3½". Otra perspectiva se muestra en 3.16b.

Las imágenes *Pirámide* tienen características que cumplen ciertas condiciones:

- La superficie de la pirámide está formada por planos con pendiente lineal (rampa) y no posee discontinuidades. La pirámide esta forrada con papel de diario para crear una textura adecuada, con patrones claramente identificables sobre su superficie. El fondo es plano y perpendicular a eje óptico de las cámaras. Está cubierto con una fotocopia de la superficie de Tritón (ver figura 3.16b), una de las lunas del planeta Neptuno, con el fin de crear un patrón aleatorio en el fondo.
- No existe oclusión de ningún objeto en la escena, ni de las caras de la pirámide, ya que desde arriba son visibles desde cualquier punto dentro de rango de movimiento que posee la cámara.
- No hay variaciones importantes de la intensidad luminosa entre ambas tomas. El papel de diario es mate y no causa reflejos intensos a menos que se lo ilumine con una luz directa muy potente o muy cercana a la superficie. Se puede considerar que los objetos en la escena son casi *lambertianos*.



Figura 3.16a: Vista de elevación de la pirámide.

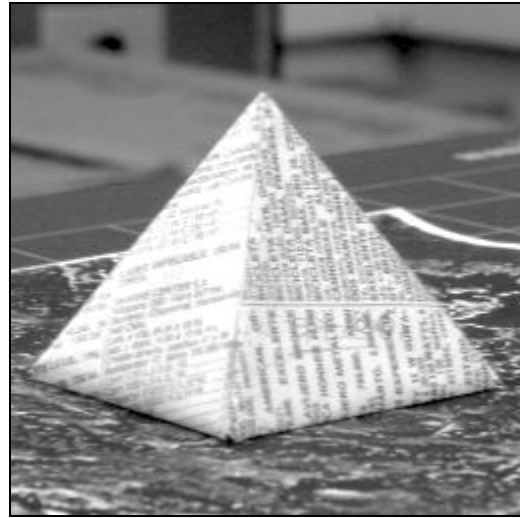


Figura 3.16b: Perspectiva de la pirámide.

La secuencia de pares estéreo está compuesta por una secuencia de vistas desde arriba de la pirámide tomadas a distinta altura. La secuencia será llamada Pira700. El primer par de Pira700 (figura 3.17) fue tomado con el lente de la cámara a una distancia de 700 mm desde el tablero cuadrulado (ver figura 3.14). Los pares siguientes fueron tomados disminuyendo la distancia entre el modelo y la cámara en pasos de 3 mm. La secuencia esta compuesta por 16 pares, el último (figura 3.18) tomado a una distancia de 655 mm del tablero.

Se empleo para las tomas un sistema binocular paralelo con *baseline* de 20 mm y distancia focal de 16 mm.

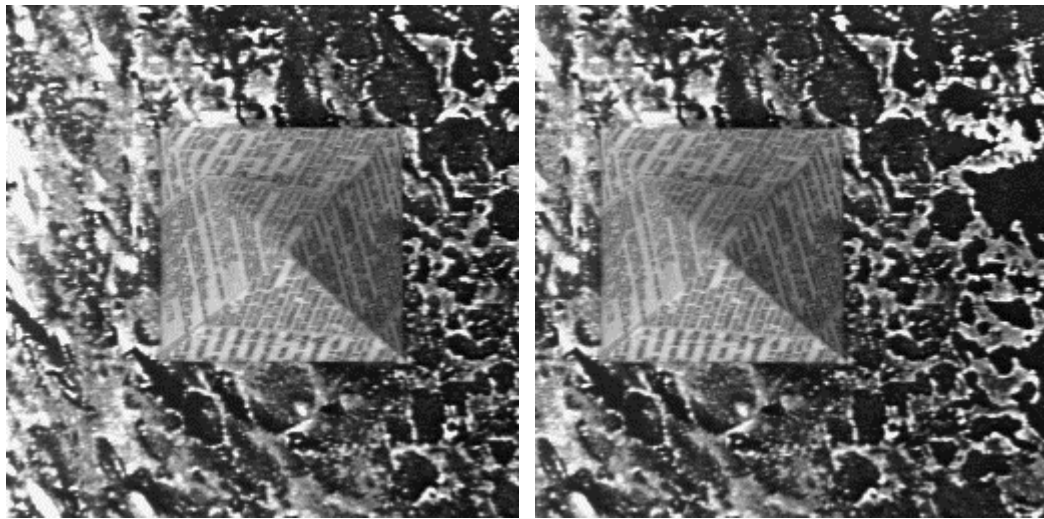


Figura 3.17: Primer par de la secuencia Pira700 (Distancia a la base: 700mm).

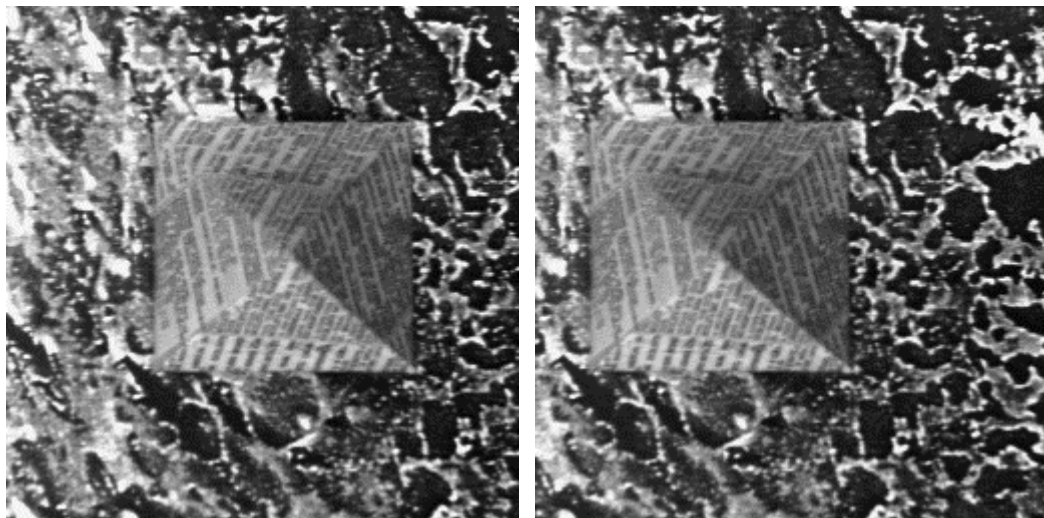


Figura 3.18: Último par de la secuencia Pira700 (Distancia a la base: 655mm).

Las características de los pares estéreo Pira700, algunas naturales otras introducidas artificialmente (como los patrones del papel), son muy favorables para el uso de la correlación normalizada debido a la gran cantidad de puntos y marcas claramente identificables en ambas imágenes (textura adecuada).

No existe una definición formal de textura, pero intuitivamente es un indicador de la suavidad, aspereza y regularidad (repetitividad de un patrón como en

las baldosas o azulejos) [Gonz93]. Existen tres enfoques distintos para cuantificar la textura de una imagen: estadísticos, estructurales y espectrales. Un indicador de que las imágenes tienen un alto contenido de texturas es la varianza bidimensional en la vecindad de cada punto de la imagen [Gonz93]. En la figura 3.19 se muestran imágenes de la desviación estándar correspondientes a la imagen derecha de la figura 3.18, para dos tamaños distintos de vecindad. Las imágenes están *pseudocoloreadas* para aumentar el contraste. Se destaca la zona verde, que corresponde a la región de mayor varianza, y en rojo las regiones de menor varianza.

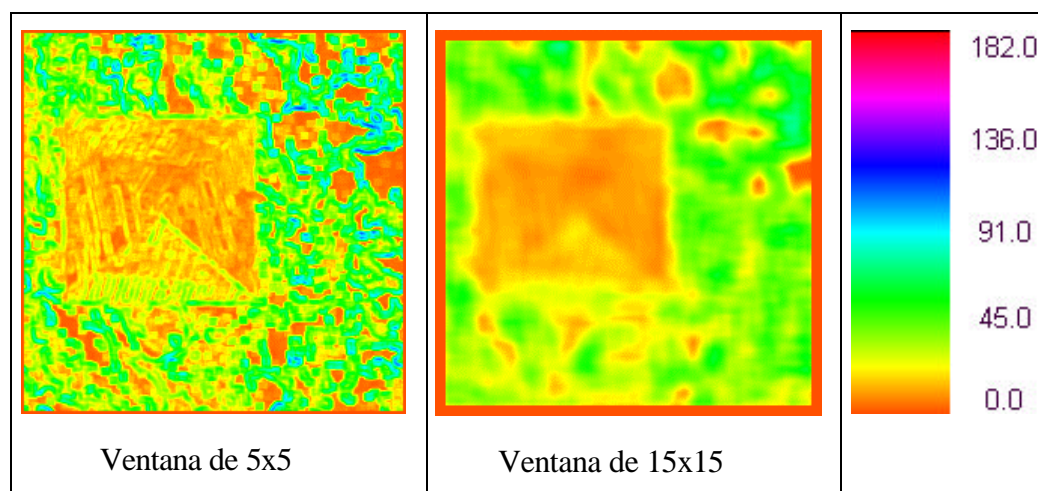


Figura 3.19: Varianza de la imagen derecha de la figura 3.18.

Otra ventaja que ofrecen estas imágenes es que corresponden a un modelo de dimensiones conocidas, permitiendo la comparación de los resultados de la reconstrucción 3D con los datos reales.

### 3.3 Procesamiento Secuencial y Análisis de las Imágenes

El objetivo del algoritmo desarrollado es la reconstrucción tridimensional en forma exacta y precisa de los objetos en la escena. Para lograr una reconstrucción tridimensional confiable se propone el procesamiento de secuencias de pares estéreo.

En la figura 3.20 se presentan las etapas del procesamiento secuencial de pares estereoscópicos. El sistema de reconstrucción tridimensional puede separarse en tres etapas principales. La primera tiene por finalidad calibrar el sistema binocular de adquisición. La calibración del sistema, ya sea en forma manual o automática, permite obtener la información de la alineación del sistema y de las distorsiones a partir de una secuencia inicial de calibración. Una vez que el sistema está calibrado, la información de la calibración  $c_b$  es utilizada para procesar el par de imágenes de la cámara izquierda  $L$ , y la derecha  $R$ , tomado en el instante  $k$ .

En la etapa de procesamiento del par  $L-R$  se distinguen tres procesos. Primero, cada punto de la imagen derecha es buscado en la izquierda (i.e. se realiza la búsqueda de correspondencias tomando como referencia la imagen derecha). Como resultado de la búsqueda de correspondencias para el  $k$ -ésimo par, se obtiene el mapa o imagen de disparidades  $D_k$ , a partir del cual se reconstruye la superficie  $Z_k$  utilizando la relación (3.28). Durante el proceso de *matching*, las disparidades son validadas y un mapa de confiabilidad  $W_k$  de los resultados se calcula en función de los valores de la curva de correlación para cada punto. El mapa  $W_k$  es un índice que describe el grado de certeza o la calidad de la correspondencia para cada punto del mapa  $D_k$ , y por lo tanto para cada punto del mapa de distancias  $Z_k$ .

La tercera etapa dentro del procesamiento del par es la de validación de resultados e interpolación de las zonas no correspondidas como se muestra en la figura (3.20). La fase de validación de los resultados debería ser capaz de detectar los errores y eliminar las correspondencias falsas. El algoritmo propuesto utiliza los índices de confiabilidad  $W_k$  (calculados para cada punto del mapa de disparidades) como estimadores de la validez de los resultados. Los resultados de distancias validados  $Z_k$  se modifican de acuerdo al movimiento realizado por el sistema en el instante  $k$  (información almacenada en la variable  $MOV_k$ ) obteniéndose  $Z'_k$ , y se fusionan con los resultados del modelo 3D obtenido en etapas anteriores  $MZ_{k-1}$ . El modelo 3D resultante se almacena en  $MZ_k$  para su uso posterior o para alimentar la tercera etapa del sistema estéreo.



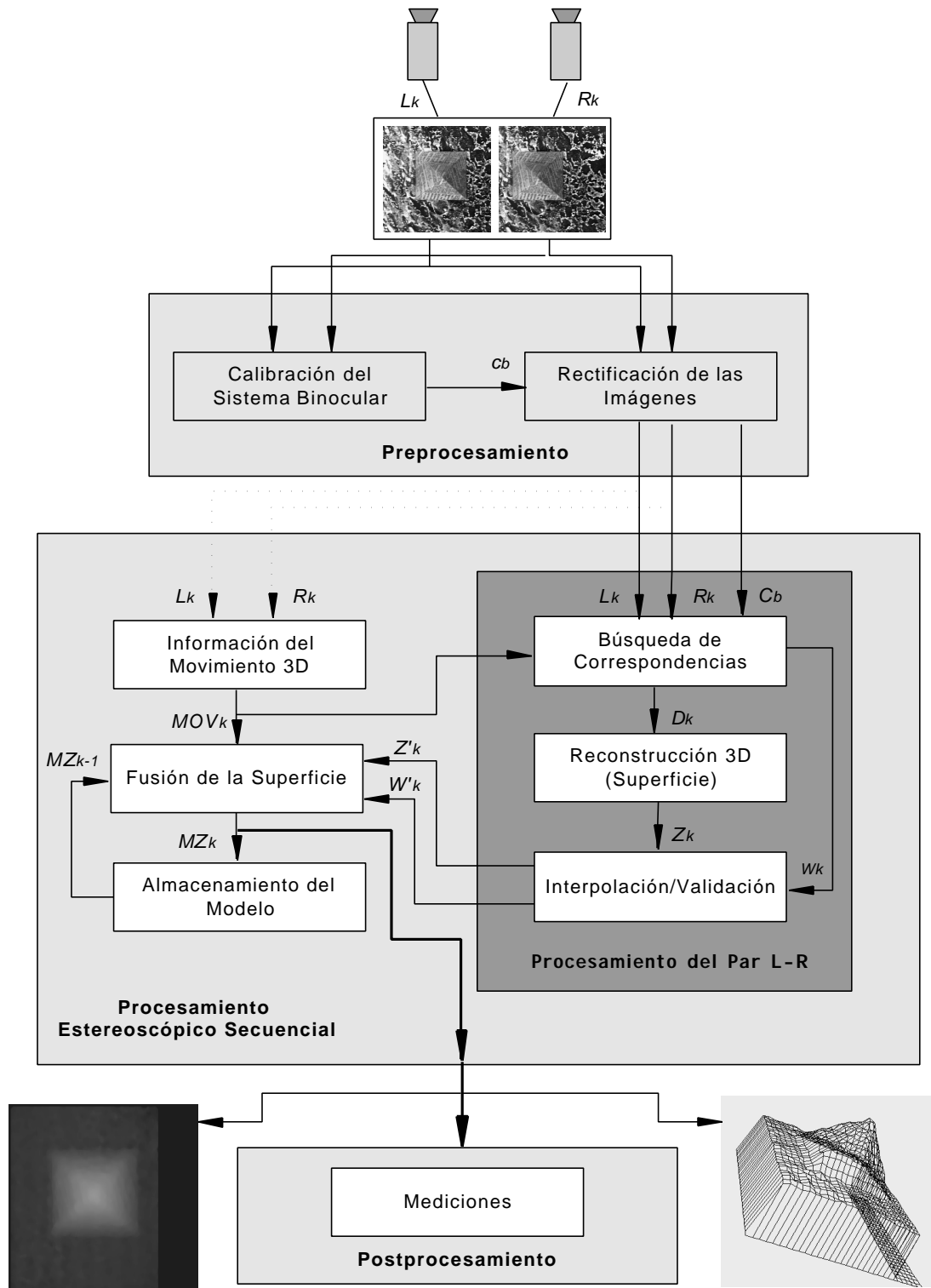


Figura 3.20: Etapas del procesamiento de secuencias estereoscópicas.

En la tercera etapa se realizan mediciones del modelo 3D resultante. Las mediciones dependen del tipo de aplicación. A partir de  $MZ_k$  se pueden medir distancias entre las cámaras y el objeto, como se suele realizar en la creación de mapas topográficos digitales. La medición de distancias también puede ser necesaria en navegación para evitar colisiones. Otras mediciones pueden ser el volumen o la forma, para aplicaciones de reconocimiento de objetos y CAD. En [Duc197] las mediciones son utilizadas para el diagnóstico clínico en oftalmología.

El trabajo desarrollado se concentra en la etapa del procesamiento secuencial, asumiendo que las imágenes han sido previamente rectificadas en la primera etapa del sistema. También se asume que la información de la geometría  $c_b$  del sistema binocular y la información del movimiento del sistema  $MOV_k$  ha sido proporcionada *a priori*.

El algoritmo propuesto pone énfasis en la reconstrucción 3D precisa del modelo y por esta razón trata de evitarse la interpolación. Se pretende mostrar que el uso de múltiples pares tomados a distintas distancias del modelo y la fusión de las superficies (*mapas de distancia o profundidad*) empleando filtros no lineales tipo FIR debería ser suficiente para lograr un modelo 3D denso (con una gran cantidad de puntos) y confiable.

A continuación se describen detalladamente las distintas etapas del procesamiento de la secuencia estereoscópica que fueron implementadas en este trabajo. La etapa de calibración y rectificación, aunque muy importante para una búsqueda de correspondencias exitosa, no recibe un énfasis particular por ser un tema bastante complejo, extenso y fuera de la línea principal de esta investigación.

### **3.3.1 Calibración del Sistema y Rectificación de las Imágenes**

En la sección 3.2 ya se describió el sistema rudimentario de calibración. En esta sección se discute y verifica la calidad de la calibración mediante el análisis de las imágenes.

## A. Rectificación Epipolar

El algoritmo de búsqueda de correspondencias requiere que los vectores epipolares en ambas imágenes coincidan con las líneas de barrido horizontal (*scanlines*). Como el soporte de la cámara permite el desplazamiento de ésta a lo largo del riel, los ejes ópticos del par resultan bastante paralelos. No hay un ángulo de elevación  $\alpha$  apreciable. Se verificó que al desplazar la cámara enfocando una línea horizontal de un patrón, la línea se mantuviese sobre la misma coordenada  $Y$ . La rotación del sensor tampoco es apreciable, ya que las líneas verticales del patrón cuadrículado se mantienen dentro de una columna de un pixel de ancho para el desplazamiento lateral.

De acuerdo al análisis de la sección 3.1, el error en la medición 3D es más sensible a errores en el ángulo de convergencia, el cual debería ser  $0^\circ$  para lograr que las rectas epipolares sean paralelas a la recta que une los centros ópticos de las cámaras. Especial cuidado debe ponerse al ajustar la fijación de la cámara al riel, ya que ajustes exagerados de las prensas introducen un ángulo de convergencia mínimo, pero que puede producir distorsiones importantes en la imagen.

Puede afirmarse que la secuencia Pira700 no requiere de rectificación. Los resultados de reconstrucción 3D para un par de Pira700, utilizando la rectificación propuesta en [Coch90], no mejoraron. La rectificación utilizada por [Coch90] consiste en una corrección lineal de la imagen, para casos en que la desalineación es mínima.

## B. Alineación de la Secuencia

La verificación de la alineación de la secuencia se realiza para todas las imágenes izquierda de Pira700 y todas las imágenes derecha por separado. Como en la primera toma aparece la pirámide vista desde una distancia  $z_0 = 700$  mm, ésta se ve más pequeña que en las tomas siguientes, en las que el par estéreo se acerca hasta una distancia  $z_{15} = 655$  mm. Por esta razón, para verificar la alineación de las imágenes  $k = 1, \dots, 15$  de la secuencia con respecto al par inicial de referencia ( $k = 0$ ), es necesario reducir el tamaño de las imágenes en un factor que se puede considerar igual a  $z_k/z_0$ , donde  $z_k$  corresponde a la distancia máxima (distancia a la base de la

pirámide) para el par  $k$ . En la figura 3.21 se muestran las imágenes  $R_0$  y  $R_{15}$ , esta última reducida en un factor  $655/700$  (0.94). Si la alineación es correcta, entonces la diferencia entre la imagen de referencia  $R_0$  y la imagen corregida por el desplazamiento  $R'_k$ , debería ser nula. Al analizar las imágenes de la figura 3.21, se confirma este supuesto.

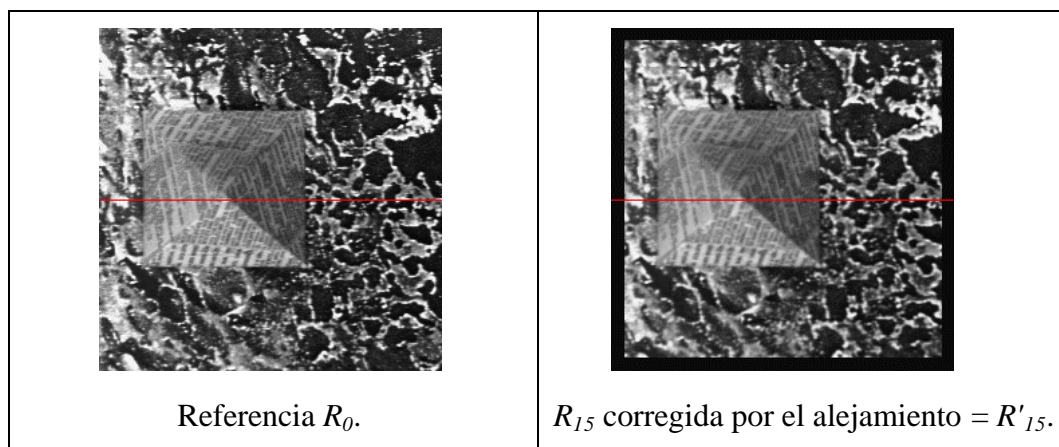


Figura 3.21: Alineación de la secuencia Pira700.

La figura 3.22 muestra como la diferencia  $R_0 - R'_{15}$  para el área de interés (la pirámide, marcada con el cuadrado rojo) es prácticamente nula. La señales correspondientes a la línea horizontal 128 de  $R_0$  y  $R'_{15}$  (línea roja de la figura 3.21) se muestran en la figura 3.23 en rojo y azul, respectivamente. La diferencia promedio de intensidades de las imágenes es inferior al 3.75% del nivel máximo de cuantización (255).

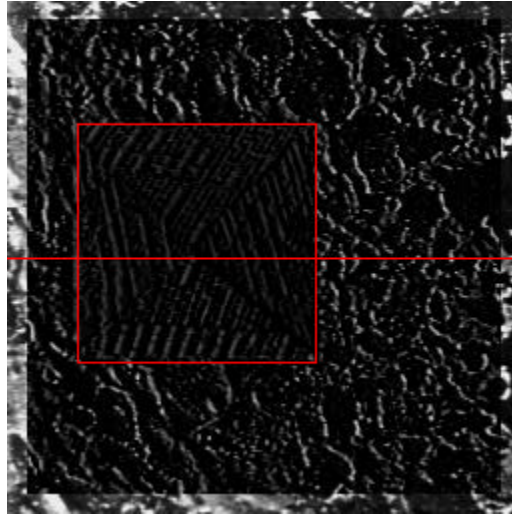


Figura 3.22: Imagen  $R_0$ - $R'_{15}$ .

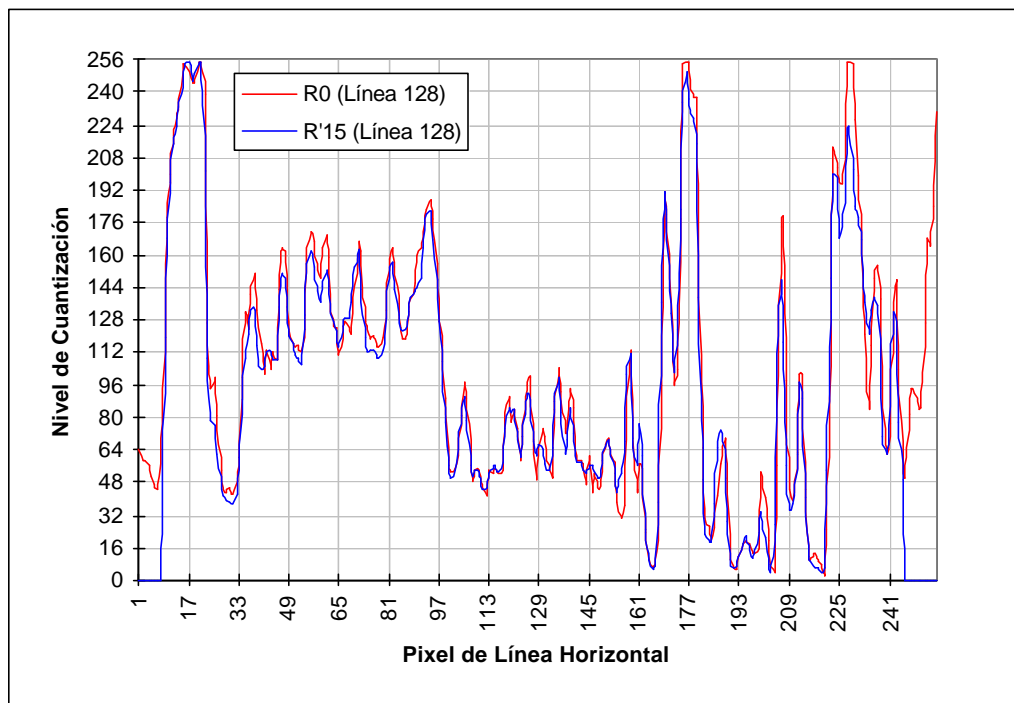


Figura 3.23: Comparación de la señal de la línea horizontal 128 de las imágenes  $R_0$  (original) y  $R'_{15}$  (rectificada) mostradas en la figura 3.21.

### 3.3.2 Procesamiento del Par Estéreo: Disparidades y Confiabilidad

#### A. Algoritmo de Búsqueda de Correspondencia

La investigación bibliográfica realizada y pruebas con los algoritmos implementados por [Duc197] permitieron definir que los métodos de *matching* que han logrado mejores resultados son: (1) aquellos basados en correlación de áreas como los de correlación normalizada NCC [Fua91, Faug93i] o los de suma de diferencias al cuadrado SSD [Kana94i] y (2) algunos basados en programación dinámica [Cox94i, Birc98].

Evaluando tanto la calidad de los resultados como el tiempo de procesamiento requerido se prefirió el primer método, utilizando la NCC como medida de similitud. Ambos métodos permiten encontrar mapas de disparidad densos, a diferencia de los métodos basados en correlación de segmentos u otras características; sin embargo, los algoritmos estereoscópicos basados en programación dinámica son más lentos.

El método basado en la NCC también fue elegido por la simplicidad inherente que presenta para el cálculo de disparidades sub-píxel. Los algoritmos desarrollados en este trabajo se basan en el algoritmo de correlación normalizada [Faug93i] acelerada según [Fua91] con resolución sub-píxel mediante interpolación cuadrática de los valores de la curva de correlación.

Es importante mencionar que si bien la correlación basada en SSD es más rápida que la NCC, la SSD es menos confiable. La desventaja de la SSD es que al no ser una correlación normalizada resultan más disparidades erróneas. Como uno de los objetivos es el de aumentar la exactitud y precisión de los resultados se descartó el uso de la SSD por su menor insensibilidad al ruido.

La correlación normalizada entre un punto de la imagen derecha  $R(x, y)$  y un punto en la imagen izquierda  $L(x+d, y)$  a una disparidad  $d$ , considerando una ventana o bloque de correlación de  $(2N+1) \cdot (2P+1)$  puntos se define como:

$$C(x, y, d) = \frac{\sum_{i=-N}^N \sum_{j=-P}^P (R(x+i, y+j) - \bar{R}(x, y))(L(x+i+d, y+j) - \bar{L}(x+d, y))}{\sqrt{\sum_{i=-N}^N \sum_{j=-P}^P (R(x+i, y+j) - \bar{R}(x, y))^2} \sqrt{\sum_{i=-N}^N \sum_{j=-P}^P (L(x+i+d, y+j) - \bar{L}(x+d, y))^2}} \quad (3.44)$$

donde

$$\begin{aligned} \bar{R}(x, y) &= \frac{\sum_{i=-N}^N \sum_{j=-P}^P R(x+i, y+j)}{(2N+1)(2P+1)} \\ \bar{L}(x, y) &= \frac{\sum_{i=-N}^N \sum_{j=-P}^P L(x+i, y+j)}{(2N+1)(2P+1)} \end{aligned} \quad (3.45)$$

Los términos  $R(x, y)$  y  $L(x, y)$  corresponden a valores de intensidad para el punto  $(x, y)$  de las imágenes derecha e izquierda, respectivamente. Los términos del denominador de (3.44) corresponden a la desviación estándar de  $R(x, y)$  y  $L(x+d, y)$  en una vecindad de  $(2N+1)$  columnas por  $(2P+1)$  filas. El valor de  $C$  puede tomar valores entre 1 y -1. La correlación es 1 en caso de que todos los puntos en la vecindad de  $R(x, y)$  sean idénticos a los puntos en la vecindad de  $L(x+d, y)$ .

El cálculo de la correlación normalizada se realiza para cada punto de la imagen derecha, buscando a lo largo del segmento epipolar de la imagen izquierda. El segmento epipolar queda definido por la mínima y máxima disparidad, como se muestra en la figura 3.24.

La búsqueda de correspondencias es muy lenta si se utiliza (3.44) directamente, por esta razón se implementó un algoritmo que calcula (3.44) y (3.45) en forma recursiva [Fua91, Faug93i], acelerando enormemente el proceso. Los detalles de esta implementación pueden encontrarse en el Anexo C.

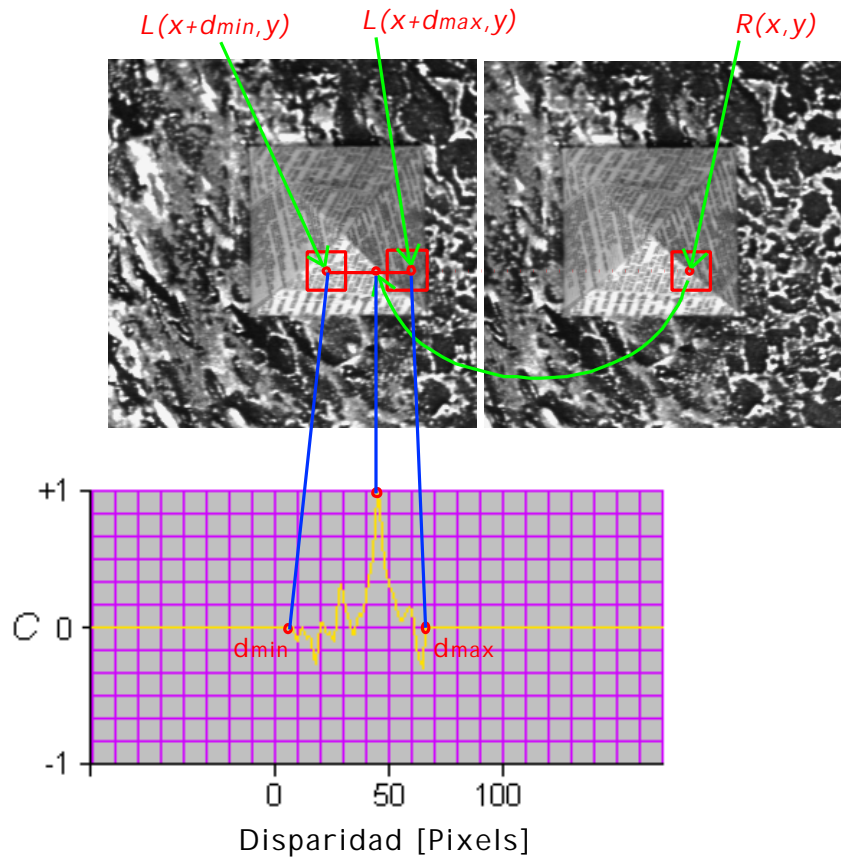


Figura 3.24: Cálculo de la correlación normalizada.

El proceso de *matching* se puede describir como la comparación del patrón de intensidades en la ventana de referencia de la imagen derecha (en torno a  $R(x,y)$ ), con cada uno de los patrones que encierra una ventana que se desplaza sobre la imagen izquierda entre las coordenadas  $x+d_{min}$  y  $x+d_{max}$ , para la misma coordenada vertical  $y$ . Al correlacionar cada patrón de referencia en la imagen derecha con un patrón en la imagen izquierda se genera una curva de *Correlación vs. Disparidad*, como se muestra en la figura 3.24. El punto de la imagen izquierda cuya vecindad es más parecida al patrón en la vecindad de referencia obtiene un *puntaje* de correlación máximo, y por lo tanto puede suponerse como el punto correspondiente a  $R(x,y)$ . Este proceso se repite para cada punto de la imagen derecha, generándose así, el mapa de disparidades.



En la figura 3.24 puede verse la curva de correlación obtenida para el pixel  $R(144, 148)$ , en un rango de disparidades que varía entre 9 y 65 pixels. Se utiliza una ventana de  $15 \times 15$  ( $N=P=7$ ). La correlación máxima es de 0.975 y ocurre para una disparidad de 45.72 pixels, calculada con una precisión significativa de una décima, es decir  $\pm 0.1$  pixels.

Para lograr una resolución *sub-pixel*, se interpola la curva de correlación entre el máximo y sus dos vecinos mediante una parábola, como lo proponen [Fuag93i], o [Deve95] para su algoritmo de detección de bordes con resolución *sub-pixel*. El valor de disparidad se toma como el punto para el cual la parábola alcanza su máximo (ver figura 3.25). En el Anexo C se encuentran las ecuaciones de interpolación.

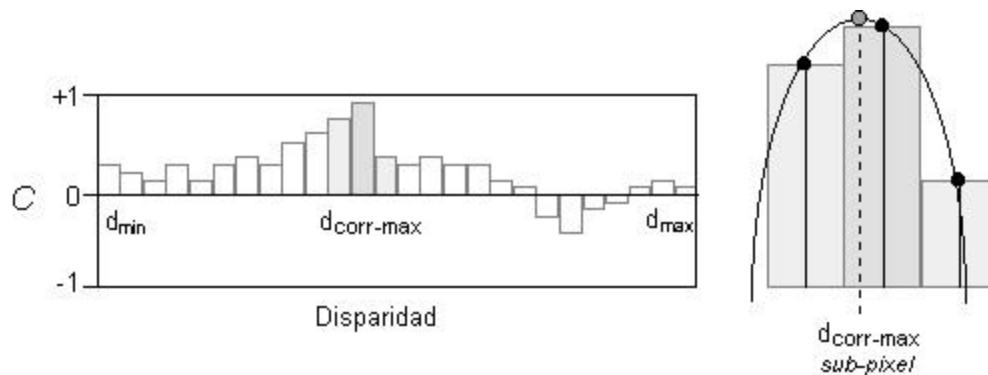


Figura 3.25: Resolución *sub-pixel* de la disparidad usando interpolación cuadrática

La elección de  $N$  y  $P$  depende del tipo de textura y detalle que presente la imagen. El trabajo de [Nish84] demuestra que teóricamente la *probabilidad de error en la búsqueda de correspondencias* decrece en la medida que aumenta el tamaño de la ventana de correlación y la textura en la escena. Sin embargo, [Kana94] demuestra que el uso bloques de correlación muy grandes causa la pérdida de precisión en la medición de la profundidad, ya que no consideran las discontinuidades en la escena (que precisamente es uno de los aspectos que se desea medir) o zonas de oclusión. En otras palabras, la *probabilidad de error en la estimación de profundidad* aumenta, aunque los bloques más grandes sean más fáciles de corresponder.

Las ventanas de correlación grandes tienen un efecto de filtro pasabajos, que *suaviza* el mapa de disparidad resultante, causando la pérdida de los detalles de la escena o el objeto reconstruido, especialmente en sus discontinuidades.

En la figura 3.26 se muestran áreas representativas de las texturas que se encuentran normalmente en la mayoría de escenas reales. Los cuadrados rojos se utilizan para indicar el tamaño máximo de bloque de correlación empleado en el análisis; éste fue de  $33 \times 33$  ( $N = P = 16$ ). En la imagen izquierda de la figura 3.26, se muestran en verde los segmentos epipolares correspondientes a los puntos analizados de la imagen derecha.

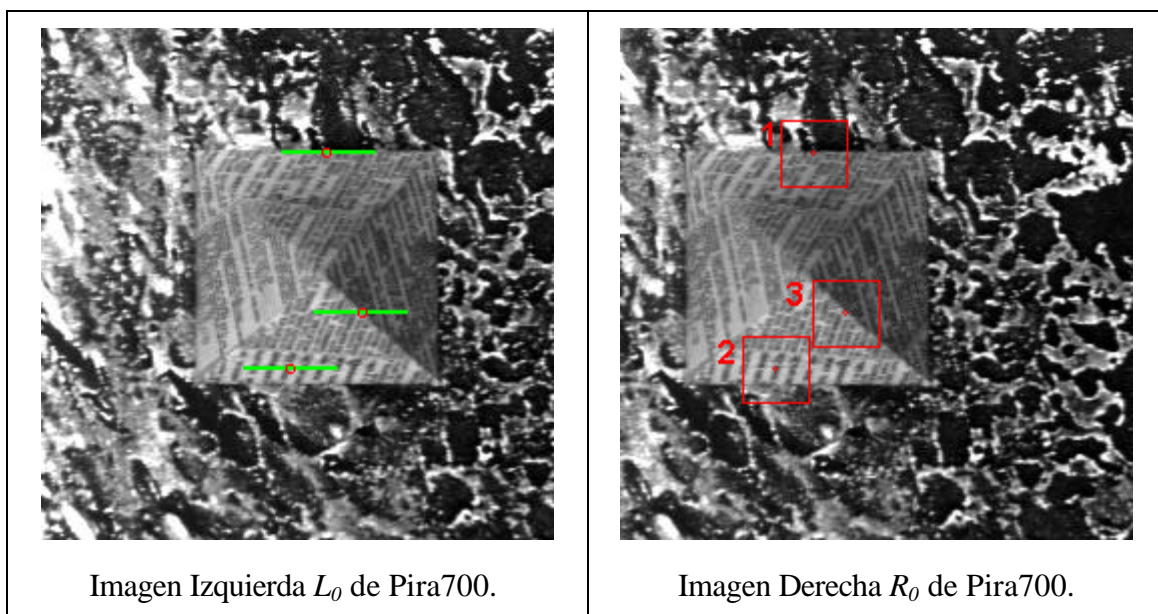


Figura 3.26: Texturas características en el primer par estéreo de Pira700.

La región 1 corresponde a una zona de baja textura y con un patrón horizontal, mientras que la región 2 corresponde a un área de la imagen con patrón periódico en el sentido horizontal. La región 3 se caracteriza por tener una textura que llamaremos aleatoria. Aunque en la región 3 se aprecian algunas líneas del diario, éstas no tienen una periodicidad en el sentido horizontal. El punto de la región 3 es único, porque se encuentra en la arista de la pirámide donde la sombra se distingue

claramente; y por esta razón la región 3 corresponde a un patrón claramente identificable en ambas imágenes, menos sujeto a ambigüedades de correspondencia que las regiones 1 y 2. Los datos correspondientes a las regiones características se encuentran resumidos en la tabla 3.2.

Tabla 3.2: Resumen de características del análisis realizado para las regiones marcadas en la imagen derecha de la figura 3.26.

	<b>Región 1</b>	<b>Región 2</b>	<b>Región 3</b>
Característica	<i>Patrón Horizontal de Baja Textura</i>	<i>Patrón Periódico</i>	<i>Patrón Aleatorio</i>
Posición de la Referencia en la Imagen Derecha	(96, 62)	(77, 170)	(112, 142)
Posición del Punto Correspondiente en la Imagen Izquierda	(140,62)	(122, 170)	(158, 142)
Disparidad Sub-Pixel (Block de Correlación: 11x11)	44.86	45.06	46.90
Rango de Búsqueda D	25-70	25-70	25-70
Distancia [mm] correspondiente al Rango D	457-1280	457-1280	457-1280

Parámetros sistema binocular paralelo: *Baseline* = 20 mm, Distancia Focal = 16 mm.

La figura 3.27 muestra las curvas de correlación que se generan al buscar los pixels de cada región sobre sus segmentos epipolares correspondientes en la imagen izquierda utilizando distintos tamaños de bloque de correlación.

Se aprecia claramente que la correlación máxima se puede detectar fácilmente para los tres casos en que se utiliza un bloque de 33x33. Sin embargo, al

disminuir el tamaño del bloque de correlación aparecen patrones similares en el rango de búsqueda, causando múltiples máximos en la curva de correlación.

Al disminuir el tamaño de la ventana, para la correlación del pixel en la región 1, se aprecia que casi toda la curva empieza a tener correlación máxima. Esto se debe a que la región 1 corresponde a un patrón horizontal con escasa textura, y casi todos los bloques comparados se parecen entre sí.

La periodicidad del patrón (líneas verticales de la pirámide) se aprecia fácilmente en las curvas de correlación de la región 2. El gran parecido entre el punto en correspondencia de la imagen derecha y los de la izquierda causa que la curva de correlación presente múltiples *peaks*, que empiezan a tener la misma magnitud a medida que se reduce el tamaño de la ventana de correlación.

<b>Región 1 - Baja Textura</b>	<b>Región 2 - Textura Periódica</b>	<b>Región 3 - Textura Aleatoria</b>
--------------------------------	-------------------------------------	-------------------------------------

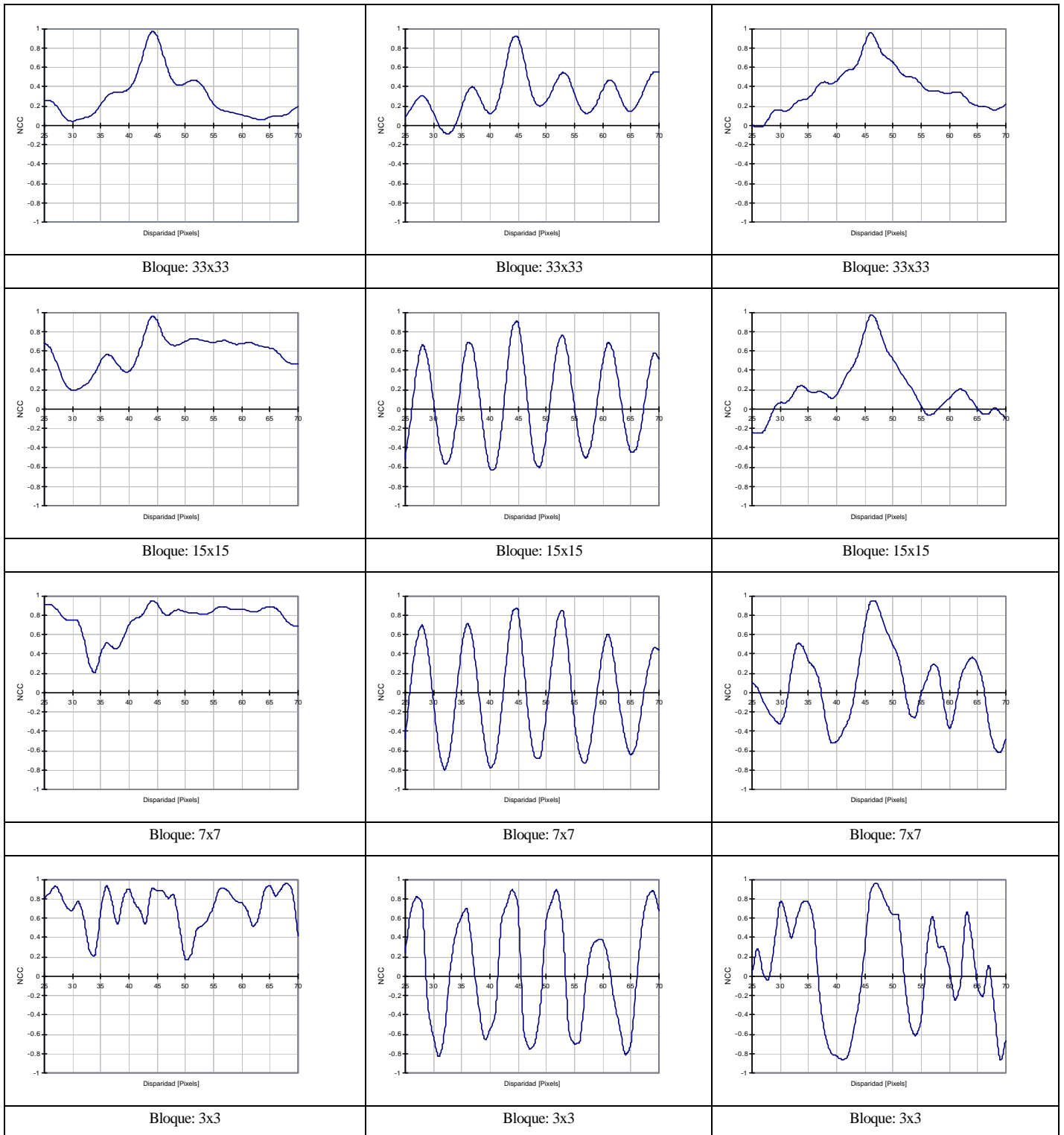


Figura 3.27: Comparación de las curvas de correlación para distintas texturas y bloques.

La ambigüedad en la correlación se puede expresar gráficamente a través de la figura 3.28, la cual muestra los *peaks* que aparecen al tratar de buscar el punto correspondiente al pixel de la región 2. La presencia de estos máximos locales complica la determinación de la disparidad correcta.

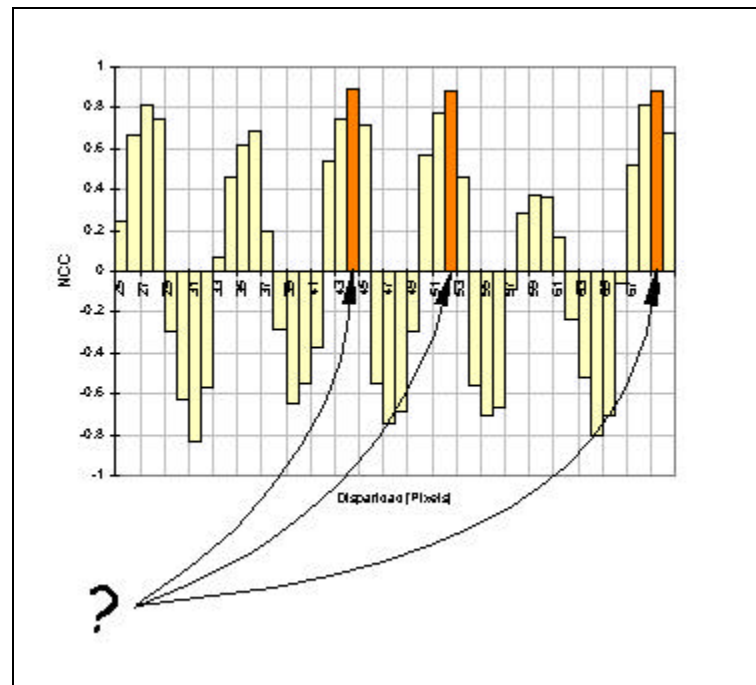


Figura 3.28: Ambigüedad en la búsqueda de correspondencias.  
(Región 2. Bloque de 3x3).

La región 3 corresponde a un sector de la imagen claramente identificable, con una textura suficiente y no periódica. Las curvas de correlación para el pixel en esta región también empiezan a mostrar otros *peaks* al reducir el tamaño del bloque; sin embargo, no alcanzan la magnitud del *peak* principal (ver tercera columna de la figura 3.27).

Con respecto a estos resultados se pueden hacer las siguientes observaciones:

- Al reducir el tamaño del bloque de correlación aumenta la cantidad de *peaks* en las curvas, por lo tanto las disparidades se determinan con un grado mayor de ambigüedad y menor confiabilidad.
- Si el rango de búsqueda de disparidades se reduce al realmente requerido: [40-50], en vez de [25-70], entonces varios máximos locales quedarán fuera del proceso de selección, y por lo tanto las disparidades encontradas corresponderán a la correcta con un mayor grado de certeza. La base de la pirámide se encuentra a 700 mm de la cámara y tiene una altura no superior a los 50 mm, razón por la cual el rango teórico de disparidades debería estar entre [45-49], siendo innecesaria la búsqueda en un rango tan amplio como [25-70] (correspondiente a un rango de distancias de 457 mm a 1280 mm). Sin embargo en la práctica las distancias no se conocen a priori, ya que es lo que se quiere determinar, y por eso debe realizarse una búsqueda en un rango amplio.

Estas observaciones son muy importantes, porque justifican la introducción del procesamiento jerárquico y secuencial de los pares estéreo: El primer par de la secuencia debería procesarse utilizando una ventana de mayor tamaño, para el caso de Pira700, 33x33 es suficiente dado que las curvas de correlación para regiones complicadas como la 1 y 2, muestran un máximo único. Los pares siguientes de la secuencia deberían procesarse con ventanas de menor tamaño para aumentar la precisión sub-píxel, pero considerando un rango de búsqueda restringido en torno a valores obtenidos para el par anterior y no una búsqueda en el rango entero inicial, que es más amplio (requiere de mayor tiempo de procesamiento) y que facilita la ocurrencia de ambigüedades.

Las observaciones además indican la importancia de introducir alguna medida para determinar la ambigüedad de los resultados, especialmente cuando la curva de correlación presenta múltiples máximos de similar magnitud.

## B. Validación de las Correspondencias y Coeficiente de Confiabilidad

Un método muy efectivo para eliminar la mayoría de los errores de disparidad consiste en calcular la correlación dos veces, invirtiendo los roles del par estéreo, es decir realizando la correlación con la imagen derecha como referencia y buscando las correspondencias en la izquierda, y luego utilizando la imagen izquierda como referencia y buscando las correspondencias en la derecha. Ambos mapas de disparidad deberían ser iguales. Si hubiesen puntos correspondientes a una misma posición con valores de disparidad distintos, estos se pueden considerar como no válidos [Fua91, Faug93i]. Normalmente los puntos no válidos son aquellos que son visibles en una imagen, pero que no aparecen en la otra debido a las oclusiones que se originan en el cambio de perspectiva.

Algunos autores han propuesto el uso de un umbral de correlación, y aceptar sólo aquellas correspondencias que tengan un valor de correlación mayor al fijado mediante el umbral. La determinación del umbral es un tanto arbitraria y se puede demostrar que su uso no garantiza la confiabilidad de los resultados, en primer lugar porque según la expresión para la correlación de (3.44), ésta toma valores entre  $-1$  y  $+1$ , independientemente del tamaño de la ventana que se utilice. En la práctica esto no es totalmente cierto porque se utiliza como promedio para el punto  $I(x,y)$  de la imagen  $I$ , los valores en una vecindad que puede contener elementos ocluidos o que pertenecen a otro objeto, distorsionando el valor promedio. Por esta razón [Kana94], utiliza un coeficiente basado en la varianza de los pixels en la ventana para determinar el tamaño de bloque óptimo.

Por otro lado como se muestra en la figura 3.29, la correlación máxima para los puntos en las regiones analizadas de la figura 3.26, se mantiene aproximadamente constante y sobre 0.8, a pesar de la variación en el tamaño de ventana de correlación utilizada.

Sin embargo, por los resultados mostrados en la figura 3.30, que muestra la disparidad encontrada en función del bloque de correlación utilizado y las disparidades reales, se puede afirmar que para ventanas pequeñas la disparidad



encontrada tiene un error elevado (notar que la correlación máxima ocurre en los extremos del intervalo de disparidades para las regiones 1 y 3), aunque la correlación tome valores cercanos a 1 como se muestra en la figura 3.29. Por lo tanto, valores de correlación cercanos a 1 no necesariamente indican que la disparidad encontrada es válida o confiable.

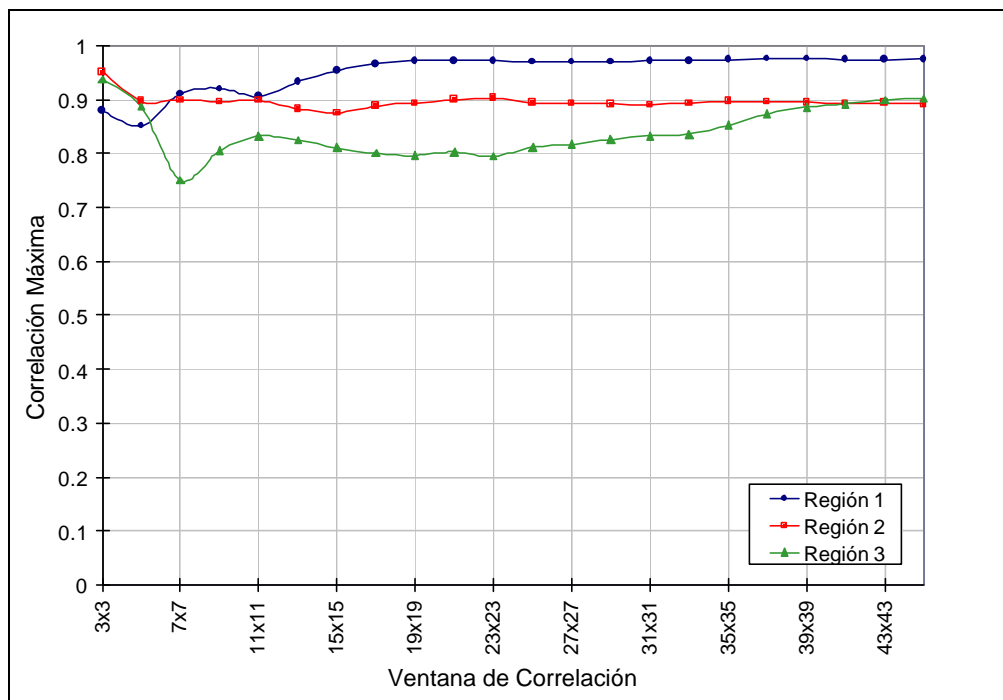


Figura 3.29: Correlación máxima en función del tamaño de ventana para los puntos de la figura 3.26.

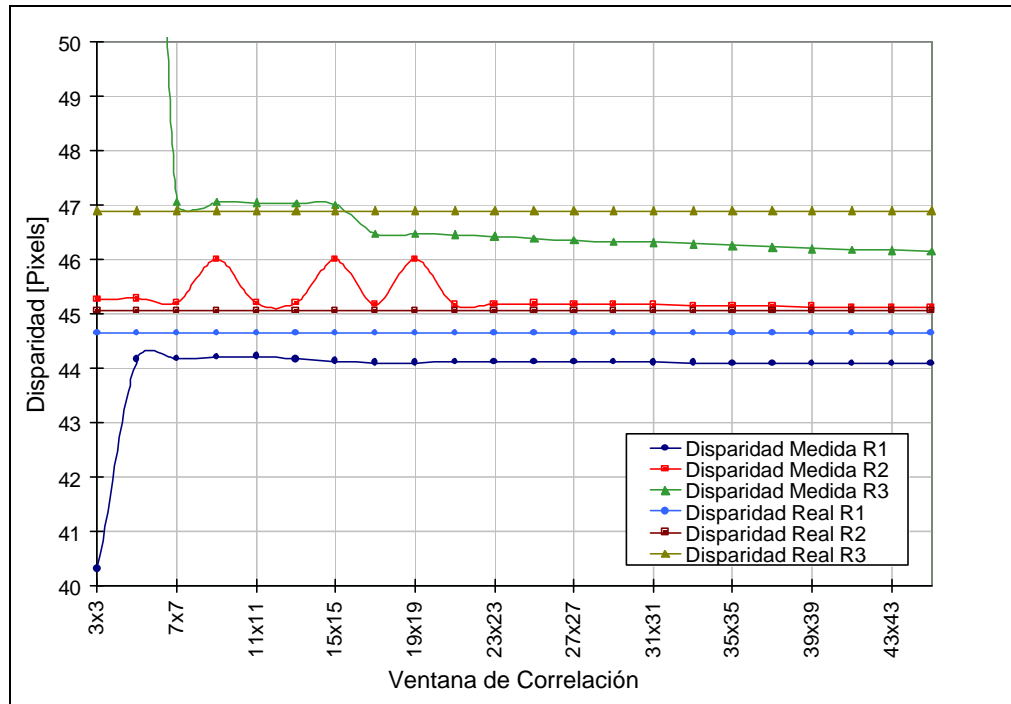


Figura 3.30: Disparidad en función del tamaño de ventana para la correlación máxima de los puntos de la figura 3.26.

Por estas razones, se propone el uso de un coeficiente de confiabilidad. Este coeficiente se calcula durante la etapa de búsqueda de correspondencias y se basa en el análisis de las curvas de correlación.

Las principales causas de error en la detección de disparidades son la oclusión, la presencia de texturas determinadas y de ruido en las imágenes. El error en el cálculo de la disparidad ocurre cuando un máximo incorrecto es seleccionado. Por lo tanto, si se detecta la presencia de varios máximos en la curva de correlación, o si la amplitud de la curva es muy reducida, podrá afirmarse que el riesgo de elegir un máximo incorrecto es alto. Basados en este razonamiento, se definieron los siguientes índices de confiabilidad:

$$W(x, y) = C_1(x, y) \cdot (C_1(x, y) - C_2(x, y)) \quad (3.46)$$

Donde  $C_1$  y  $C_2$  corresponden al máximo global y al mayor de los máximos locales de la curva de correlación (el segundo *peak* más alto) para el punto  $(x,y)$  del mapa de disparidades.

La determinación de (3.46) se hizo en forma heurística. El factor de confiabilidad  $W(x,y)$ , *valora* aquellas correspondencias que tienen un máximo global alto y único, por esta razón se multiplican el valor de la correlación máxima  $C_1$ , con la distancia que existe entre el máximo y el segundo máximo más cercano ( $C_1-C_2$ ). Si en la curva hay dos *peaks* iguales, entonces la diferencia entre el máximo y el segundo máximo es cero, y por lo tanto la disparidad encontrada no es confiable y debe eliminarse. Si el máximo es único, entonces la diferencia entre el máximo y el segundo *peak* es máxima, valorándose más este resultado. Pero aunque el máximo sea único, si éste no tiene un valor suficientemente alto, se considerará poco confiable.

En teoría el valor máximo de la confiabilidad podría ser 1, pero como en la práctica la correlación máxima no alcanza a superar valores cercanos a 0.99, fue necesario definir un factor de confiabilidad normalizado según:

$$W(x, y) = a \cdot C_1(x, y) \cdot (a \cdot C_1(x, y) - a \cdot C_2(x, y)) \quad (3.47)$$

$$\text{donde } a = \frac{1}{C_1(x, y)}.$$

De esta manera (3.47) se reduce a:

$$W(x, y) = 1 - \frac{C_2(x, y)}{C_1(x, y)} \quad (3.48)$$

El factor de confiabilidad normalizado (3.48) toma valores entre 0 (disparidades no válidas) y 1 (confiabilidad máxima). Las ventajas del uso de (3.48) se explicarán en la sección 3.3.4 sobre el procesamiento secuencial de los pares estéreo.

Al graficar el coeficiente de confiabilidad no normalizado en función del tamaño de ventana de correlación se puede apreciar que la confiabilidad es máxima al

utilizar bloques de correlación que varían entre  $21 \times 21$  y  $31 \times 31$  pixels, como se muestra en la figura 3.31.

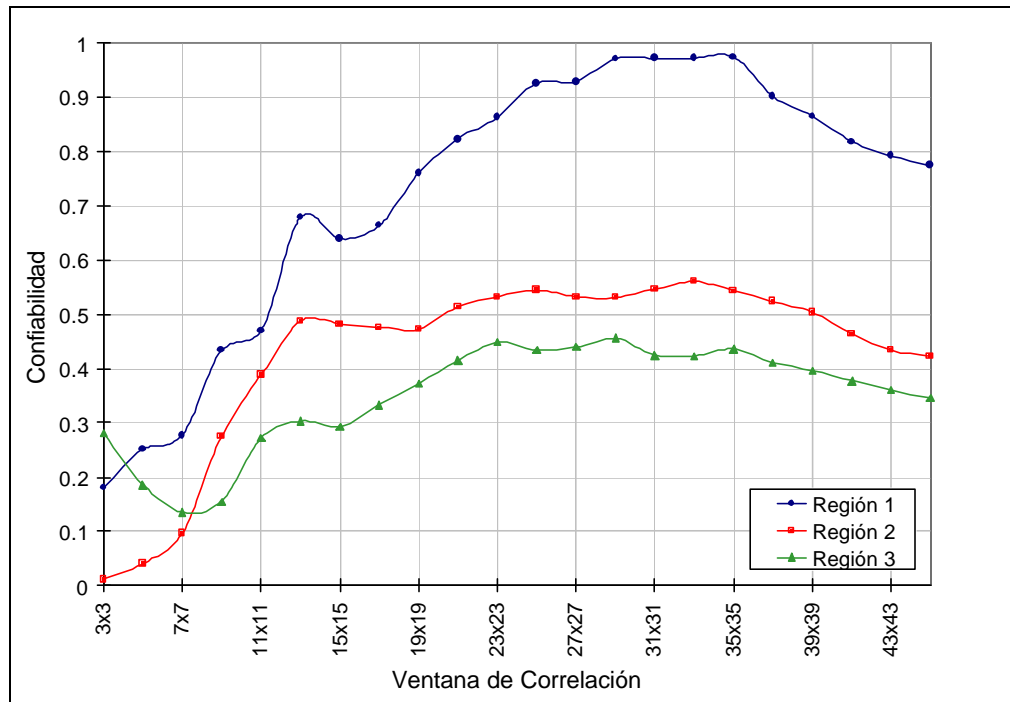


Figura 3.31: Confiabilidad en función del tamaño de ventana para los puntos de la figura 3.26.

En la figura 3.32 se muestran el mapa de disparidades, validadas invirtiendo el rol de los pares, y el mapa de confiabilidades en la primera y segunda columna, respectivamente. En la tercera columna de la figura 3.32, se muestran las disparidades obtenidas sólo para aquellos puntos de la escena cuyo respectivo factor de confiabilidad normalizada (según (3.48)) fue igual a 1.

Puede notarse que la confiabilidades son altas para la mayoría de los puntos de la escena cuando se utilizan ventanas de correlación grandes ( $33 \times 33$  y  $15 \times 15$ ), exceptuando los segmentos horizontales de poca textura en los bordes de la base de la pirámide. Sin embargo, el número de puntos confiables disminuye

drásticamente al reducir el tamaño del bloque de correlación, como se aprecia en los mapas de confiabilidad para ventanas de  $7 \times 7$  y  $3 \times 3$ .

También se aprecia que al disminuir el tamaño de la ventana de correlación aumenta el número de puntos que no aprueba la validación realizada invirtiendo los roles de las imágenes del par. Los puntos no válidos se muestran en color rojo oscuro sobre los mapas de disparidades de la figura 3.32.

Pero la desventaja de usar ventanas demasiado grandes es que la precisión de los resultados de reconstrucción tridimensional disminuye ya que los bloques de correlación grandes tienen un efecto de filtro pasabajos. Como puede observarse en la primera columna de la figura 3.32, la ventaja de utilizar ventanas pequeñas se encuentra en el nivel de detalle que puede obtenerse de la escena (las aristas de la pirámide no se vuelven *borrosas* como al utilizar bloques de  $15 \times 15$  o  $33 \times 33$ ).

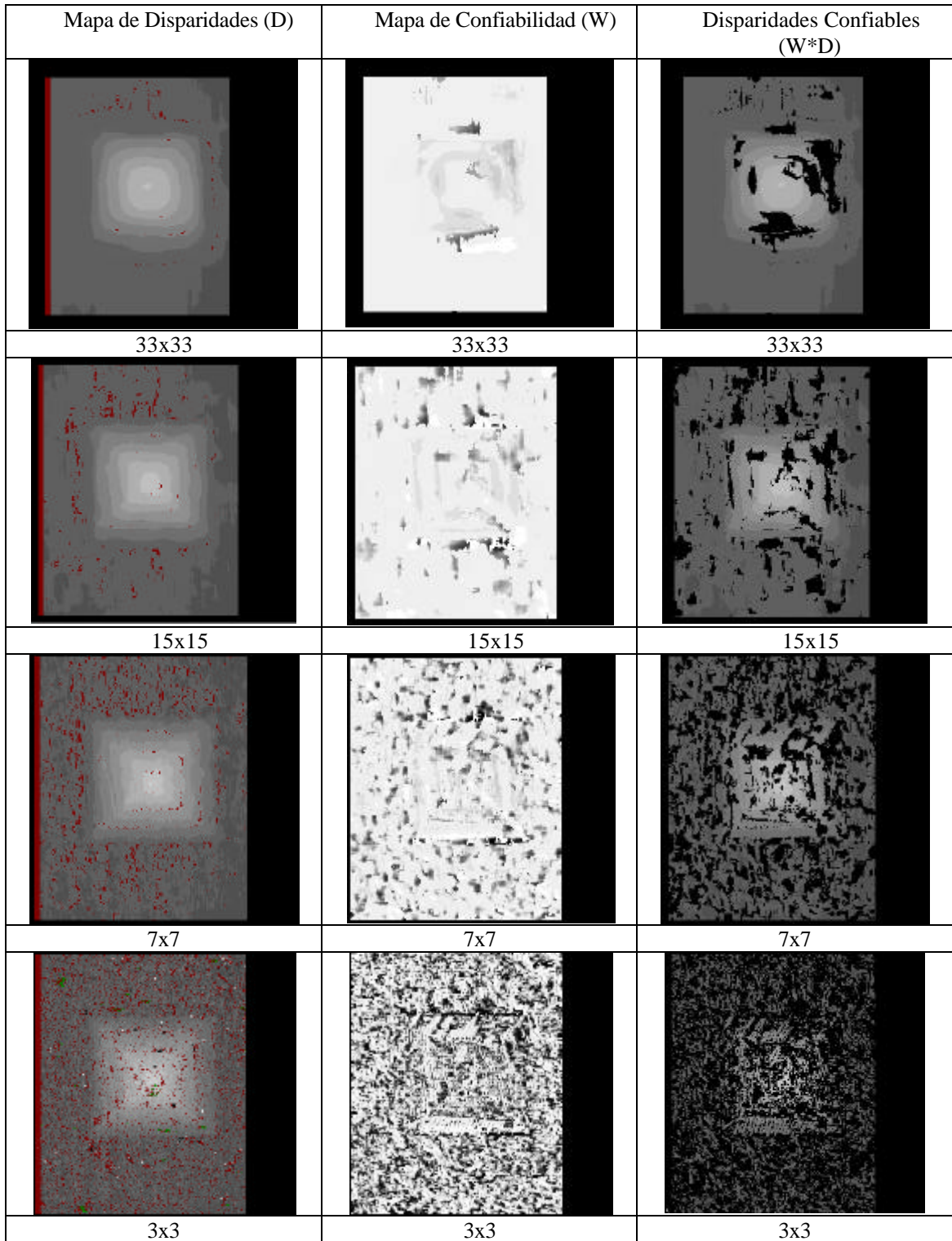


Figura 3.32: Validación y confiabilidad para distintos tamaños de ventana de correlación.

### 3.3.3 Interpolación

La mayoría de algoritmos de correlación deben realizar una etapa de interpolación para atenuar posibles errores de correspondencia o para rellenar aquellas zonas para las cuales no pudo calcularse su disparidad. Estas zonas inevitablemente ocurren en la práctica, debido a texturas que causan ambigüedad en la búsqueda de correspondencias o a la escasez de ella (e.g. paredes de ladrillos o paredes lisas, respectivamente).

Una interpolación sencilla consiste en utilizar un filtro de mediana, frecuentemente utilizado para eliminar ruido impulsivo (ruido que no tiene una distribución uniforme o una distribución Gaussiana, sino aquel que toma valores en los extremos del rango de cuantización). En la figura 3.33 se muestra un mapa de disparidades calculado para el primer par de Pira700 utilizando un bloque de correlación de  $3 \times 3$ . Se puede apreciar que hay una gran cantidad de puntos no correlacionados (en color negro) y algunos puntos cuya correlación máxima ocurrió para la máxima disparidad (en color blanco). En este caso los puntos no correlacionados pueden considerarse como ruido impulsivo, y pueden eliminarse bastante bien utilizando un filtro de mediana. Utilizar este filtro equivale a aplicar una restricción sobre el gradiente de disparidades máximo permitido. En Pollard, Mayhew y Frisby [Faug93], se utiliza una restricción que limita el valor absoluto del gradiente del mapa de disparidades a un umbral máximo, evitando que ocurran cambios demasiado repentinos.

La ventaja de un filtro no lineal como la mediana, frente al uso de filtros pasabajos o de promedios ponderados, es que introduce una distorsión menor en los bordes de los objetos de la escena.

Otros métodos de interpolación filtran el mapa de disparidades incorporando información sobre las sombras en la escena [Fua91]. El supuesto en que se basan estos métodos es que la luminosidad cambia donde se producen discontinuidades en la escena, y por lo tanto, incorporando la información de la variación de la luminosidad se podrían interpolar superficies, sin suavizar los bordes de los objetos en la escena (i.e. manteniendo discontinuidades).

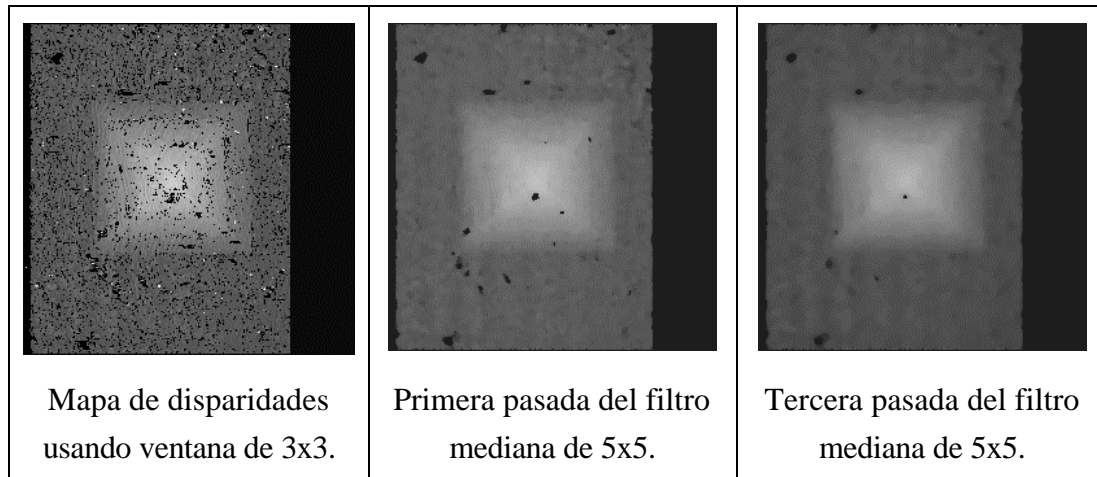


Figura 3.33: Eliminación de puntos no válidos usando filtro de mediana.

En la práctica, lo más frecuente es usar algoritmos de interpolación basados en la minimización de diferencias al cuadrado, bajo restricciones de continuidad y curvatura, como la interpolación cúbica o la interpolación mediante *splines* [Wolb92]. Dentro de éste tipo de algoritmos se encuentra el de [Szel90], que trata de minimizar la energía de unos resortes que unen una superficie tipo membrana a los datos reales. Este método tiene la ventaja de que las constantes de los resortes pueden ser modificadas permitiendo un ajuste menos o más ceñido a los datos interpolados.

La desventaja de la interpolación es que tiende a reducir la precisión de las mediciones, especialmente donde ocurren las discontinuidades en los elementos de la escena [Moha89, Hoff89]. Por esta razón no se aplica una interpolación a la superficie entera. Recordemos que uno de los objetivos del algoritmo secuencial propuesto es completar los vacíos de los mapas de profundidad en forma precisa. Sin embargo, se propone la interpolación local de la superficie en las áreas para las cuales no se halló una correspondencia válida.

Se implementó la interpolación lineal de las áreas no correspondidas de la escena, utilizando como puntos de referencia, los cuatro vecinos más cercanos para los cuales se encontró una disparidad. Este método, mostró ser muy eficiente para



completar los vacíos de información que ocurren en las zonas ocluidas y aquellas con texturas no apropiadas. Gráficamente el esquema de interpolación bilineal propuesto se muestra en la figura 3.34.

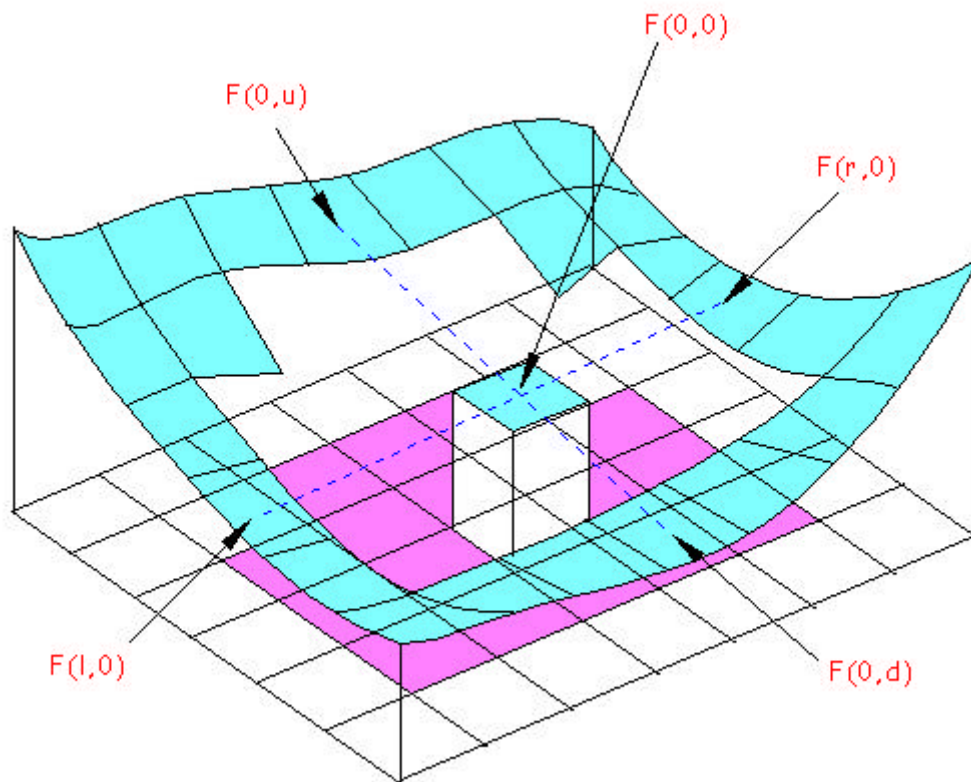


Figura 3.34: Interpolación bilineal en las zonas no correspondidas.

Al buscar las correspondencias los pixels no correlacionados se marcan de modo que puedan ser identificados en las etapas posteriores del procesamiento (pixels violeta en la figura 3.34). La interpolación se realiza tomando en cuenta los cuatro vecinos más cercanos al pixel marcado en las direcciones Norte, Sur, Oeste, Este ( $(0, u)$ ,  $(0, d)$ ,  $(l, 0)$  y  $(r, 0)$ , respectivamente). Se ponderan las disparidades de los pixels vecinos en forma proporcional a su distancia con respecto al pixel cuya disparidad se desea estimar. La expresión para la interpolación bilineal utilizada es:

$$F(0,0) = \frac{r \cdot F(l,0) + l \cdot F(r,0) + d \cdot F(0,u) + u \cdot F(0,d)}{l + r + u + d} \quad (3.49)$$

En la figura 3.35 se muestra la interpolación de las zonas no correlacionadas del par estéreo de la figura 2.6. El método permite completar en forma satisfactoria los vacíos sobre el fondo y sobre el muñeco. El filtro de mediana aplicado sobre el mapa de disparidades interpolado permite reducir los cambios repentinos que no pudieron ser eliminados en la etapa de validación.

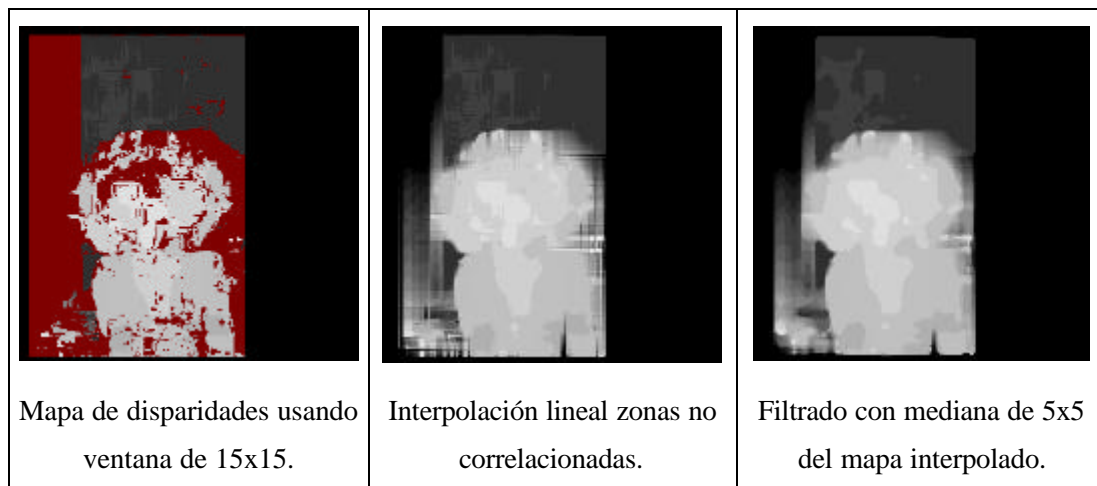


Figura 3.35: Aplicación del método de interpolación bilineal al mapa de disparidades de la figura 2.6.

### 3.3.4 Procesamiento de la Secuencia de Pares Estereoscópicos: Fusión de la Superficie con el Modelo 3D

Al utilizar una secuencia de pares estéreo tomados desde distintas posiciones, las disparidades calculadas para un par de imágenes pueden emplearse para predecir las disparidades (o las distancias) del par siguiente de acuerdo al movimiento que realizó el sistema binocular en el espacio. De este modo, puede restringirse el rango de búsqueda de disparidades para el par siguiente a un rango adaptado de acuerdo al movimiento y a la información sobre las disparidades encontradas previamente. Al reducir el rango de búsqueda, deberían disminuir los

errores producto de las ambigüedades que ocurren en intervalos de búsqueda muy amplios; y además debería reducirse el tiempo de procesamiento

El considerar en forma novedosa la integración de los mapas de profundidad de cada par de la secuencia mediante un filtro FIR no lineal permite lograr resultados más precisos y densos. Las disparidades de cada par de la secuencia se obtienen con bloques de correlación cada vez más reducidos; esta estrategia tiene por finalidad encontrar una estimación confiable, aunque de menor precisión, para guiar la búsqueda en las etapas siguientes con ventanas pequeñas que permiten una mayor precisión, pero que no permiten obtener resultados confiables en la primera iteración.

Debe recordarse que el trabajo se desarrolló bajo varios supuestos, los más importantes de todos son: (1) La secuencia de imágenes se tomó con cámaras paralelas calibradas, razón por la cual las imágenes no requieren de una rectificación. Esto simplifica enormemente el proceso de búsqueda de correspondencia a líneas de barrido horizontal. (2) Se tiene un *feedback* preciso de los movimientos realizados por las cámaras.

A continuación se describen las técnicas de procesamiento secuencial para reconstrucción tridimensional implementadas.

### A. Procesamiento IIR no lineal

La fusión de los mapas de disparidad de la secuencia en este caso se realiza empleando la ecuación de un filtro IIR de primer orden, cuyos coeficientes están dados por los elementos del mapa de confiabilidad. La confiabilidad de la disparidad calculada para cada punto del par varía entre tomas, razón por la cual el filtro es no lineal. La estimación de la profundidad  $MZ_k$  en la etapa  $k$  está dada por:

$$MZ_k = W_k \cdot Z'_k + (1 - W_k) \cdot MZ_{k-1} \quad (3.49)$$

donde,  $W_k$  corresponde al mapa de confiabilidad de las disparidades del  $k$ -ésimo par estéreo y  $Z'_k$  corresponde a la profundidad estimada en la etapa actual corregida por el desplazamiento. Como el movimiento de las cámaras  $MOV_k$  es de translación,  $Z'_k$  puede expresarse según:

$$Z'_k = Z_k - \sum_{i=1}^k MOV_i \quad (3.50)$$

para referir  $Z'_k$  a  $Z_0$ . Las expresiones anteriores se utilizan para calcular la profundidad de cada punto en la imagen  $(x,y)$  correspondiente a un punto  $(X,Y)$  de la escena, pero por simplicidad en la notación, los índices correspondientes se han omitido (i.e.  $I = I(x,y)$  ).

Los coeficientes  $W_k$  corresponden a la confiabilidad normalizada y se calculan utilizando la ecuación (3.48). La ventaja de emplear la confiabilidad normalizada radica en que al tomar valores entre 0 y 1 permite eliminar los valores estimados antiguos  $MZ_{k-1}$  cuando la confiabilidad actual  $W_k$  de un punto es 1, ya que el término  $(1-W_k)$  será 0.

Sin embargo, el uso de este filtro no es apropiado si los objetos en la escena cambian de posición con respecto a las cámaras y entre sí, ya que se empeoraría la reconstrucción al fusionar una escena previa  $(k-1)$  distinta en su composición a la actual  $(k)$ .

## B. Procesamiento FIR no lineal

Para evitar los problemas del filtro anterior se propone utilizar un filtro FIR no lineal. Este sólo consideraría un número limitado de pares en la secuencia, según la ecuación:

$$MZ_k = \frac{\sum_{i=k-1-(S+1)}^{k-1} W_i \cdot Z'_i}{\sum_{i=k-1-(S+1)}^{k-1} W_i} \quad (3.51)$$

donde,  $S$  corresponde al número de pares de la secuencia utilizados para obtener las distancias estimadas a la escena  $MZ_k$ . El estimador de las distancias de cada punto expresado en la ecuación (3.51) corresponde a un filtro de promedio ponderado móvil con coeficientes variables. En este caso los coeficientes también se obtienen del mapa de confiabilidades calculado utilizando (3.48).

Los resultados de las reconstrucciones utilizando estos filtros se describen en el capítulo 4.

### 3.3.5 Ajuste del Rango de Disparidades

A partir de una estimación inicial de las disparidades de una escena  $d_0(x, y)$ , encontradas en un rango amplio de búsqueda con una ventana normalmente mayor a 5x5 pixels, se puede repetir el proceso de búsqueda en una segunda iteración sobre un rango de búsqueda adaptado localmente para cada punto de la imagen. Con esta estrategia no sólo se lograría reducir el tiempo de procesamiento en la siguiente etapa, sino también una reducción de los errores y un aumento de la confiabilidad de los resultados, ya que muchos de los *peaks* adyacentes al máximo quedarían fuera del nuevo rango restringido de búsqueda.

El rango local de búsqueda  $d_1(x, y)$ , para cada punto de la imagen puede estimarse como:

$$\begin{aligned} d_1(x, y)_{max} &= d_0(x, y) + \Delta d_{max} \\ d_1(x, y)_{min} &= d_0(x, y) - \Delta d_{max} \end{aligned} \quad (3.52)$$

donde,  $\Delta d_{max}$  corresponde al máximo error en las disparidades.

Como se mostrará en el capítulo 4, las pruebas realizadas permitieron establecer que el error promedio en la determinación de disparidades es de aproximadamente 1.5 pixels e incluso menor. Estos resultados concuerdan con los de [Falk97i], quien además sostiene que en las zonas de oclusión y discontinuidades este rango debería ser mayor.

Se propone en esta investigación utilizar una estimación del rango ajustado localmente para las zonas de discontinuidad o zonas no correlacionadas según:

$$\begin{aligned} d_1(x, y)_{max} &= d_0(x, y)_{max} + \Delta d_{max} \\ d_1(x, y)_{min} &= d_0(x, y)_{min} - \Delta d_{max} \end{aligned} \quad (3.53)$$

donde,  $d_0(x, y)_{max}$  y  $d_0(x, y)_{min}$  corresponden a la disparidad máxima y mínima más cercanas punto al indeterminado  $(x, y)$  sobre el segmento epipolar, respectivamente.

### 3.3.6 Mediciones y Reconstrucción Tridimensional

La reconstrucción de la superficie se realiza a partir de las relaciones derivadas en el capítulo 2: (2.16), (2.18) y (2.19), las cuales permiten estimar las distancias a cada punto en la escena y generar los mapas de profundidades que son integrados utilizando las ecuaciones (3.49) o (3.51). Además de la información de disparidades para cada punto en la escena, se requiere información sobre la geometría del sistema binocular, para poder recuperar la posición de los puntos en el espacio 3D en unidades métricas. La información de la geometría del sistema se obtiene de la etapa de calibración.

Una vez obtenido el modelo reconstruido final  $MZ_k$ , pueden realizarse distintas mediciones, como por ejemplo área, volumen, centro de masa, etc. Las mediciones que se realicen dependerán del objetivo de cada aplicación en particular. En esta investigación se realizan mediciones de:

- Distancias mínima y máxima al modelo.
- Volumen promedio.

Estas mediciones se discuten en el capítulo siguiente, y tiene por finalidad evaluar la calidad de la reconstrucción 3D.

## IV. RESULTADOS EXPERIMENTALES

### 4.1 Evaluación de los Algoritmos Estereoscópicos

El análisis de los resultados experimentales tiene por finalidad evaluar la calidad de los algoritmos y métodos de reconstrucción tridimensional implementados.

La mejor manera para establecer el éxito de un método de estereoscopía es comparar sus resultados con los de la información de distancias reales medidas mediante otros medios que no sean los estereoscópicos (e.g. *laser range finders*). La adquisición de información total y exacta es costosa, y por consiguiente prohibitiva. Imágenes sintetizadas (escenas creadas artificialmente) pueden entregar mucha información sobre distancias reales, aunque carecen de las complicaciones de las imágenes tomadas a escenas reales. Por esta razón se utilizarán dos tipos de imágenes:

- Imágenes de una escena artificial generadas con el computador. En este caso se conoce con exactitud el mapa de disparidad, y la precisión del algoritmo de reconstrucción puede ser estimada.
- Imágenes de prueba adquiridas en laboratorio de objetos a distancias y con dimensiones conocidas. Estas imágenes permiten una evaluación principalmente cualitativa del proceso de reconstrucción tridimensional de escenas reales.

La evaluación de cualquier método experimental requiere de la caracterización de la exactitud y precisión de sus resultados. Por *exactitud* se entiende la concordancia o similitud entre el valor real y el resultado obtenido. La *precisión* es el número significativo de dígitos del resultado. Decir que la disparidad es de 1.44 [pixels] puede ser muy preciso, pero probablemente poco exacto. En cambio, el resultado podría entregarse con mayor exactitud, aunque con menor precisión, como  $0.77 \pm 0.85$  [pixels].

Para la visualización tridimensional de los resultados se generaron superficies enmalladas (también conocidas como *mesh* o *wireframe*). Además se presentan algunas superficies tridimensionales con la textura de la escena *mapeada*

(*texture mapping*) no sólo para dar una idea de como se vería la escena, sino también para poder realizar un análisis cualitativo.

#### 4.1.1 Imágenes Sintéticas

Las imágenes sintéticas se generan aplicando una textura a una cierta superficie tridimensional cuyas dimensiones son conocidas y luego realizando una proyección en perspectiva del modelo tridimensional sobre el plano imaginario correspondiente a cada cámara. Al realizar la proyección inversa se pueden determinar todas las disparidades con exactitud y sin errores, y emplearlas para medir la precisión de los resultados de estimación de la posición para cada punto de la escena.

Se utilizaron dos pares estéreo sintéticos. El primero se generó a partir de un *objeto artificial* (que podría ser la pieza de una máquina, por ejemplo). Las dimensiones del objeto se detallan en el diagrama de la figura 4.1.

Los pasos para crear el par sintético se resumen en:

- Generación de la superficie de coordenadas  $X_b, Y_b, Z_b(X_b, Y_b)$  (con resolución de 1024x1024 puntos) relativa al sistema de coordenadas de la cámara B.
- Proyección sobre la cámara B de la superficie  $Z_b(X_b, Y_b)$ .
- Proyección sobre la cámara A de la superficie  $Z_b(X_b, Y_b)$ , considerando la translación entre los sistemas de coordenadas definida por el *baseline* según:

$$X_a = X_b + b, Y_a = Y_b, Z_a(X_a, Y_a)$$

- Determinación de la disparidad y generación del mapa de disparidad.

Las ecuaciones de proyección necesarias son:

$$u_b = \frac{f}{r_x} \cdot \frac{X_b}{Z_b(X_b, Y_b)} \quad u_a = \frac{f}{r_x} \cdot \frac{X_b + b}{Z_b(X_b + b, Y_b)}$$

(4.1)

Estas corresponden a la aplicación de (2.3) y permiten definir la disparidad para cada punto utilizando la relación:  $d(u_b, v_b) = u_a - u_b$ .



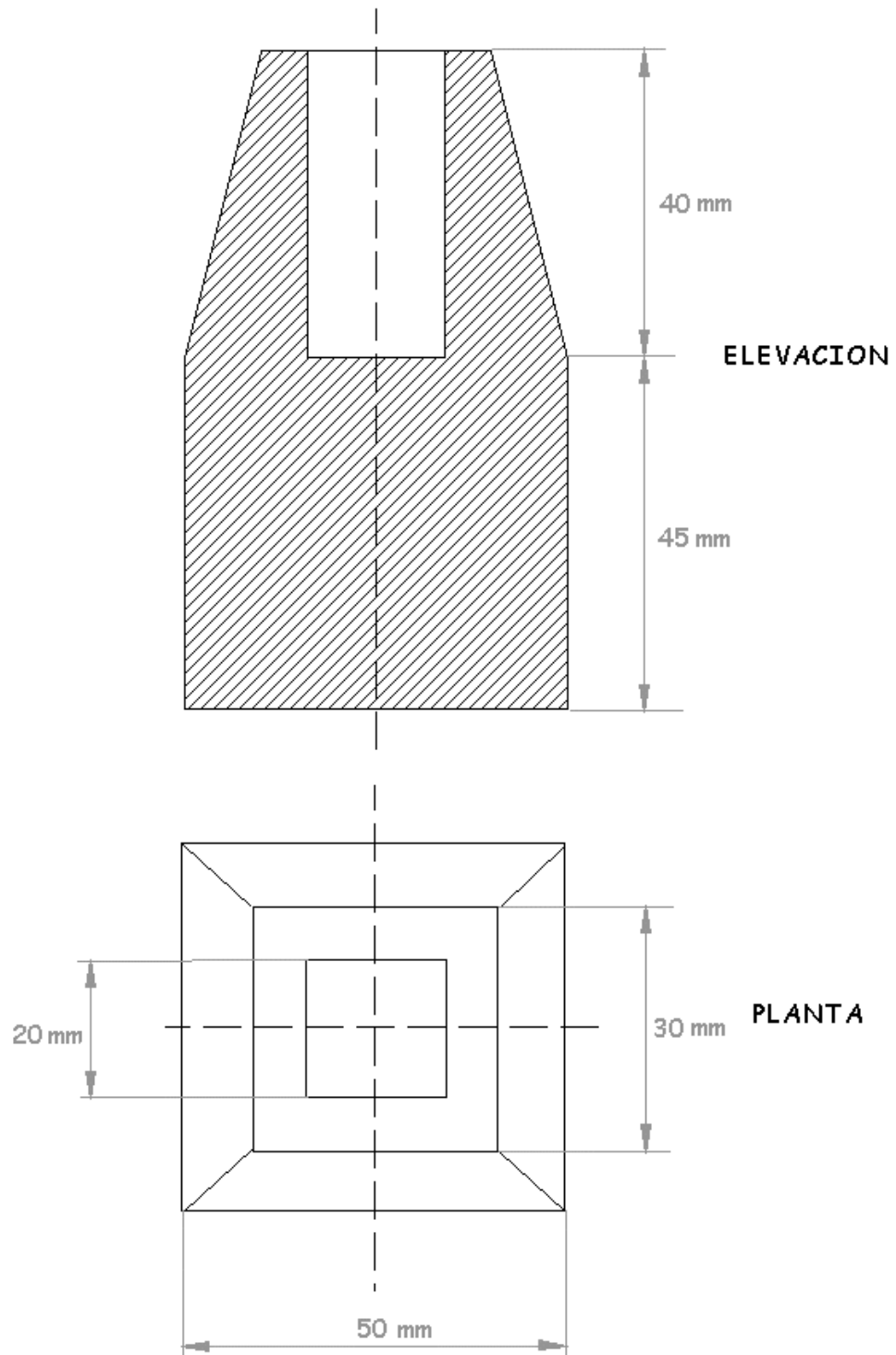


Figura 4.1: Vistas de elevación y planta del modelo artificial.

Al proyectar cada punto del modelo sobre los planos imaginarios de las imágenes se utilizó como textura un patrón aleatorio cuya intensidad se definió a partir de una distribución uniforme  $U \sim U[-100+Z, 100+Z]$ , de media proporcional a la altura de la superficie. Este rango de valores se eligió para permitir una mejor visualización de la textura en la imágenes cuantizadas en 256 niveles.

Los parámetros del sistema binocular paralelo simulado comprenden: un *baseline* de 50 mm, distancia focal 16mm, resolución del sensor de 256x256, tamaño de las celdas de 0.01 mm/pixel. Distancia máxima al objeto 640 mm. Dados estos parámetros, las disparidades varían entre 10 y 20 pixels.

El par estéreo creado mediante el procedimiento descrito y su mapa de disparidades se muestran en la figura 4.2 y 4.3, respectivamente.

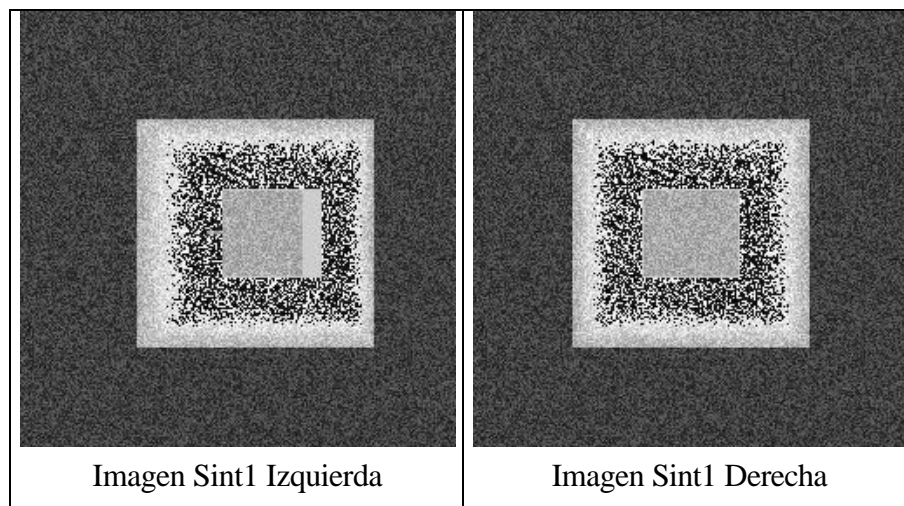


Figura 4.2: Par estéreo sintético Sint1.

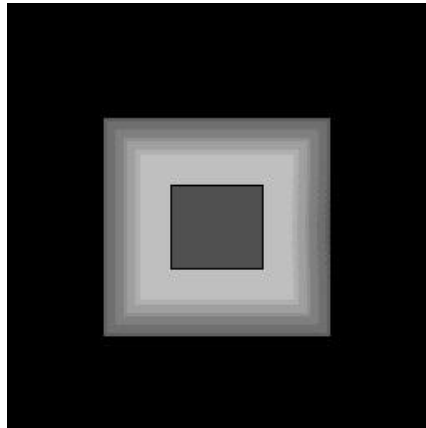


Figura 4.3: Mapa de disparidad correspondiente al par Sint1.

La representación tridimensional de la superficie del modelo Sint1 se muestra en la figura 4.4. El objeto fue creado intencionalmente con discontinuidades marcadas y textura no completamente aleatoria para elevar el grado de complejidad que deberá enfrentar el algoritmo de reconstrucción, a un nivel similar al que encontraría en la reconstrucción de un objeto real.

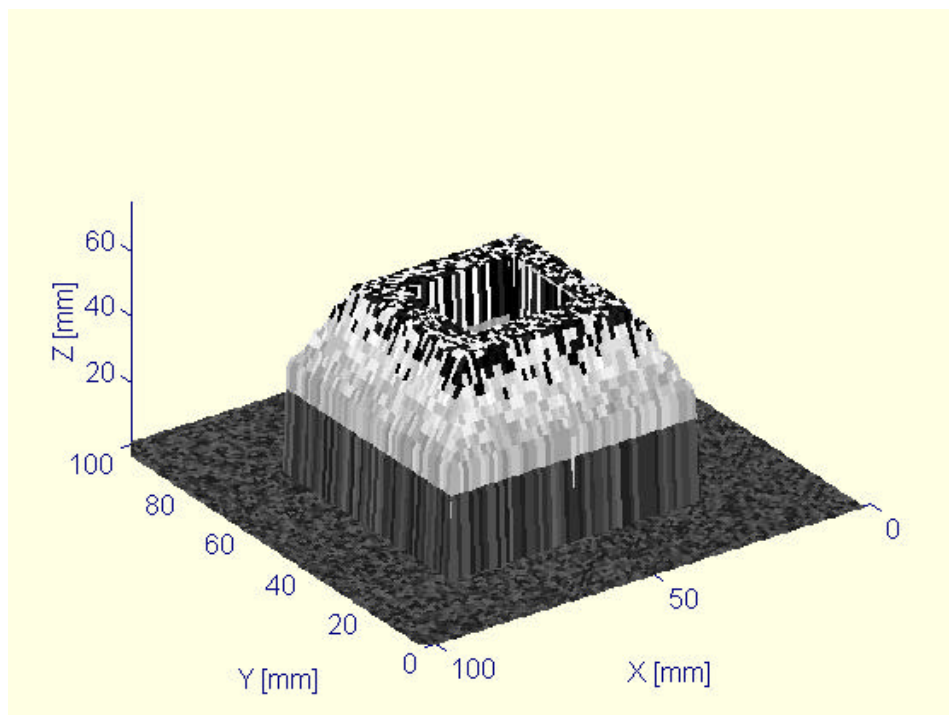


Figura 4.3: Modelo 3D del objeto en Sint1.

El segundo par estéreo se muestra en la figura 4.5, fue generado mediante un programa de *ray-tracing* y se obtuvo en [web17]. Este par estéreo simula un pasillo con objetos que se pueden distinguir fácilmente, como una esfera y un cono.

En este caso, aunque no se cuenta con información del sistema binocular necesaria para la recuperación de las distancias en unidades métricas, se tiene el mapa de disparidades (figura 4.6), también generado con precisión de punto flotante. Comparando el mapa de disparidades reales con aquellas obtenidas por los algoritmos desarrollados podrá evaluarse el nivel de error en la estimación de las disparidades.

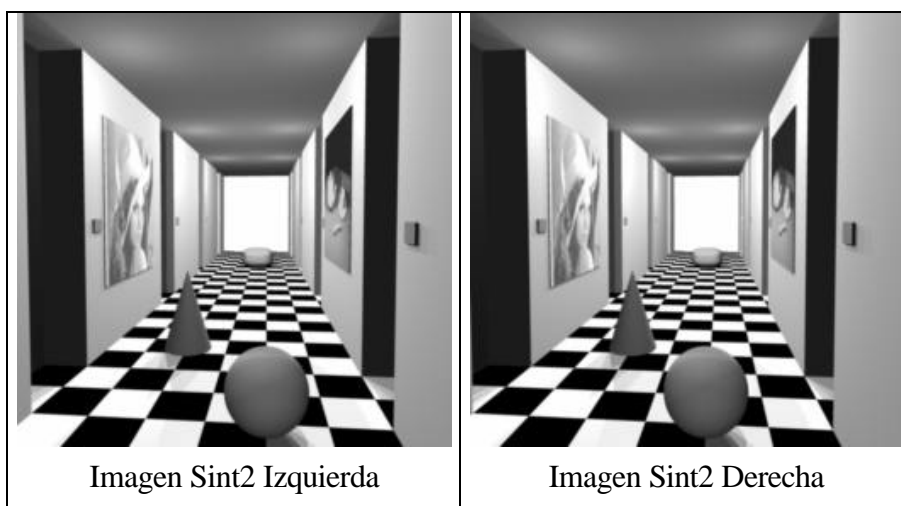


Figura 4.5: Par estéreo sintético Sint2.

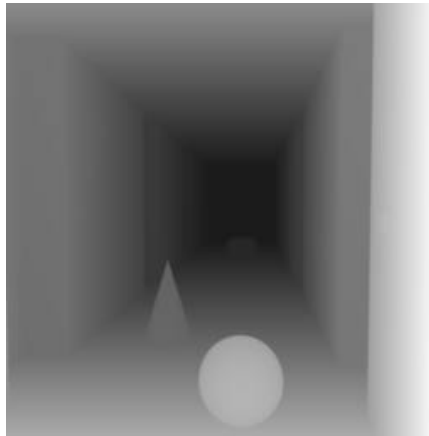


Figura 4.6: Mapa de disparidad correspondiente a Sint2.

Los resultados de la aplicación de los algoritmos a los pares Sint1 y Sint2 se encuentran resumidos en la tabla 4.1. Se aplicó un procesamiento después de la etapa de búsqueda de correspondencias, el cual llamaremos *procesamiento post-matching*. En la tabla se denotan con la iniciales I, M y V los procesamientos de interpolación en las zonas no correspondidas, filtraje con mediana y validación, respectivamente (explicados en las secciones 3.3.2 y 3.3.3). En todos las pruebas se validaron los resultados, en algunas pruebas además se emplearon las etapas de interpolación y filtraje con mediana.

De los resultados obtenidos para el primer par, Sint1, se destacan las siguientes observaciones:

- Al utilizar la ventana de correlación grande (15x15) el error promedio en la determinación de las disparidades es 1.8 veces más grande, que al utilizar el bloque de 7x7. Este hecho se debería a que el modelo posee discontinuidades abruptas y que al utilizar el bloque de 15x15, éstas se suavizan.
- La aplicación del ajuste de rango de búsqueda a partir de una etapa de búsqueda inicial permite reducir el tiempo de procesamiento hasta un 23%. La prueba 1 de la tabla 4.1 requirió de 6.1 segundos al buscar las disparidades de cada pixel en el rango 0-25, mientras que sólo necesitó de 4.7 segundos en la prueba 2 para el rango ajustado a partir de una estimación inicial utilizando un bloque de 15x15 (rango A/15x15). Esta disminución de un 23% del tiempo de procesamiento

también se puede apreciar entre la prueba 5 y la 6, que requirieron de 6.5 y 4.9 segundos, respectivamente.

- La interpolación de las zonas no correspondidas permite reducir el error medio y su desviación estándar de manera significativa, sin un costo en el tiempo de procesamiento elevado. El tiempo de procesamiento sólo se incrementa en un 3.2%, mientras que el error promedio disminuye en un 27.8%.
- La aplicación del filtro de mediana puede tener un costo de procesamiento más elevado entre 2 a 5 segundos adicionales dependiendo de si se utiliza un bloque de 5x5 o 7x7. Sin embargo, el error promedio puede disminuir adicionalmente en un 26%, como lo demuestran los resultados de las pruebas 5 y 8.
- El error promedio en las disparidades puede alcanzar valores inferiores a 1 pixel. En la prueba 10 se utilizó un bloque de correlación de 5x5 para la búsqueda en el rango local ajustado. Además se aplicaron las etapas de validación, interpolación y doble pasada de mediana 5x5, obteniéndose un error medio de 0.73 pixels. Si suponemos que el error tiene una distribución normal, entonces podemos afirmar que, el error medio estará en el intervalo  $[0.73-2.13, 0.73+2.13]$  con una confianza del 68.3% ( $P\left(\left\|\frac{x - m}{s}\right\| \leq s\right) = 0.683$ ).

Tabla 4.1: Evaluación de los Algoritmos Estereoscópicos.

Prueba	Imagen	Ventana	Rango de Búsqueda	Procesamiento Post-Matching	Error Promedio de la Disparidad [pixels]	Desviación Estándar de la Disparidad [pixels]	Tiempo de Procesamiento [s]
1	<b>Sint 1</b>	15x15	0-25	V	2.05	4.30	6.1
2		15x15	A/15x15	V	1.92	4.09	4.7
3		15x15	0-25	VI	1.48	3.34	6.3
4		15x15	A/15x15	VI	1.47	3.34	4.7
5		7x7	0-25	VI	1.15	2.41	6.5
6		7x7	A/7x7	VI	1.14	2.94	4.9
7		7x7	A/15x15	VI	1.15	3.07	5.0
8		7x7	A/7x7	VI-M7	0.84	2.43	10.1
9		5x5	A/5x5	VI-M5	0.81	2.27	7.1
10		5x5	A/5x5	VI-2M5	0.73	2.13	9.1
11	<b>Sint 2</b>	21x21	0-10	VI	0.69	0.74	3.0
12		21x21	0-10	VI-M5	0.65	0.61	5.0
13		15x15	0-10	VI	0.76	0.94	3.2
14		15x15	A/15x15	VI	0.83	1.13	2.9
15		15x15	A/21x21	VI	0.69	0.76	2.9
16		7x7	A/15x15	VI-M5	0.87	1.20	5.2

Al aplicar la textura de la imagen derecha del par estéreo Sint1 sobre el mapa de alturas como se muestra en la figura 4.7, se puede obtener una representación más realista de objeto, muy similar a la que se presentó en la figura 4.3, correspondiente al objeto original.

La representación tridimensional del resultado de la prueba 10 empleando Sint1 se puede comparar a la representación tridimensional del mapa de disparidades reales como se muestran en las figuras 4.9 y 4.8, respectivamente.

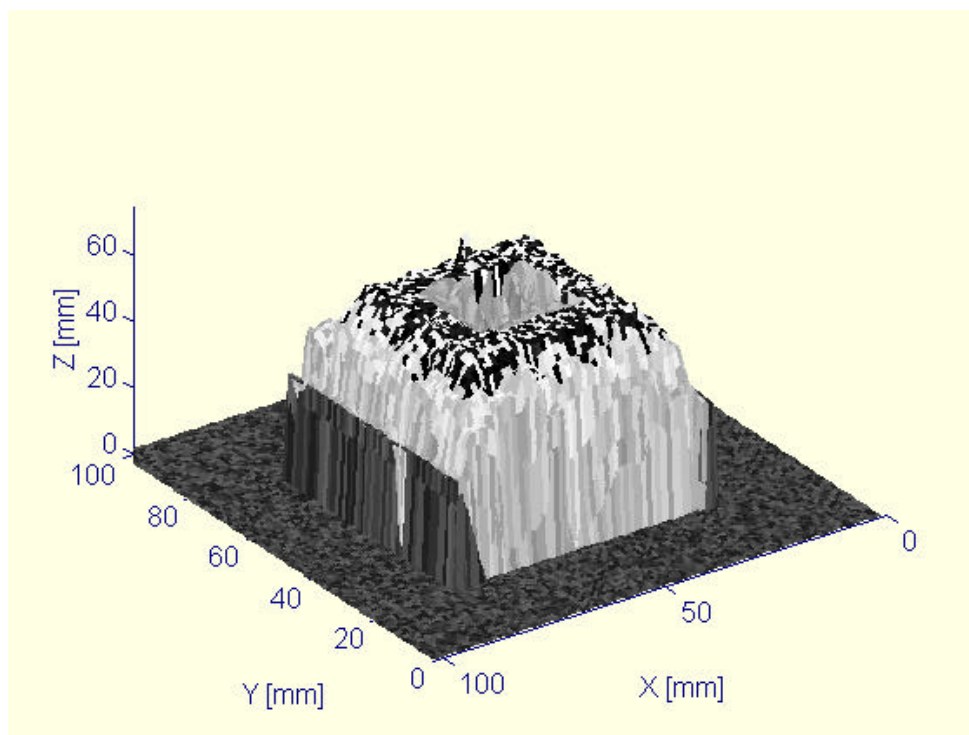


Figura 4.7: Representación 3D del objeto Sint1 reconstruido.

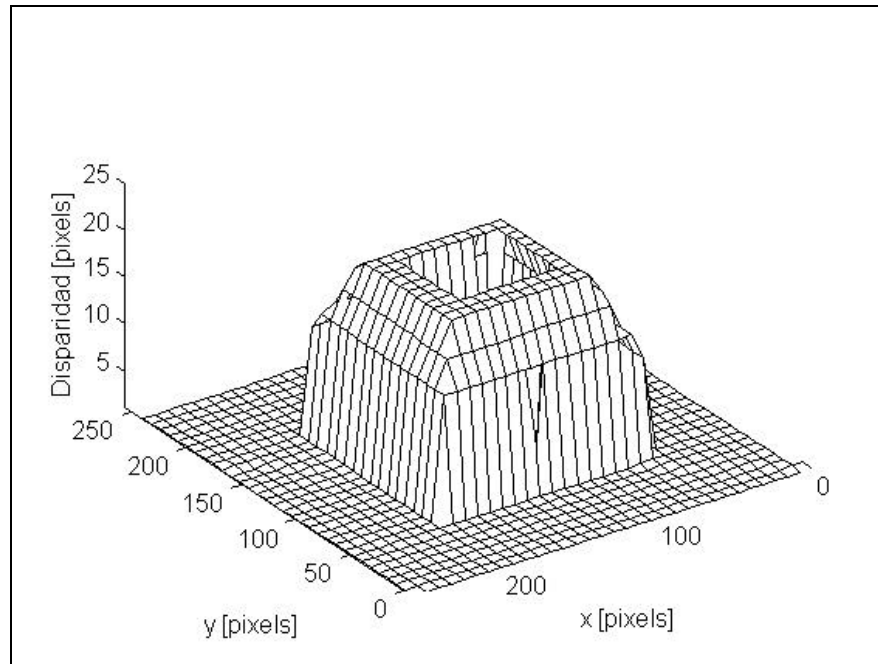


Figura 4.8: Representación tridimensional del mapa de disparidades reales de Sint1.

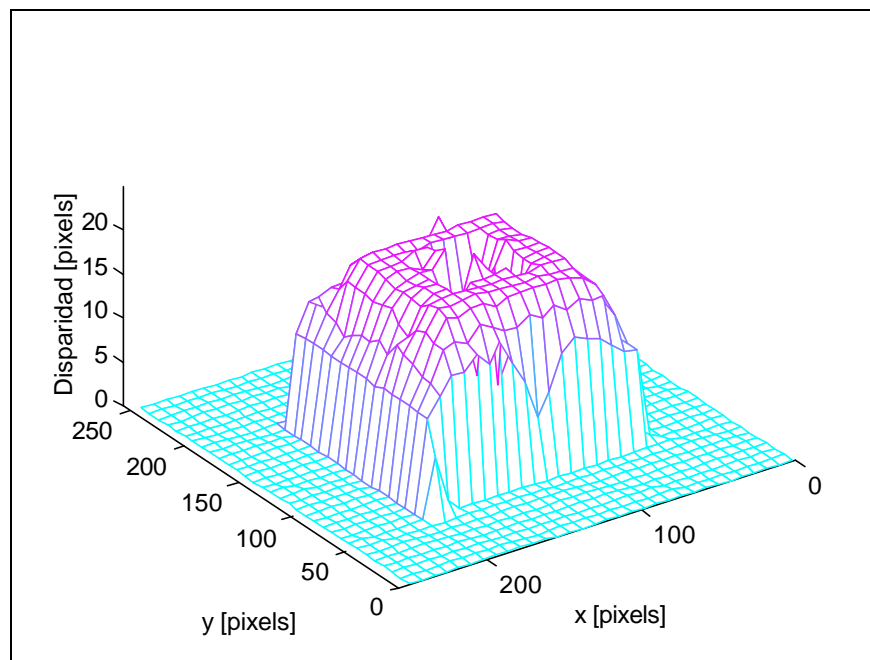


Figura 4.9: Mapa de disparidades medidas en la prueba 10 a partir del par Sint1.



A pesar de la escasa textura de la escena en el par Sint2, el algoritmo logra resultados con un error incluso menor que aquellos obtenidos al procesar Sint1, como se aprecia en la tabla 4.1. Esto se debe a que las superficies de la escena son suaves y no presentan mayores discontinuidades. El error medio es de 0.65 pixels para el mejor caso (prueba 12), con una desviación estándar de 0.61 pixels. Los tiempos de procesamiento requeridos por Sint2 son menores que aquellos requeridos por Sint1 debido a que el rango de búsqueda de disparidades es menor (0-10). Sin embargo, para este par sintético la reducción del tiempo de procesamiento es tan solo del 9.4% entre la prueba 13 y la 14, o la 15. Esto se debe a que el rango inicial de búsqueda ya es bastante pequeño y no puede reducirse a una vecindad inferior a los 4 pixels.

En la figura 4.10 se muestran los mapas de disparidad resultantes de la búsqueda de correspondencias utilizando bloques de 15x15 sin interpolación de las zonas no válidas (color rojo oscuro) y con interpolación. Se puede afirmar que la interpolación local de las áreas rojas permite completar los vacíos de información en forma efectiva (reducciones del error medio del orden del 25%).

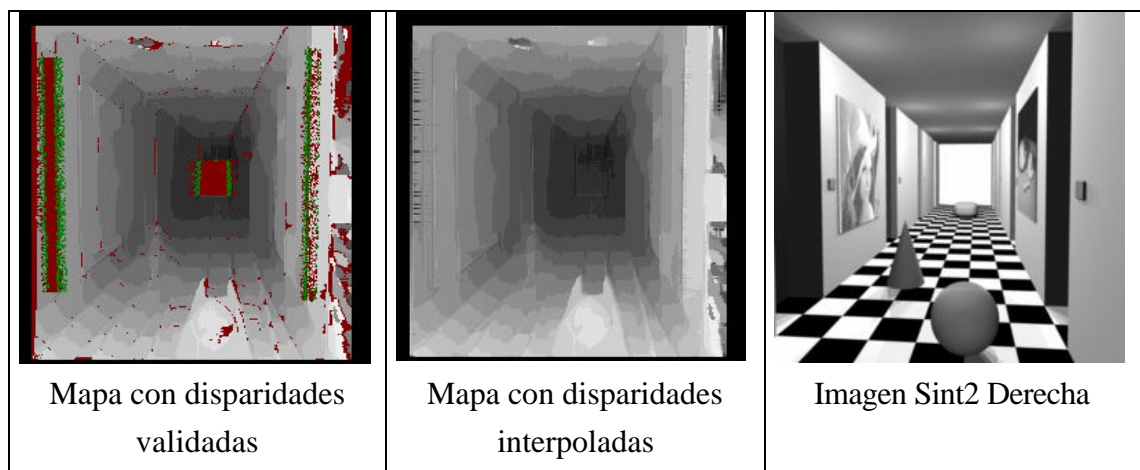


Figura 4.10: Resultados de la búsqueda de correspondencias empleando un bloque de 15x15, con validación e interpolación.

La figura 4.11 muestra el mapa de confiabilidad de los resultados obtenidos en la primera iteración (imagen izquierda), y el mapa de confiabilidad que se logra al ajustar el rango de búsqueda de disparidades localmente. Puede apreciarse que el número de puntos poco confiables (grises oscuros y negros) disminuye considerablemente. Por lo tanto, se verifica el supuesto de que al restringir la búsqueda en forma local a partir de una estimación inicial de las disparidades, las ambigüedades en la correlación de la siguiente iteración disminuyen.

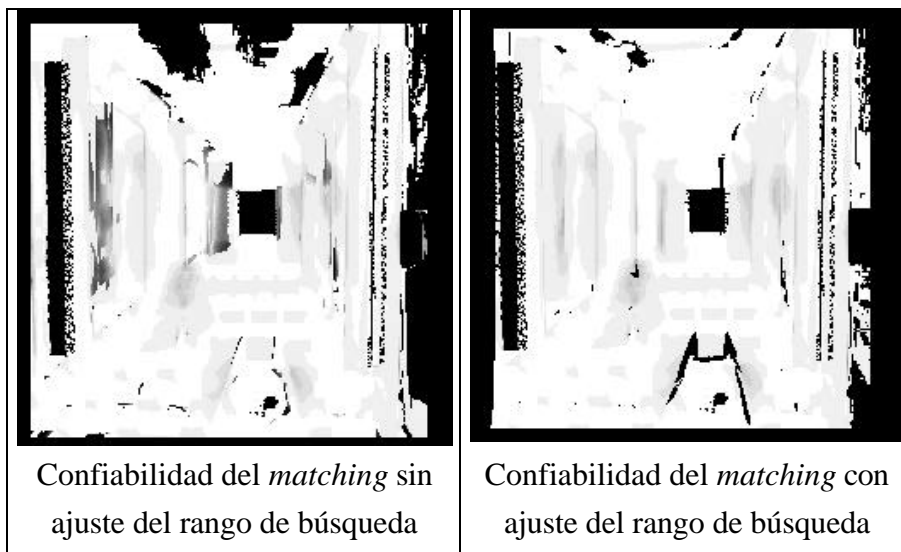


Figura 4.11: Mapas de confiabilidad del par Sint2 procesado con un bloque de 7x7.

#### 4.1.2 Imágenes de Prueba

El resultado anterior también se verifica en las imágenes de Pira700. En la figura 4.12 se muestra la confiabilidad obtenida al utilizar un rango inicial de búsqueda entre 41 y 49 pixels, y aquella obtenida al restringir el rango localmente a partir de los valores del mapa de disparidades anterior. En este caso el aumento de confiabilidad también es significativo.

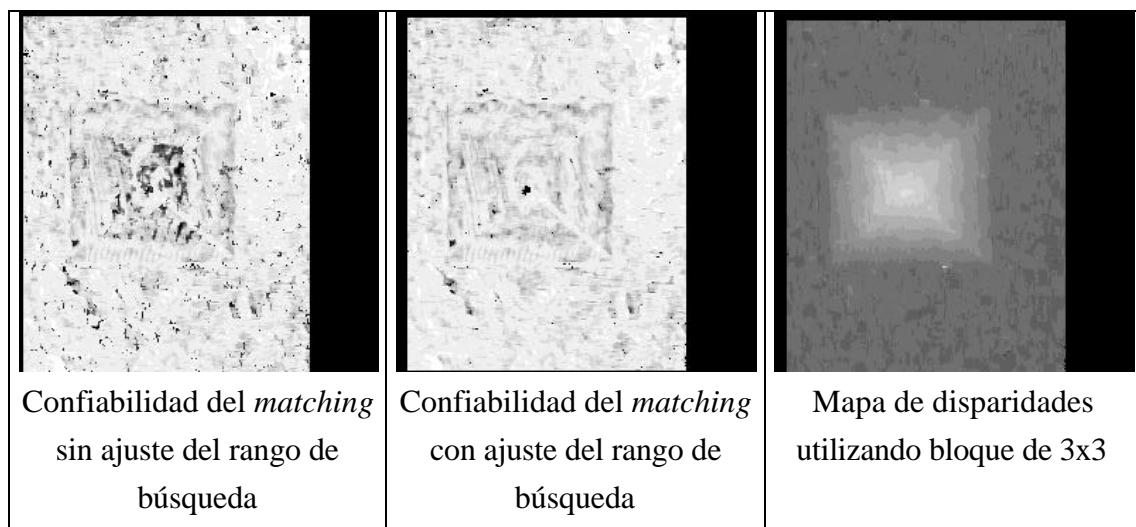


Figura 4.12: Mapas de confiabilidad de Pira700.

El análisis del error en la reconstrucción para las imágenes de prueba tomadas en el laboratorio se presenta con mayor detalle en la sección 4.2.

## 4.2 Reconstrucción y Mediciones Tridimensionales

Antes de analizar los resultados obtenidos mediante el procesamiento de secuencias de pares estéreo de una escena, en la tabla 4.2 se demuestran las ventajas de utilizar la correlación normalizada en vez del criterio SSD.

Aunque el algoritmo con resolución sub-píxel basado en la NCC es 2.2 veces más lento que el basado en SSD, el primero tiene tan solo un 2.29% de error relativo al rango en la pruebas realizadas con Pira700, mientras que el segundo tiene un error del 5.67%.

Tabla 4.2: Cuadro Comparativo de la Correspondencia SSD vs. NCC.

Método	SSD	NCC
Bloque Inicial	7x7	7x7/5x5
Disparidad Mín. [pixels]	44.44	44.53
Disparidad Máx. [pixels]	48.17	47.98
Zmax [mm]	654.88	653.22
Zmin [mm]	605.96	606.32
Zmax-Zmin [mm]	50.72	46.90
Volumen [cm <sup>3</sup> ]	50.72	46.41
Error Relativo al Rango Z = 48 mm	5.67%	2.29%
Error Volumen %	4.79%	4.11%
Rango de Búsqueda	41-49	41-49
Tiempo de Procesamiento [s]	1.1	2.4

En la tabla 4.3 se muestran los resultados obtenidos para la reconstrucción tridimensional y la medición del volumen de Pira700 aplicando las técnicas de procesamiento secuencial descritas. De estos resultados se destaca lo siguiente:

- Al utilizar el procesamiento secuencial se logran menores errores que en el caso del procesamiento de un solo par (tabla 4.2), especialmente en la determinación de las distancias a los puntos en la escena, los cuales están relacionados al error relativo al rango.
- Al emplear filtros de más etapas se logran mejores resultados.
- La reducción del error utilizando el procesamiento FIR con ventana de tamaño variable (decreciente) frente al nivel de error del procesamiento IIR con ventana de tamaño fijo son similares, excepto para la secuencia de 16 pares, que es notablemente mejor tanto en la estimación de distancias como en la medición del volumen.
- La utilización de bloques de correlación menores a 11x11 permiten obtener mejores resultados que al utilizar bloques de dimensiones mayores como el de 33x33 al procesar la secuencia Pira700. Estos últimos filtran las componentes de alta frecuencia de las imágenes y reducen el nivel de detalles del mapa de disparidad.

Tabla 4.3: Evaluación de los métodos de procesamiento secuencial aplicados a la secuencia Pira700.

Método	FIR			IIR		
	4	8	16	4	8	16
Bloque Inicial	33x33v	11x11v	7x7v	33x33c	11x11c	7x7c
Disparidad Mín. [pixels]	44.09	44.07	43.87	43.85	43.89	43.67
Disparidad Máx. [pixels]	47.26	47.42	47.36	46.79	47.29	47.33
Zmax [mm]	659.73	660.15	663.17	663.43	662.89	666.11
Zmin [mm]	615.62	613.46	614.28	621.68	615.22	614.60
Zmax-Zmin [mm]	44.12	46.69	48.89	41.75	47.66	51.51
Volumen [cm <sup>3</sup> ]	43.13	44.02	48.92	43.06	50.43	49.61
Error Relativo al Rango Z = 48 mm	8.08%	2.73%	1.85%	13.02%	0.71%	7.31%
Error Volumen %	10.89%	9.05%	1.07%	11.03%	4.19%	2.50%

(Indice v: denota tamaño de bloque variable (decreciente). Índice c: denota tamaño de bloque constante)

La representación tridimensional de la pirámide con su textura y la malla de superficie 3D se muestran respectivamente en la figuras 4.13 y 4.15 para un par procesado de Pira700.

La reconstrucción tridimensional de la pirámide obtenida procesando los 16 pares de la secuencia Pira700 con un bloque de correlación inicial de 7x7, y empleando el método FIR de coeficientes variables con rango de búsqueda de disparidades restringido en forma local, se muestra en las figuras 4.14 y 4.16 para la superficie con textura y la superficie enmallada, respectivamente.

Las figuras 4.17 y 4.18 muestran la escena del muñeco de la figura 2.6 reconstruida tridimensionalmente a partir del mapa de disparidades interpolado. Aunque la escala de profundidades se ha comprimido para efectos del gráfico, se aprecian los tres niveles principales de profundidad de la escena: el muñeco, la caja y el librero.

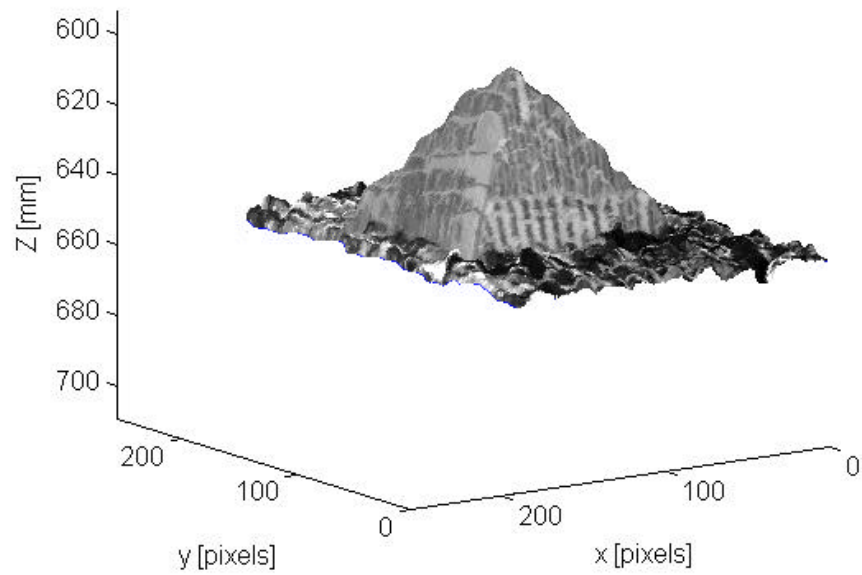


Figura 4.13: Representación con *texture mapping* de la pirámide de Pira700,  $k=0$ .

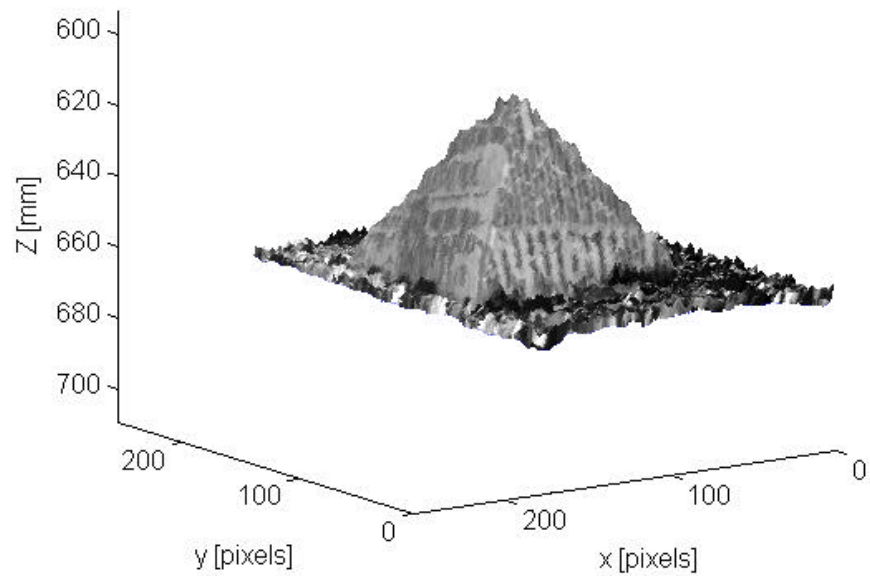


Figura 4.14: Representación con *texture mapping* de la pirámide de Pira700,  $k=0-15$ .

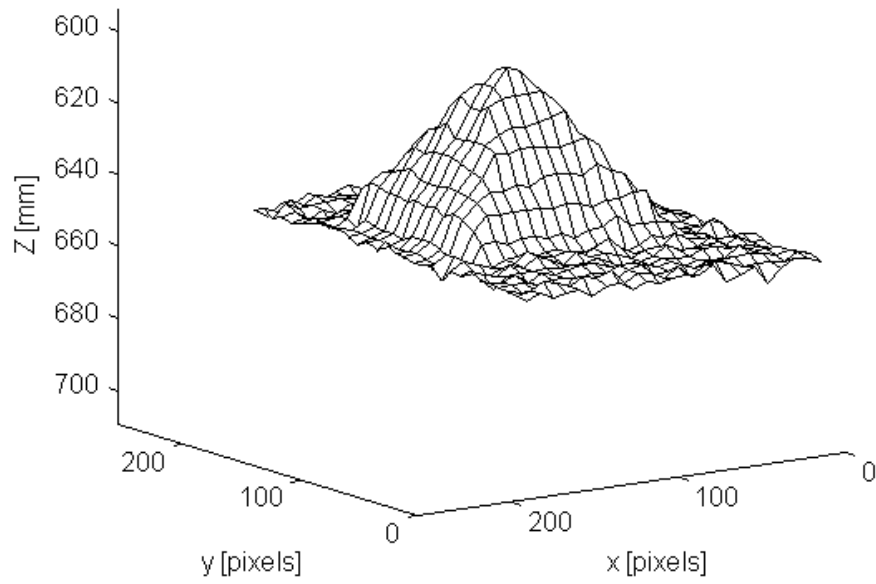


Figura 4.15: Representación *mallado de superficie* de la pirámide de Pira700,  $k=0$ .

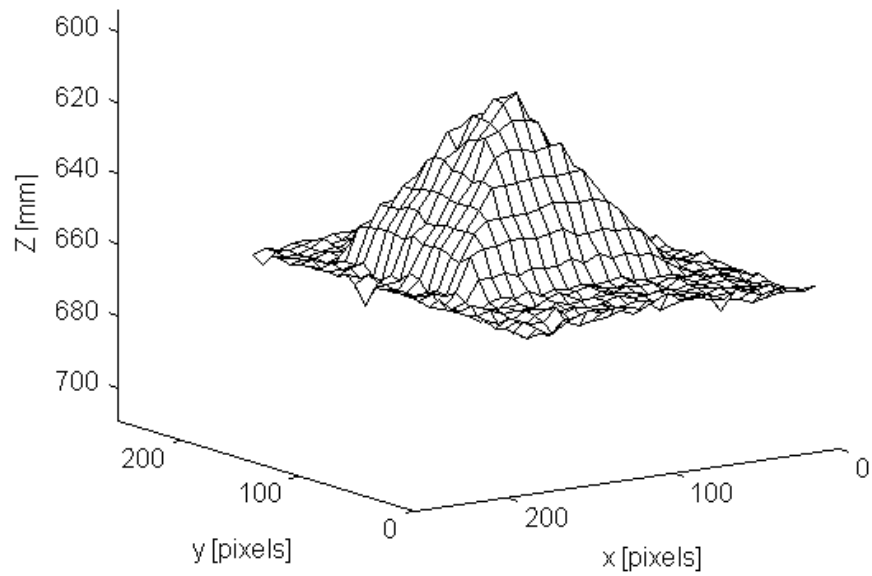


Figura 4.16: Representación *mallado de superficie* de la pirámide de Pira700,  $k=0-15$ .

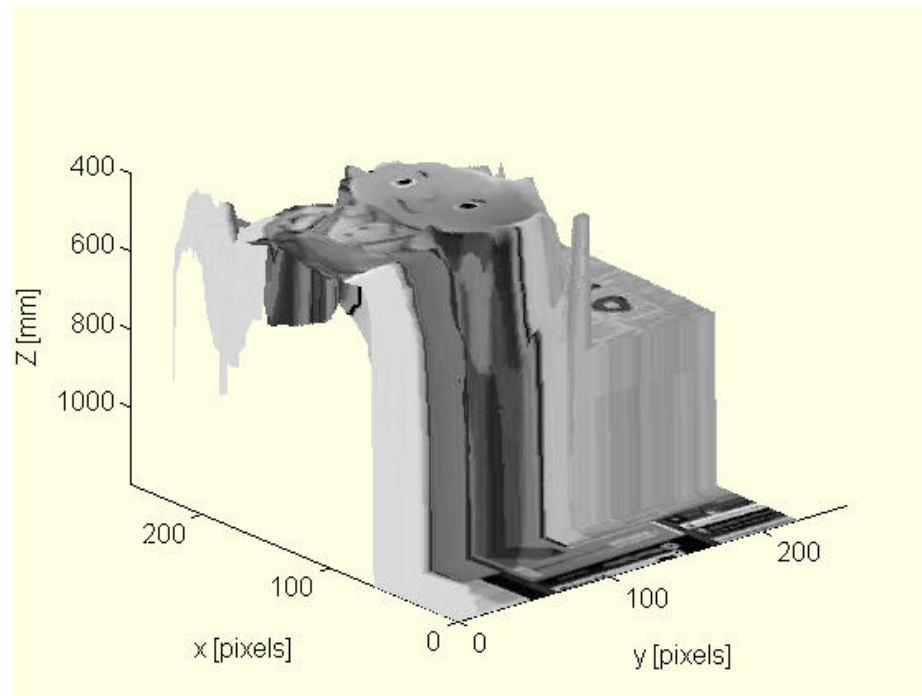


Figura 4.17: Reconstrucción 3D de la escena del muñeco.

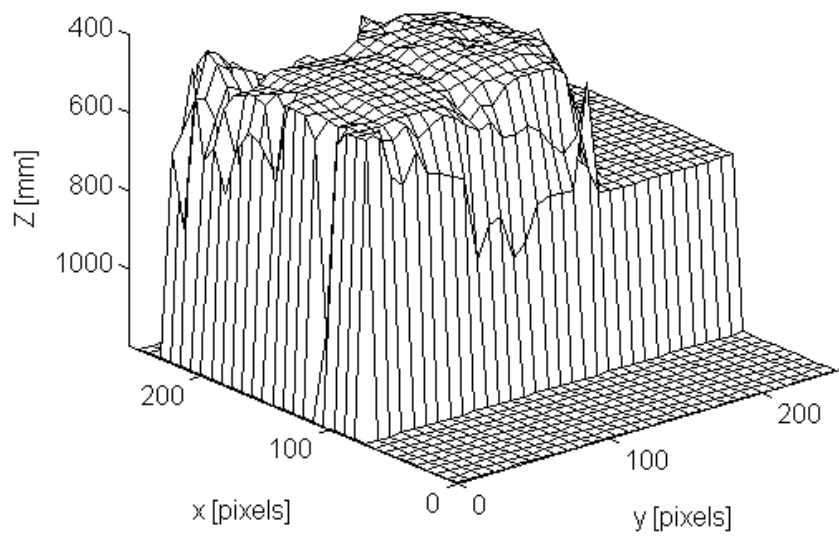


Figura 4.18: Malla 3D de la escena del muñeco.



### 4.3 Implementación de los Algoritmos

Los algoritmos descritos se implementaron en un computador IBM-PC con un procesador Pentium de 233 MHz y 32MB de RAM, utilizando el lenguaje C para Windows 95 [Schi95, Fast95]. El consumo de memoria y la utilización de la CPU durante el procesamiento de un par por el programa desarrollado se muestra en la figura 4.19. Se puede apreciar que el consumo máximo de memoria no supera los 5 MB, y no se requiere de *swaping* de memoria a disco, ya que las únicas lecturas y escrituras que se producen corresponden a la apertura y cierre de archivos de imágenes y resultados.

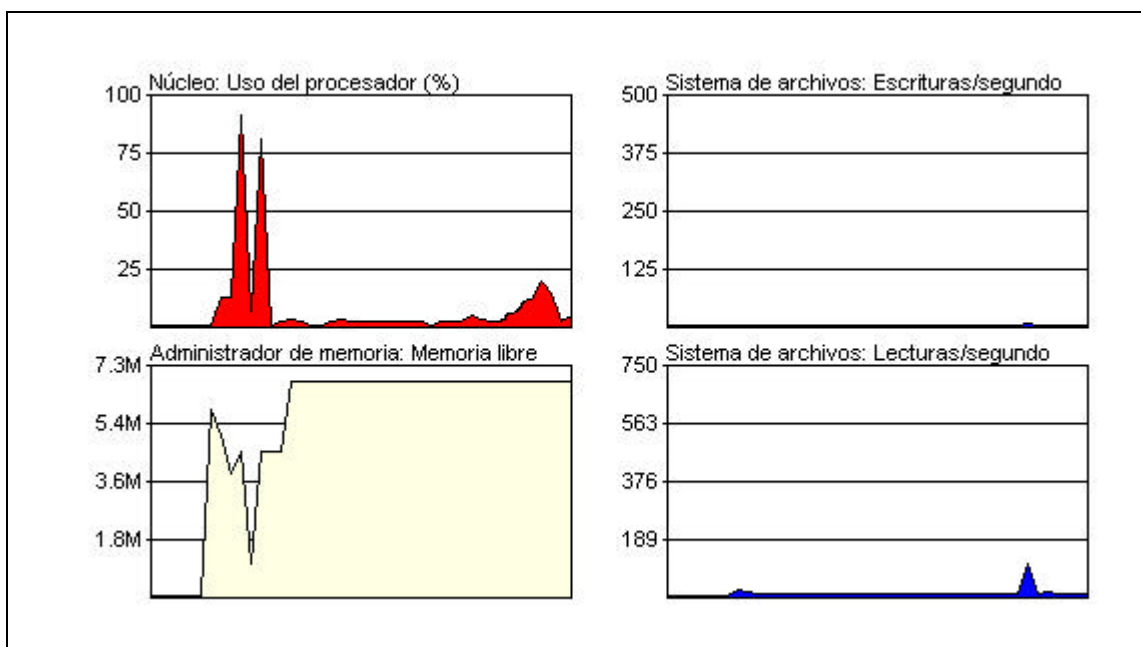


Figura 4.13: Consumo de recursos del sistema por el programa para estereoscopia.

La interfaz del programa se muestra en la figura 4.14. El programa posee algunas herramientas para el análisis gráfico de las curvas de correlación, además posee filtros para el preprocesamiento de las imágenes y algunas rutinas para la realización de mediciones (en unidades métricas) de la superficie reconstruida en función de los parámetros del sistema binocular.

La interfaz se desarrolló en forma estructurada, de modo que la incorporación de código adicional sea en lo posible sencilla, y pueda emplearse en futuras aplicaciones de análisis 3D.

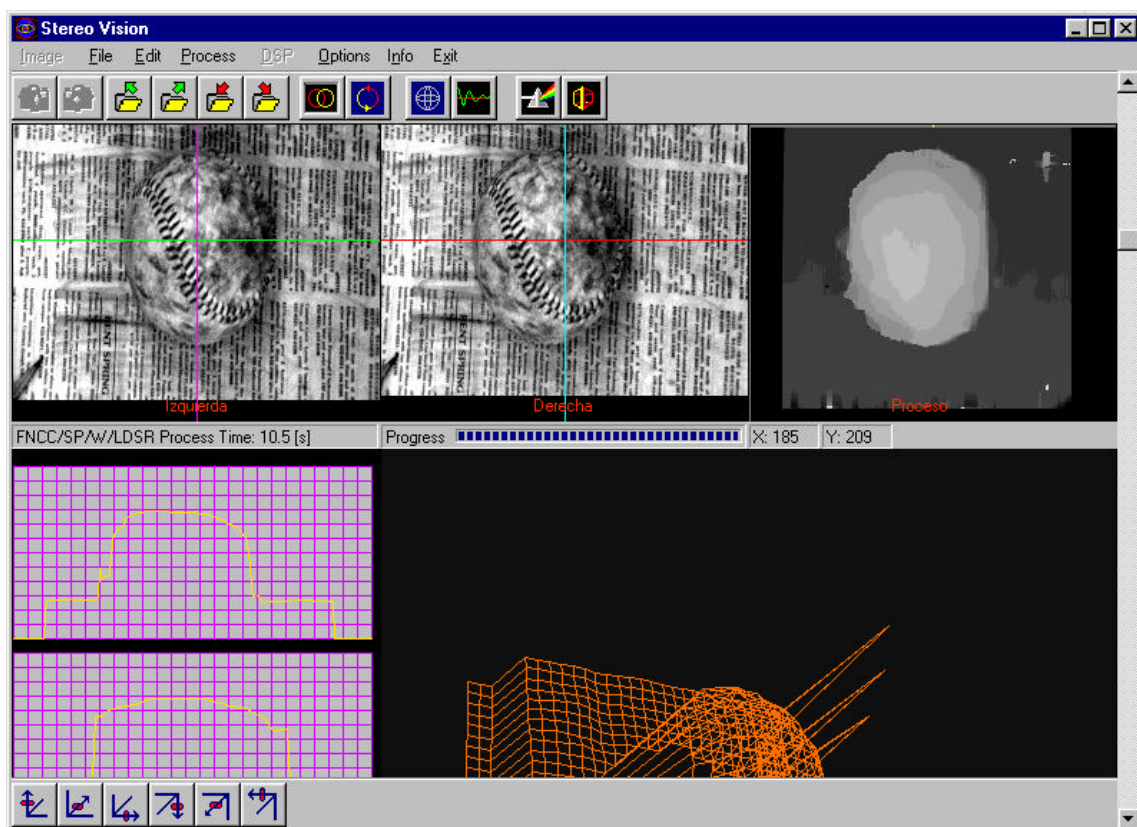


Figura 4.14: Interfaz del programa para análisis estereoscópico.

## V. CONCLUSIONES Y DISCUSION

Se realizó una investigación sobre los métodos existentes para la medición de la profundidad. Se determinó que los métodos basados en estereoscopia son los que ofrecen mayores ventajas prácticas, aunque en el futuro los costos de los *laser range finders*, superiores en exactitud y precisión, podrían reducirse y desplazar a los métodos estéreo. La evaluación de los métodos estéreo permitió establecer que los algoritmos basados en correlación de áreas son hasta la fecha los mejores.

Para lograr el objetivo primario de esta investigación: reducir el error en la estimación de la posición de los puntos de la escena en el espacio tridimensional, fue necesario analizar las fuentes de error en la estimación. Esto permitió establecer expresiones matemáticas sobre los límites superiores de la resolución y confiabilidad de los resultados del sistema. De este análisis se desprende que el error depende del rango de operación del sistema y que el parámetro que más contribuye a aumentar el error en la estimación de la profundidad es el error en la calibración del ángulo de convergencia.

Se implementó el algoritmo de *correlación normalizada acelerada sub-pixel*, introduciendo una etapa de análisis de las curvas de correlación y el procesamiento de secuencias de pares estéreo como métodos para reducir los errores producto de las ambigüedades que se generan en la etapa de búsqueda de correspondencias. Como el análisis de los resultados experimentales lo demuestra, se logró una disminución del error y un aumento en la confiabilidad de la estimación de distancias.

A diferencia de otros algoritmos, la correlación normalizada se hace sobre los niveles de intensidad de la imagen y no involucra la segmentación *a priori* de ésta o la extracción de características. En pruebas realizadas, éstos demostraron ser más rápidos y entregaron mapas de disparidad más densos. Se prefirió la correlación normalizada acelerada según [Faug93i], ya que se comprobó que es menos sensible al ruido de cuantización que la correlación acelerada basada en la suma de diferencias al cuadrado, aunque entre un 70% y un 120% más lenta.

La introducción del rango ajustado de búsqueda de correspondencias permitió reducir el error en la determinación de las disparidades en al menos un 6% y una reducción del tiempo de procesamiento de los pares siguientes de un 23%. Si además se introduce la interpolación de las zonas no correspondidas o determinadas inválidas, entonces el error promedio disminuye alrededor de 25%.

A continuación se presenta un resumen con los principales aportes, las limitaciones del sistema desarrollado y los desarrollos futuros.

### **5.1 Resumen de Contribuciones**

Los aportes novedosos de este trabajo son:

- El desarrollo de un algoritmo estereoscópico que permite la reconstrucción tridimensional con mayor precisión al incorporar el ajuste de rango de búsqueda a través del procesamiento secuencial.
- La incorporación del estimador de confiabilidad de los resultados a partir del análisis de las curvas de correlación.
- La determinación de las expresiones matemáticas que permitan el diseño de un sistema estereoscópico binocular que cumpla una cierta tolerancia.
- La interpolación bilineal sólo en las áreas no correspondidas.

### **5.2 Limitaciones de las Técnicas Utilizadas**

Las limitaciones de las técnicas de estimación de distancias para reconstrucción tridimensional utilizando estereoscopía están dadas fundamentalmente por las texturas de la escena, o la escasa visibilidad en ambientes con neblina o mucho polvo.

La sensibilidad de los métodos a la presencia de ciertas texturas o patrones puede reducirse empleando sistemas trinoculares. En aplicaciones en interiores, proyectando un patrón de manchas (con algo tan simple como un proyector

de diapositivas como en [Nish84]) se puede lograr una textura adecuada para la escena.

Recordemos que las ventajas de la estereoscopía frente a otros métodos descritos en la sección 1.2, son fundamentalmente la densidad de los resultados, la precisión, la fácil adaptabilidad del sistema binocular a distintos rangos y su bajo costo.

### 5.3 Desarrollos Futuros

Como se demostró mediante el análisis realizado en el capítulo 3, la adquisición del par estéreo con un sistema calibrado en forma precisa, es el factor que más influye en la etapas de posteriores del proceso de *matching* y reconstrucción tridimensional.

La adquisición de las imágenes de prueba en el laboratorio se realizó con extremo cuidado de modo que las imágenes no requiriesen de una rectificación. Sin embargo, en los sistemas prácticos las cámaras están sujetas a las vibraciones que pueden introducir severas distorsiones sobre los parámetros del sistema binocular, las cuales causarán mediciones incorrectas. Especialmente sistemas dinámicos como robots y vehículos poseen vibraciones que obligan a la calibración continua del sistema de visión estéreo. Por esta razón la calibración del sistema (i.e. la obtención de los parámetros extrínsecos e intrínsecos) debería ser automática y realizarse a partir del procesamiento de los mismos pares estéreo.

Además la *autocalibración* es muy importante en procesos industriales, ya que permitiría la calibración en línea del sensor sin la necesidad de detener el proceso.

Futuras aplicaciones prácticas deberían considerar procesos de *autocalibración* y determinación del movimiento, tal vez usando técnicas de estimación del *flujo óptico* [Horn86, Roy92]. Lecturas recomendadas en esta área son [Horn86, Tsai87, Roy92, Luon93, Zelle96].

La *autocalibración* y rectificación de la imágenes es un tema que, aunque ha recibido bastante atención, aún puede ser motivo para investigaciones futuras, ya

que será parte necesaria de cualquier sistema automatizado de visión artificial con fines prácticos.

Un trabajo que podría ser parte del curso de Diseño Eléctrico sería la implementación en hardware de los algoritmos de estereoscopia y otras rutinas que permitan una reconstrucción tridimensional más rápida aprovechando las capacidades de procesamiento del DSP TI C40 que posee la tarjeta de adquisición de imágenes Dipix XPG-1000.

Desarrollos futuros podrían considerar la implementación de un *sistema de visión inteligente*, que permita a un robot decidir que acción tomar con respecto al medio que lo rodea. En la industria, el sistema binocular podría utilizarse para el control de inventario en silos, medición de volumen, estimación de la masa de material almacenado, control de calidad y operación remota. En el área del diagnóstico clínico los algoritmos desarrollados podrían emplearse para el análisis oftalmológico como en [Duc197].

En el figura 5.1 se muestra de manera simplificada una posible aplicación futura en el área de robótica. En el diagrama se ilustra la relación del sistema de visión con el resto de los componentes del sistema de control y planificación de tareas.

Para alcanzar los beneficios de una producción verdaderamente automatizada se requiere programación de alto nivel del movimiento del robot, la cual se conoce como planificación de tareas. Sin embargo, en la actualidad los sistemas aún no han alcanzado una flexibilidad adecuada en términos de capacidad de adaptación a cambios en el espacio de trabajo, lo cual implica que la planificación de tareas no logra ser un cien por ciento efectiva. En las personas la vista es fundamental para poder interactuar con el medio. Por ejemplo, sin iluminación adecuada, una persona en un espacio que desconoce tendrá que dar pasos tentativos para saber si hay escalones o paredes próximas. La inseguridad por la carencia de un *feedback* apropiado limitará su velocidad de acción. Por esta razón los sistemas de visión constituyen una base clave para aumentar la inteligencia del sistema, ya que proporcionan una gran cantidad de información del entorno y de los cambios que

ocurren en éste. Su uso se justifica también en el hecho de que la tecnología existente permite procesar la información visual en forma cada vez más eficiente.

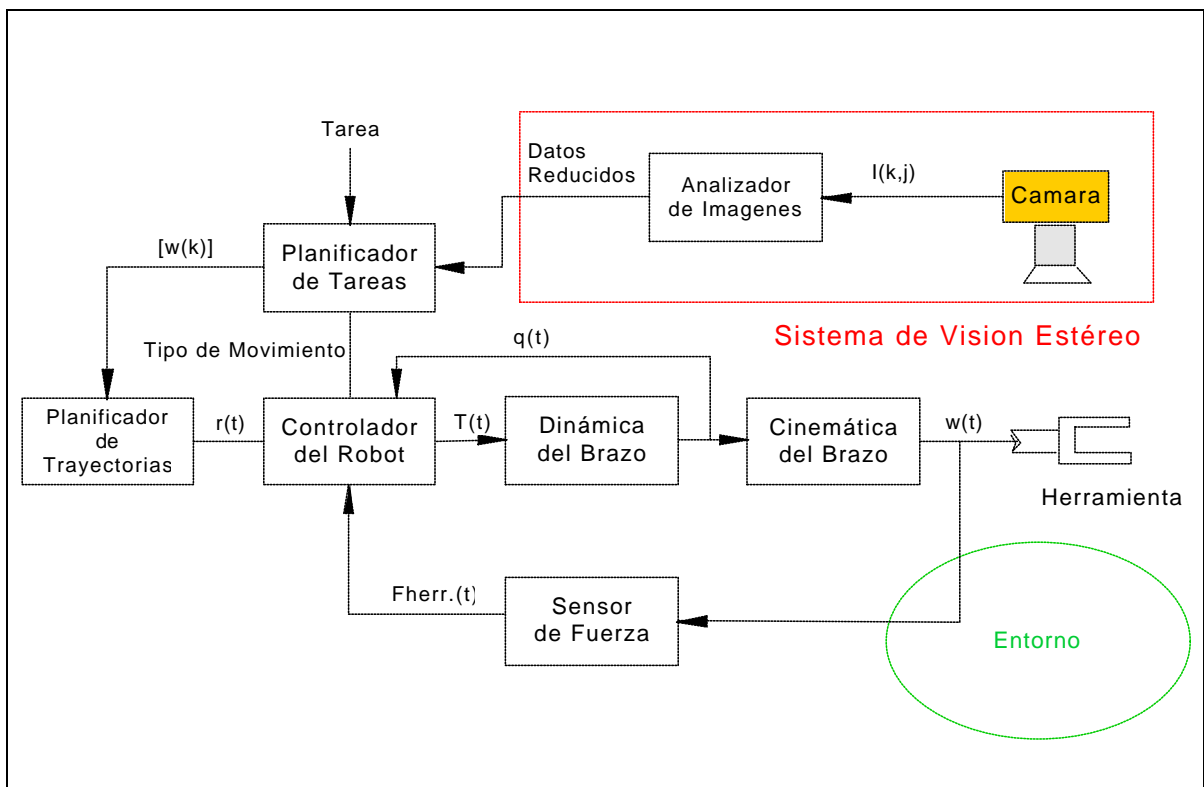


Figura 5.1: La estereoscopia como componente esencial de un sistema robótico.

Como se muestra en la figura 5.1, el *input* primario es la especificación de la tarea. Para planificar la tarea, el planificador usa un modelo interno del entorno, además de datos *on-line* provenientes de los sensores, específicamente del sistema de visión. Las imágenes  $I(k,j)$  son procesadas por el analizador de imágenes para reducir los datos a una forma utilizable por el planificador de tareas, por ejemplo, distancia a objetos. Un movimiento es planificado como una trayectoria específica discreta  $w(k)$ , la cual es enviada al planificador de trayectorias, que mediante técnicas de interpolación y de cinemática inversa convierte  $w(k)$  a un movimiento continuo equivalente en el espacio de las articulaciones del robot. La trayectoria en el espacio de las articulaciones sirve luego para que el controlador de torque genere el perfil de torque requerido,  $T(t)$ , para cada articulación del brazo robótico. Los datos  $F_{herr.}$ ,

*feedback* de los sensores de movimiento y fuerza, son usados por el controlador para lograr movimientos resguardados (que eviten colisiones) y que cumplan con la trayectoria exigida.

En este contexto, la estereoscopía constituye un método para aprovechar al máximo la información visual del entorno cambiante.



## BIBLIOGRAFÍA

Nota: ☺ Recomendada

☺☺ Muy Recomendada

☺☺☺ Imprescindible

### Bibliografía Básica

#### I. Estereoscopia

##### *Revisiones Generales y Teoría*

- [Barn82] S. Barnard y M. Fischler. **Computational Stereo**. Computing Surveys, 14(4): 553-570. Diciembre, 1982. ☺☺
- [Dhon89] U. Dhond, J.K. Aggarwal. **Structure from Stereo - A Review**. IEEE Transactions on Systems, Man, and Cybernetics, 19(6): 1489-1510. Noviembre/Diciembre 1989. ☺☺
- [Faug93] O. Faugeras. **Three-Dimensional Computer Vision: A Geometric Viewpoint**. The MIT Press. Cambridge, Massachusetts, U.S.A., 1993. ☺☺☺
- [Horn86] B. K. P. Horn. **Robot Vision**. The MIT Press, Cambridge, Massachusetts, U.S.A., 1986. ☺
- [Jarv83] R. A. Jarvis. **A Perspective on Range Finding Techniques for Computer Vision**. IEEE Transactions on Pattern Analysis and Machine Intelligence, 5(2): 122-139. Marzo 1983. ☺☺
- [Klet95] R. Kettle, A. Koschan, K. Schlüns and Volker Rodehorst. **Evaluation of Surface Reconstruction Methods**. Computer Vision Group, FR 3-11. Computer Science Department, Berlin Technical University. 1995.
- [Lane96] R.A. Lane, N.A. Thacker. **Stereo Vision Research: An Algorithm Survey**. Research Report, Electronic Systems Group, Department of Electronic and Electrical Engineering, University of Sheffield, U.K., Enero, 1996. ☺☺

- [Wood93] A. Woods, T. Docherty, R. Koch. **Image Distortion in Stereoscopic Video Systems**. Stereoscopic Displays and Applications IV, Proceedings of the SPIE Volume 1915, San Jose, California, U.S.A., Feb., 1993. ☺☺

### *Desarrollos basados en Correlación*

- [Faug93i] O. Faugeras, B. Hotz, H. Mathieu, T. Viéville, Z. Zhang, P. Fua, E. Théron, L. Moll, G. Berry, J. Vuillemin, P. Bertin, C. Proy. **Real Time Correlation-Based Stereo: Algorithm, Implementations and Applications**. Rapport de recherche N° 2013, INRIA, Sophia-Antipolis, Francia, Abril, 1993. ☺☺☺
- [Fua91] P. Fua. **Combining Stereo and Monocular Information to Compute Dense Depth Maps that Preserve Depth Discontinuities**. Proceedings of the 12<sup>th</sup> International Joint Conference on Artificial Intelligence, págs. 1292-1298, Agosto, 1991. ☺☺☺
- [Nish84] H.K. Nishihara. **PRISM: A Practical Real-Time Imaging Stereo Matcher**, A.I. Memo 780. Massachusetts Institute of Technology, Artificial Intelligence Laboratory. Mayo 1984. ☺☺☺

### *Desarrollos basados en Análisis del Flujo Óptico*

- [Roy92] S. Roy. **Analyse d'images Stéréoscopiques Basée sur la Détermination du Flux Optique**. Master Thesis, University of Montreal, Montreal, Canada, 1992.

### *Análisis de Múltiples Imágenes Estéreo*

- [Roy93] S. Roy, J. Meunier. **Stereoscopic Analysis of Multiple Images**. Département d'informatique et de recherche opérationnelle. Université de Montréal, Montreal, Canada, 1993. Submitted August 7<sup>th</sup> 1993 to the International Journal of Computer Vision. ☺

[Tuca96] V. Tucakov, D. G. Lowe. **Temporally Coherent Stereo: Improving Performance Through Knowledge of Motion.** Dept. of Computer Science, Univeristy of British Columbia, Vancouver, B.C., Canada, Junio 1996. ☺☺☺

[Falk95] Ver **III. Reconstrucción 3D**: L. Falkenhagen. ☺☺☺

[Koch96i, Koch95, Koch94, Koch94j] Ver **III. Reconstrucción 3D**: R. Koch. ☺☺☺

## II. Error en la Reconstrucción 3D usando Estereoscopia

[Blos87] S.D. Blostein, T.S. Huang. **Error Analysis in Stereo Determination of 3-D Point Positions.** IEEE Transactions on Pattern Analysis and Machine Intelligence, (9)6: 752-765, Noviembre, 1987.

[Chan94] C. Chang, S. Chatterjee, P.R. Kube. **A Quantization Error Analysis for Convergent Stereo.** Proceedings of the 1994 IEEE International Conference on Computer Vision, U.S.A., (2): 735-739, 1994. ☺☺

[Keht94] N.D. Kehtarnavaz, W. Sohn. **Analysis of Camera Movements in Stereo Vision-Based Vehicle Tracking.** Proceedings of the 1994 IEEE International Conference on Computer Vision, U.S.A., (2): 710-714, 1994. ☺

[McVe82] E. S. McVey, J.W. Lee. **Some Accuracy and Resolution Aspects of Computer Vision Distance Measurements.** IEEE Transactions on Pattern Analysis and Machine Intelligence, (4)6: 646-649, Noviembre, 1982. ☺☺☺

[Moha89] R. Mohan, G. Medioni, R. Nevatia. **Stereo Error Detection, Correction, and Evaluation.** IEEE Transactions on Pattern Analysis and Machine Intelligence, (2)2: 113-120, Febrero, 1989.

[Rodr90] J.J. Rodríguez, J.K. Aggarwal. **Stochastic Analysis of Stereo Quantization Error.** IEEE Transactions on Pattern Analysis and Machine Intelligence, (12)5: 467-470, Mayo, 1990. ☺☺

[Zhao95] W. Zhao, N. Nandhakumar. **Effects of Camera Aligment Errors on Stereoscopic Depth Estimates.** Journal of Pattern Recongnition. págs. 2115-2126, U.S.A., Diciembre, 1996. ☺

### III. Reconstrucción 3D

- [Koch96i] R. Koch. **3-D Modeling of Human Heads from Stereoscopic Image Sequences**. Conference of the German Pattern Recognition Society 96 (Deutsche Arbeitsgemeinschaft Mustererkennung), Heidelberg, Germany, Sept. 1996.
- [Koch96] R. Koch. **Surface Segmentation and Modeling of 3-D Polygonal Objects from Stereoscopic Image Pairs**. International Conference on Pattern Recognition 96, Vienna, Austria, Agosto, 1996.
- [Koch95] R. Koch. **3-D Surface Reconstruction from Stereoscopic Image Sequences**. International Conference of Computer Vision ICCV '95, Cambridge, MA., USA, Junio, 1995. ☺☺☺
- [Koch94i] R. Koch. **Model-based 3-D scene analysis from stereoscopic image sequences**. ISPRS Journal of Photogrammetry and Remote Sensing 49 (5), pp. 23-30, Elsevier Science Publishers B.V., 1994. ☺☺☺
- [Koch94] R. Koch. **3-D Scene Modeling from Stereoscopic Image Sequences**. Paker, Y. and Wilbur, S. (Ed.), Image Processing for Broadcast and Video Production, Hamburg 1994, pp. 128-135, Springer series on Workshops in Computing, ISBN 3-540-19947-0, Springer Great Britain, 1994.
- [Koch93i] R. Koch. **Automatic Reconstruction of Buildings from Stereoscopic Image Sequences**. Computer Graphics Forum Vol. 12 (3), Proceedings of Eurographics '93, Barcelona, Spain, Sept. 1993.
- [Koch93] R. Koch. **Dynamic 3D Scene Analysis through Synthesis Feedback Control**. IEEE Trans. Patt. Anal. Mach. Intell., Special issue on analysis and synthesis, 15(6), pp. 556-568, Junio 1993.
- [Falk97i] L. Falkenhagen. **Hierarchical Block-Based Disparity Estimation Considering Neighbourhood Constraints**. International workshop on SNHC and 3D Imaging, September 5-9, 1997, Rhodes, Greece. ☺☺☺

- [Falk97] L. Falkenhagen. **Block-Based Depth Estimation from Image Triples with Unrestricted Camera Setup**. IEEE Workshop on Multimedia Signal Processing, Junio 23 - 25, 1997, Princeton, New Jersey, USA.
- [Falk95] L. Falkenhagen. **3D Object-Based Depth Estimation from Stereoscopic Image Sequences**. International Workshop on stereoscopic and three-dimensional imaging, Septiembre 6-8 1995, Fera Congress Center, Santorini, Greece, 1995.
- [Falk94] L. Falkenhagen. **Depth Estimation from Stereoscopic Image Pairs Assuming Piecewise Continuous Surfaces**. European Workshop on Combined real and synthetic image processing for broadcast and video productions, Noviembre 23-24, 1994, Hamburg, Germany
- [Falk94i] L. Falkenhagen. **Depth Estimation from Stereoscopic Image Pairs Assuming Piecewise Continuous Surfaces**. Paker, Y. and Wilbur, S. (Ed.), Image Processing for Broadcast and Video Production, Hamburg 1994, pp. 115-127, Springer series on Workshops in Computing, ISBN 3-540-19947-0, Springer Great Britain, 1994.

#### IV. Información en Internet

##### *Scanning Laser Range Finders*

- [web1] <http://www.sick.de> (Sick AEG)
- [web2] <http://www.frc.ir.cmu.edu> (Field Robotics Center/Institute of Robotics/CMU)
- [web3] <http://www.riegl.co.at> (Riegl)
- [web4] <http://www.helpmaterobotics.com> (Helpmate Robotics)

***Frame Grabbers***

[web5] <http://www.dipix.com> (Dipix)

[web6] <http://www.matrox.com> (Matrox)

***Sistema Trinocular***

[web7] <http://www.ptgrey.com> (Point Gray Research)

***Radar de Apertura Sintética (SAR)***

[web8] [http://innpin.phy.hw.ac.uk/~peckham/envphy\\_4/sar/sar.html](http://innpin.phy.hw.ac.uk/~peckham/envphy_4/sar/sar.html)  
(Heriot-Watt  
University)

[web9] [http://www.gis.wau.nl/sar/sig/sar\\_reso.htm](http://www.gis.wau.nl/sar/sig/sar_reso.htm) (Remote Sensing Society)

[web10] <http://southport.jpl.nasa.gov/desc/imagingradarv3.html> (JPL)

[web11] <http://www.gfz-potsdam.de/pb1/pg5/research/INSAR/introduction.html>  
(GFZ/Potsdam)

***Calibración de las Cámaras***

[web12] <http://www.ius.cs.cmu.edu/afs/cs.cmu.edu/user/rgw/www/TsaiCode.html>  
(CMU)

***Topografía y DEM (Digital Elevation Model)***

- [web13] <http://pds.jpl.nasa.gov:8005/mgddf/chap4/chap4.htm> (Sonda Magellan)
- [web14] <http://terraweb.wr.usgs.gov/TRS/projects/Flagstaff/photos/> (Satélites SPOT)

***3D Graphics Algorithms: Mesh Generation, Shading, Texture Mapping and Rendering***

- [web15] <http://www.cs.cmu.edu/~ph/> (Paul S. Heckbert/CMU)

***Visión: Imágenes y Links Relacionados***

- [web16] <http://www.cs.cmu.edu/afs/cs/project/cil/ftp/html/vision.html> (Computer Vision CMU)
- [web17] <http://www-dbv.informatik.uni-bonn.de/> (Comp. Vision University of Bonn)

## Bibliografía Complementaria

### I. Algoritmos y Aplicaciones de Estereoscopía

#### *General*

- [Coch90] S. Cochrane. **Surface Description from Binocular Stereo**. Tesis de Doctorado, Southern California University. California, U.S.A., Noviembre, 1990. ☺☺
- [Coch92] S. Cochrane, G. Medioni. **3-D Surface Description from Binocular Stereo**. IEEE Transactions on Pattern Analysis and Machine Intelligence. (14)10:981-944, Octubre, 1992.
- [Duc197] J. P. Duclos. **Aplicación de Técnicas Estereoscópicas en la Reconstrucción Tridimensional de la Papila**. Tesis de Magister. Pontificia Universidad Católica de Chile, Santiago, Chile, 1997. ☺☺

#### *NCC*

- [Lan95] Z. Lan, R. Mohr. **Robust Matching by Partial Correlation**. Rapport de recherche N° 2643, INRIA, Sophia-Antipolis, Francia, Agosto, 1995.

#### *SSD*

- [Kana93] M. Okutomi, T. Kanade. **A Multiple Baseline Stereo**. IEEE Transactions on Pattern Analysis and Machine Intelligence, (15)4: 353-363, Abril, 1993. ☺☺
- [Kana94] T. Kanade, M. Okutomi. **A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment**. IEEE Transactions on Pattern Analysis and Machine Intelligence, (16)9: 920-932, Septiembre 1994. ☺☺



- [Kana94i] T. Kanade. **Development of a Video-Rate Stereo Machine**. Proceedings of 94 ARPA Image Understanding Workshop, Monttey, California, págs. 549-558, U.S.A., Noviembre, 1994. ☺

### *Piramidal*

- [Hoff89] W. Hoff, N. Ahuja. **Surfaces from Stereo: Integrating Feature Matching, Disparity Estimation, and Contour Detection**. IEEE Transactions on Pattern Analysis and Machine Intelligence, (11)2: 121-136, Febrero 1989.

- [Falk97i] Ver *L. Falkenhagen*

### *Adaptive Window Size Stereo*

- [Kana94] Ver *SSD*

- [Lotti94] J. Lotti, G. Giraudon. **Correlation Algorithm with Adaptive Window for Aerial Image in Stereo Vision**. Rapport de recherche N°2121, INRIA, Sophia-Antipolis, Francia, 1994. ☺☺☺

### *Programación Dinámica*

- [Birc98] S. Birchfield, C. Tomasi. **Depth Discontinuities by Pixel-to-Pixel Stereo**. Department of Electrical Engineering Department of Computer Science, Stanford University. Stanford, California. Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India. ☺
- [Cox92] I. J. Cox, S. Hingorani, B. Maggs, S. Rao. **Stereo without Regularisation**. British Machine Vision Conference, Leeds, UK, págs. 733-739, David Hogg&Roger Boyle (ed.), Springer Verlag, 1992.

- [Cox94] I. J. Cox. **A Maximum Likelihood N-camera Stereo Algorithm.** IEEE Conference On Computer Vision and Pattern Recognition. págs. 733-739, 1994.
- [Cox94i] I. J. Cox, S. Hingorani, B. Maggs, S. Rao. **A Maximum Likelihood Stereo Algorithm.** Computer Vision and Image Understanding, (63)3: 542-567, Mayo, 1996.

### *Técnicas de Relajación de Restricciones*

- [Marr76] D. Marr, T. Poggio. **Cooperative computation of stereo disparity.** Science, Vol. 194, págs. 283-287, 1976.

### *Fase*

- [Fleet91] D. Fleet, A. Jepson, M. R. M. Jenkin. **Phase-Based Disparity Measurement.** CVGIP: Image Understanding, 53(2): 198-210, 1991.
- [Maim96] M. Maimone. **Characterizing Stereo Matching Problems using Local Spatial Frequency.** Tesis de Doctorado. School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213 3891, U.S.A., Mayo, 1996.
- [Smit96] P. W. Smith, N. Nandhakumar. **An Improved Power Cepstrum Based Stereo Correspondence Method for Textured Scenes.** IEEE Transactions on Pattern Analysis and Machine Intelligence, (18)3:338-348, Marzo, 1996.
- [Yesh89] Y. Yeshurun, E. Schwartz. **Cepstral Filtering on a Columnar Image Architecture: A Fast Algorithm for Binocular Stereo Segmentation.** IEEE Transactions on Pattern Analysis and Machine Intelligence, (11)7:759-767, Julio, 1989.

### *Area/Segment Stereo*

- [Mara89] S. Marapane, M. Trivedi. **Region-Based Stereo Analysis for Robotic Applications**. IEEE Transactions on Systems, Man , and Cybernetics, (19)6: 1447-1464, Noviembre/Diciembre, 1989.

### *Segment Stereo (Zero-Crossings)*

- [Kim87] Y.C. Kim, J.K. Aggarwal. **Positioning Three-Dimensional Objects Using Stereo Images**. IEEE Journal of Robotics and Automation, (RA-3)4: 361-373, Agosto 1987.

### *Active Stereo*

- [Ahu93] N. Ahuja, A.L. Abbott. **Active Stereo: Integrating Disparity, Vergence, Focus, Aperture, and Calibration for Surface Estimation**. IEEE Transactions on Pattern Analysis and Machine Intelligence, (15)10: 1007-1029, Octubre, 1993.

### *Sistemas Trinoculares y Estereoscopia con Múltiples Baselines*

- [Ayac91] N. Ayache, F. Lustman. **Trinocular Stereo Vision for Robotics**. IEEE Transactions on Pattern Analysis and Machine Intelligence, (13)1: 73-84, Enero, 1991.
- [Blak93] A. Blake, D. McCowen, H.R. Lo, P.J. Lindsey. **Trinocular Active Range-Sensing**. IEEE Transactions on Pattern Analysis and Machine Intelligence, (15)5: 477-483, Mayo, 1993.
- [Chio95] Ver *Configuración de las Cámaras*
- [Kana93] Ver *SSD*
- [Stew89] Charles V. Stewart. **Trinocular stereo. Theoretical advantages and a new algorithm**. Proceedings of SPIE - The International Society for Optical Engineering. Publ. by Int Soc for Optical Engineering, Bellingham, WA, USA. v 1198. p 377-391. Sensor Fusion II: Human and Machine Strategies. Philadelphia, PA, USA, 1989.

### *Algoritmos Acelerados para Estereoscopia*

[Kosc95] A. Koschan and V. Rodehorst. **Towards Real-Time Stereo Employing Parallel Algorithms for Edge-Based and Dense Stereo Matching.** Proc. of the IEEE Workshop on Computer Architectures for Machine Perception CAMP'95, Como, Italy, 18-20 September 1995, pp. 234-241.

[Faug93i], [Fua91] Ver en **Bibliografía Básica.**

### *Implementaciones en Hardware*

[Erte96] G. Erten. **Analog VLSI Implementation for Stereo Correspondence Between 2-D Images.** IEEE Transactions on Pattern Analysis and Machine Intelligence, (7)2: 266-277, Marzo, 1996.

[Hors93] I. Horswill, M. Yamamoto. **A US\$1000 Active Stereo Vision System.** Artificial Intelligence Lab. MIT, Cambridge, MA, U.S.A., 1993.

[Nish84] Ver **Bibliografía Básica.**

[Kana94i] y [Faug93i] Usando DSP's. Ver **Bibliografía Básica.**

### *Configuración de las Cámaras*

[Alve89] N. Alvertos, D. Brzakovic, R.C. González. **Camera Geometries for Image Matching in 3D Machine Vision.** IEEE Transactions on Pattern Analysis and Machine Intelligence, (2)9: 897-914, Septiembre, 1989.

[Chio95] R.N. Chiou, C.H. Chen, K.C. Hung, and J.Y. Lee. **The Optimal Camera Geometrie and Performance Analysis of a Trinocular Vision System** IEEE Transactions on Pattern Analysis and Machine Intelligence, (25)8: 1207-1220, Agosto, 1995.

## II. Calibración de las Cámaras

- [Tsai87] R. Tsai. **A Versatile Camera Calibration Technique for High Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses.** IEEE Journal of Robotics and Automation, 3(4): 323-344. Agosto, 1987. ☺☺☺
- [Luon93] Q. Luong, O. Faugeras. **Self-Calibration of a Stereo Rig from Unknown Camera Motions and Point Correspondences.** Rapport de recherche N°2014, INRIA, Sophia-Antipolis, Francia, Julio, 1993.
- [Zelle96] C. Zeller, O. Faugeras. **Camera Self-Calibration from Video Sequences: the Kruppa Equations Revisited.** Rapport de recherche N°2793, INRIA, Sophia-Antipolis, Francia, Febrero, 1996.
- [Zhan93] Z. Zhang, Q. Luong, O. Faugeras. **Motion of an Uncalibrated Stereo Rig: Self-Calibration and Metric Reconstruction.** Rapport de recherche N° 2079, INRIA, Sophia-Antipolis, Francia, Octubre, 1993.
- [View94] T. Viéville. **Auto-Calibration of Visual Sensor Parameters on a Robotic Head.** INRIA, Sophia, Francia. Image and Vision Computing, Diciembre, 1994.
- [Stein96] G.P. Stein. **Lens Distortion Calibration Using Point Correspondences.** A.I. Memo 1595. Massachusetts Institute of Technology, Artificial Intelligence Laboratory. Noviembre, 1996.

### *Geometría Epipolar - Rectificación*

- [Papa96] D.V. Papadimitriou, T.J. Dennis. **Epipolar Line Estimation and Rectification for Stereo Image Matching Pairs.** IEEE Transactions on Pattern Analysis and Machine Intelligence, (5)4: 672-676, Abril, 1996.
- [Zhan94] Z. Zhang, R. Deriche, O. Faugeras, Q. Luong. **A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry.** Rapport de recherche N° 2273, INRIA, Sophia-Antipolis, Francia, Mayo, 1994.

### III. Optic Flow - Shape from Motion - Motion Estimation

- [Bab94] A. Bab-Hadiashar, David Suter, Ray Jarvis. **Image-Interpolation Based Optic Flow Technique**. Intelligent Robotics Research Centre, Department of Electrical & Computer Systems Engineering. Monash University, Australia. 1994.
- [Yi96] J.Yi, J. Oh. **Estimation of depth and 3D motion parameter of moving object with multiple stereo images**. Image And Vision Computing, Elsevier Science B.V. (ed.), 14(7), págs. 501-516, Julio 1996.
- [Yi97] J.Yi, J. Oh. **Recursive resolving algorithm for multiple stereo and motion matches**. Image And Vision Computing, Elsevier Science B.V. (ed.), 15(73), págs. 181-196, Marzo, 1997.

### IV. Aplicaciones basadas en Sonar, Range Finders, Dead Reckoning, GPS

- [DFR96] G. Dudek, P. Freedman, I. Rekleitis. **Just-In-Time Sensing: Efficiently Combining Sonar and Laser Range Data for Exploring Unknown Worlds**. IEEE International Conference in Robotics and Automation, (1), págs. 667-671, Abril, 1996.
- [Kelly94] A. Kelly. **Modern Inertial and Satellite Navigation Systems**. CMU-RI-TR-94-15. The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A., Mayo, 1994.
- [Kelly95] A. Kelly. **Concept Design of a Scanning Laser Rangefinder for Autonomous Vehicle**. CMU-RI-TR-94-21. The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A., Enero, 1995.
- [Thor92] C. Thorpe, T. Kanade, M. Hebert, D. Pomerleau. **Third Annual Report for Perception for Outdoor Navigation**. CMU-RI-TR-92-16. The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A., Septiembre, 1992.

- [Krot94] E. Krotkov, M. Herbert, R. Simmons. **Stereo Perception and Dead Reckoning for a Prototype Lunar Rover**. CMU-RI-TR-92-16. The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A., 1994.

## V. Gráficos en Computador

- [Hear97] D. Hearn, M.P. Baker. **Computer Graphics - C Version**. 2<sup>nd</sup> Edition, Prentice Hall Inc., U.S.A., 1997.
- [Irwi95] J. Irwin, W. T. Hewitt, T. L. J. Howard. **Computer Graphics and Visualization. Lighting and Shading** / Student Notes. ITTI Gravigs Project. Computer Graphics Unit, Manchester Computing Centre, University of Manchester, U.K., 1995. ☺☺☺
- [Itti95] **Computer Graphics and Visualization. Standards for Computer Graphics** / Student Notes. ITTI Gravigs Project. Computer Graphics Unit, Manchester Computing Centre, University of Manchester, U.K., 1995.
- [Lill95] C. Lilley, F. Lin, W. T. Hewitt, T. L. J. Howard. **Computer Graphics and Visualization. Colour in Computer Graphics** / Student Notes. ITTI Gravigs Project. Computer Graphics Unit, Manchester Computing Centre, University of Manchester, U.K., 1995. ☺☺
- [Lin95] F. Lin, K. Wyrwas, J. Irwin, C. Lilley, W. T. Hewitt, T. L. J. Howard. **Computer Graphics and Visualization. Geometry for Computer Graphics** / Student Notes. ITTI Gravigs Project. Computer Graphics Unit, Manchester Computing Centre, University of Manchester, U.K., 1995. ☺☺

### *Texture Mapping*

- [Heck86] P. Heckbert. **Survey of Texture Mapping**. IEEE Computer Graphics and Applications, págs. 207-212, Noviembre, 1986.
- [Heck89] P. Heckbert. **Fundamentals of Texture Mapping and Image Warping**. Tesis de Magister. Universidad de California, Berkeley. U.S.A., Junio, 1989.

### *Mesh Generation*

- [Boss96] F. Bossen, P. Heckbert. **A Pliant Method for Anisotropic Mesh Generation**. 5<sup>th</sup> International Meshing Roundtable. Octubre, 1996.

## **VI. Procesamiento de Imágenes**

### *General*

- [Chin82] R.T. Chin, C.A. Harlow. **Automated Visual Inspection: A Survey**. IEEE Transactions on Pattern Analysis and Machine Intelligence, (4)6: 557-572, Noviembre, 1982.
- [Pav182] T. Pavlidis. **Algorithms for Graphics and Image Processing**. Computer Science Press Inc., U.S.A., 1982.
- [Slam80] C.C. Slama, C. Theurer, S.W. Henriksen. **Manual of Photogrammetry**. American Society of Photogrammetry, 1980.
- [Gonz93] R. C. González, R.E. Woods. **Digital Image Processing**. Addison Wesley, 1<sup>a</sup> ed., U.S.A., Septiembre 1993. ☺☺☺
- [Prat91] W. K. Pratt. **Digital Image Processing**. John Wiley and Sons, 2<sup>a</sup> ed., New York, U.S.A., 1991. ☺☺☺
- [Wolb92] G. Wolberg. **Digital Image Warping**. IEEE Computer Society Press Inc., U.S.A., 1992. ☺

### *Detección de Bordes*

- [Cann86] J. Canny. **A Computational Approach to Edge Detection**. IEEE Transactions on Pattern Analysis and Machine Intelligence, (8)6: 679-698. Noviembre, 1986. ☺☺
- [Deri90] R. Deriche. **Fast Algorithms for Low-Level Vision**. IEEE Transactions on Pattern Analysis and Machine Intelligence, (12)1: 78-87, Enero, 1990. ☺☺



- [Deve95] F. Devernay. **A Non-Maxima Suppression Method for Edge Detection with Sub-Pixel Accuracy.** Rapport de recherche N° 2724, INRIA, Sophia-Antipolis, Francia, 1994. ☺☺

### *Interpolación*

- [Szel90] R. Szeliski. **Fast Surface Interpolation Using Hierarchical Basis Functions.** IEEE Transactions on Pattern Analysis and Machine Intelligence, (12)6: 513-528, Junio, 1990.

### **VII. Instrumentación y Control**

- [PIIC93] D.M. Considine. **Process/Industrial Instruments and Control Handbook.** 4ª Ed. McGraw-Hill, Inc., U.S.A., 1993.

### **VIII. Procesamiento Digital de Señales:**

- [Oppe75] A. Oppenheim, R. Schaffer. **Digital Signal Processing.** John Wiley and Sons, 2ª ed., New York, U.S.A., 1975. ☺☺

### **IX. Robótica**

- [Fu87] K.S. Fu, R.C. González, C.G.S. Lee. **Robotics: Control, Sensing, Vision and Intelligence.** McGraw-Hill Inc., U.S.A., 1987. ☺☺☺

### **X. Manuales**

- [Dipix96] **Power Grabber: Dipix XPG-1000 Frame Grabber.** Ver. 2.50. Dipix Technologies Inc. Board Products Division 1051 Baxter Road, Ottawa, Ontario, Canada., Abril 1996.

- [Scor91] **Scorbot-ER VII User's Manual.** Eshed Robotec Ltd. Tel Aviv, Israel, September 1991.

- [Schi95] H. Schildt. **Programación en C y C++ en Windows 95**. McGraw-Hill, Inc., España, 1995.
- [Fast95] **Fastgraph for Windows**. Ver. 1.0. Programmer's Graphics Library. Ted Grubber Software, Inc., Las Vegas, U.S.A., Octubre, 1995.

**ANEXOS**

## **ANEXO A: METODOS DE BUSQUEDA DE CORRESPONDENCIAS**

Se describen distintos métodos de búsqueda de correspondencias para el cálculo de disparidades (basado en la clasificación de [Faug93] y de [Duc197]).

### **A.1 Métodos Basados en Area**

En [Lan95] se afirma que existen tres métodos de *matching* basados en el área: Flujo Óptico, Transformada de Fourier, y Correlación. Sin embargo los dos primeros bien pueden ser considerados por sí solos como una categoría distinta por el tipo de procesamiento que involucran.

Los métodos basados en área (también llamados *métodos densos*) buscan encontrar para todos los puntos de la imagen A un punto correspondiente en la imagen B. La mayoría de estos métodos utilizan la restricción epipolar y en general aplican un criterio de continuidad en las disparidades. Dentro de los métodos basados en área pueden distinguirse tres enfoques distintos para resolver el problema de *matching*.

#### **A.1.1 Correspondencia Basada en Correlación**

Este método calcula una medida de similitud entre un punto de la imagen A y sus homólogos potenciales en la imagen B. Para ello se define para cada punto en la imagen A una vecindad, generalmente un área rectangular (ventana o bloque), que es comparada con un área idéntica en la imagen B, desplazada para cada nivel de disparidad, cubriendo así todo el segmento epipolar. Utilizando alguna medida de similitud, se comparan la ventana original de la imagen A con las ventanas desplazadas de la imagen B, eligiéndose la de mayor similitud. Generalmente se utiliza una medida de la correlación basada en la intensidad luminosa. En algunos casos las imágenes se filtran previamente usando pasabajos, frecuentemente del tipo Gaussiano, para eliminar errores introducidos en la cuantización o debido a los posibles reflejos y brillos en la escena.

La determinación de la ventana de correlación es una tarea difícil. En efecto, si la ventana es muy grande, los detalles se pierden por lo que la disparidad no será precisa, pero el método robusto; si la vecindad es muy pequeña, la posibilidad de

correspondencia inexacta aumenta fuertemente. Se han desarrollado algoritmos que usan ventanas adaptivas [Kana94, Lotti94], o múltiples resoluciones de imágenes, que equivale a utilizar ventanas de distintos tamaños [Falk97i, Coch92, Fua91].

La desventaja de los métodos de correlación es su excesivo requerimiento computacional. Esto ha hecho que se investiguen soluciones de computación paralela [Kosc95] y algoritmos optimizados [Fua91, Faug93i].

### **A.1.2 Correspondencia a Múltiples Niveles de Resolución**

Para optimizar la búsqueda de puntos correspondientes se puede utilizar una estimación de la disparidad. Aplicando distintos niveles de filtraje a las imágenes se puede obtener disparidades en las imágenes de resolución menor y utilizarlas como estimación de las imágenes de resolución superior, hasta llegar a la resolución original.

Este método reduce considerablemente el tiempo de cálculo (hasta un factor de 32 para una imagen de  $256 \times 256$  con 4 niveles de filtraje [Roy92]). Sin embargo, algunos autores mantienen sus reservas ante este procesamiento *jerárquico* de la imagen, ya que un error en un nivel inferior de resolución puede propagarse y aumentar los errores en niveles de mayor resolución [Faug93].

### **A.1.3 Programación Dinámica**

Aparte de buscar los pares de puntos correspondientes uno por uno, es posible asociar directamente dos conjuntos de puntos situados sobre vectores epipolares compatibles de cada imagen. El problema de correspondencia se convierte entonces en el problema de búsqueda de un camino de costo mínimo sobre el plano formado por ambos vectores epipolares, como se ve en la Figura A.1.3.

Si se aplica la hipótesis de que los objetos son opacos, es posible encontrar un camino óptimo en este mapa de costos. El problema se puede resolver mediante el uso de programación dinámica, método conocido por su rapidez de ejecución. La elección de la función de costo es crítica para la precisión de las disparidades obtenidas.

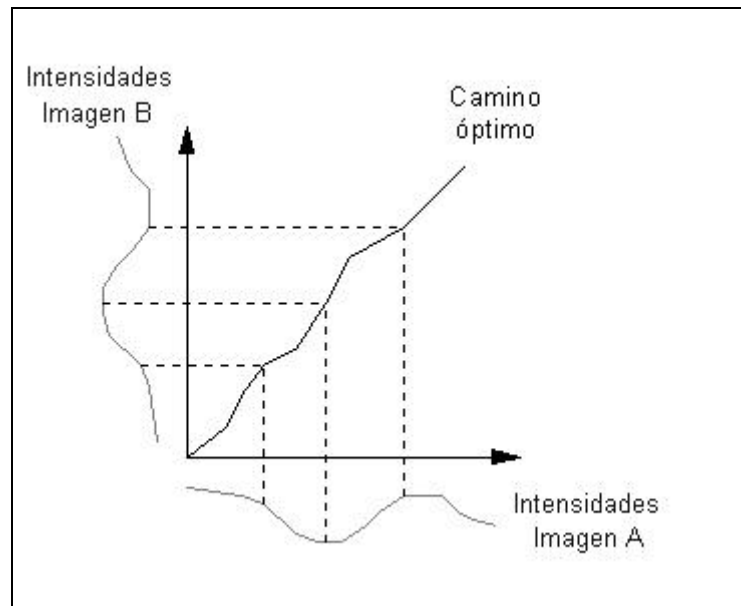


Figura A.1: Mapa de costos para dos segmentos epipolares.

La función de costo utilizada comúnmente es la diferencia de la intensidad luminosa y una componente que castiga las situaciones probables de oclusión. La explicaciones claras de la aplicación de este método se encuentran en [Birc98, Cox94i].

## A.2 Métodos Basados en Características

Los métodos no densos, es decir, que generan información de profundidad para algunos puntos de la escena, proceden inicialmente a la extracción de componentes específicos de ambas imágenes para luego ponerlos en correspondencia y obtener la profundidad. Debido a que la profundidad sólo se recupera para los componentes, el campo de profundidades no es completamente determinado y es necesario interpolar la profundidad para el resto de la imagen.

Estos métodos se ven afectados por las oclusiones y por las texturas que pueden hacer variar considerablemente el número de componentes a analizar. Si el número es muy grande, la búsqueda de correspondencia se complica; si es muy pequeña, la información sobre la profundidad queda demasiado dispersa, perdiendo utilidad y dependiendo demasiado de la calidad de los métodos de interpolación [Horn86].

### **A.3 Algoritmos Basados en Relajación**

El fundamento básico de esta clase de técnicas es permitir a los puntos que son puestos en correspondencia reorganizarse según patrones establecidos de “correspondencia correcta”, propagando en forma iterativa ciertas restricciones. Este tipo de técnicas fueron populares en la década del 70 y principios del 80, destacando el trabajo de [Marr76].

Sin embargo, estos algoritmos son extremadamente ineficientes y se han utilizado casi exclusivamente para respaldar teorías sobre visión biológica.

### **A.4 Algoritmos Basados en el Dominio de la Frecuencia**

Recientemente se han utilizado métodos basados en el dominio de la frecuencia, como los algoritmos que utilizan las diferencias de fase y la transformada *Cepstrum*. En el primer caso, se puede utilizar el resultado de un filtro complejo del tipo *Gabor* para obtener la diferencia de fase entre ventanas de las imágenes A y B, como se utiliza en [Fleet91, Maim96]. La diferencia de fase es proporcional a la disparidad, por lo que se pueden obtener resultados densos, similares a los métodos basados en área.

La transformada *Cepstrum* [Yesh89, Smit96], o más bien el espectro de potencia de una transformada *Cepstrum*, permite obtener disparidades ya sea en forma directa al reconocer picos en el resultado o al calcular varias transformadas en ventanas de ambas imágenes y escoger la de mejores coeficientes.

Aunque son métodos densos, los resultados obtenidos con este enfoque no logran superar a los algoritmos basados en el área. Por otro lado son más lentos aún si se usa la transformada rápida de Fourier (FFT).

## ANEXO B: APLICACIONES DE LA ESTEREOSCOPIA

La estereoscopia con finalidades recreativas se remonta casi a los inicios de la fotografía (s. XIX), en la figura B.1 se ilustra un par estereoscópico de 1872 (Publicado por Kilburn Bros, Littleton, NH).

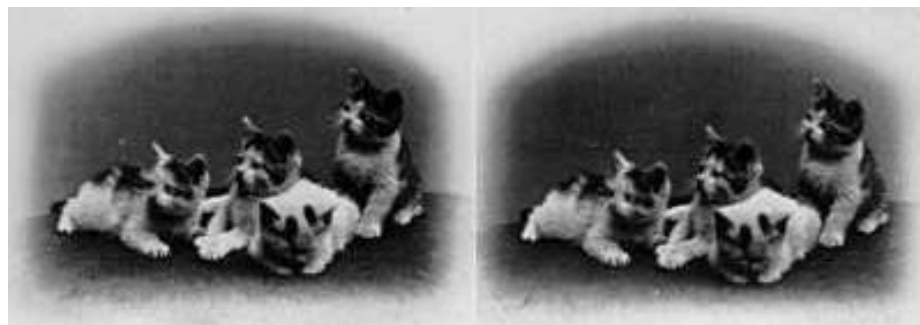


Figura B.1: Par estéreo: Who goes there! © 1878. Por J.P. Soule.

El físico francés Joseph Nicéphore Niépce (1765-1833) hizo la primera fotografía permanente en 1826, y se asoció con Louis Jacques Mandé Daguerre (1789-1851) en 1829, éste último perfeccionó el proceso fotográfico conocido como *Daguerreotipo* en 1837. Un año más tarde, el físico e inventor británico Sir Charles Wheatstone (1802-1875) inventaría el estereoscopio en 1838, aunque se hizo más famoso por aplicar un circuito para la medición de resistencias inventado por el británico Samuel Hunter Christie.

El astrónomo británico Sir John Herschel (1792-1871) dió el nombre a la fotografía e inventó un fijador para el proceso fotográfico. La astrofotografía nació virtualmente junto con la fotografía, fue William Draper quien retrató el primer rostro humano y el que tomó la primera fotografía del cielo. El físico británico James Clerk Maxwell (1831-1879) logró exitosamente la primera fotografía en color en 1861.



A continuación se ilustran algunas de las aplicaciones prácticas de la estereoscopía en la actualidad. Algunas de las áreas más comunes de aplicación son:

- Evasión de Obstáculos y Detección de Distancias (e.g. sistemas de navegación autónoma desarrollados por CMU, rovers como el Nomad o el Mars Pathfinder (ver figuras B.2.1, B.2.2, B.2.3)).
- Fotogrametría (e.g. pareja de satélites franceses SPOT (ver figura B.3), sonda Magellan enviada a Venus [web13, web14]).
- Reconstrucción 3D para aplicaciones de automatización industrial, CAD, y teleoperación.

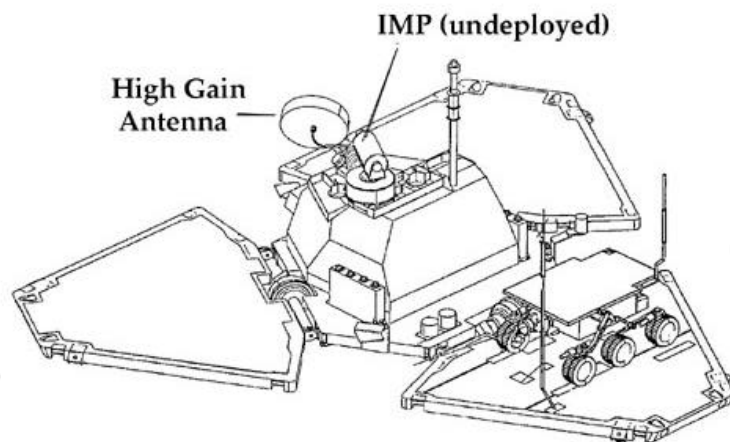


Figura B.2.1: Módulo espacial con el Imager Stereo (IMP) y el Pathfinder.

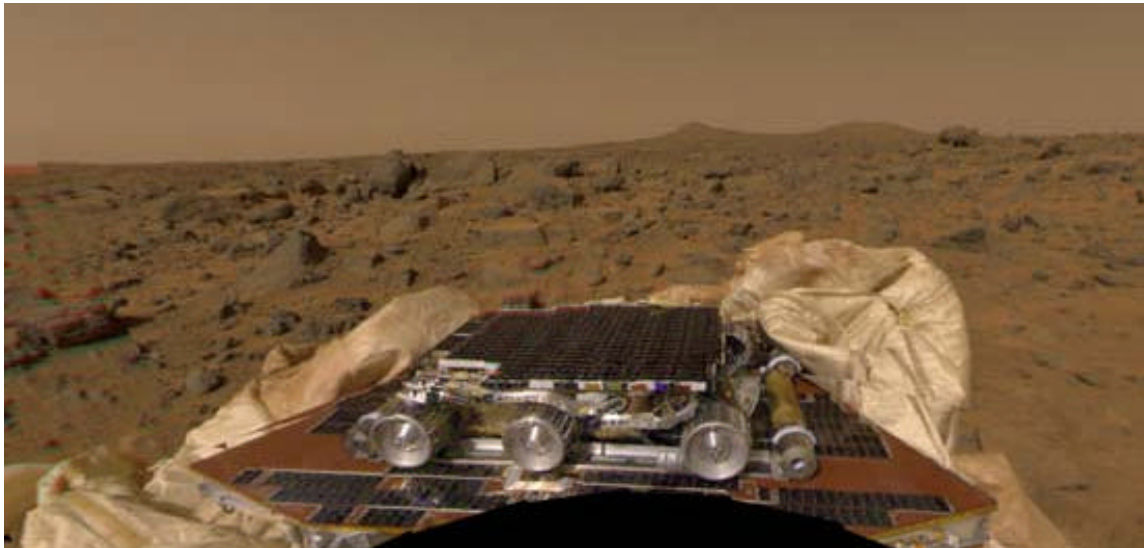


Figura B.2.2: Mars Pathfinder

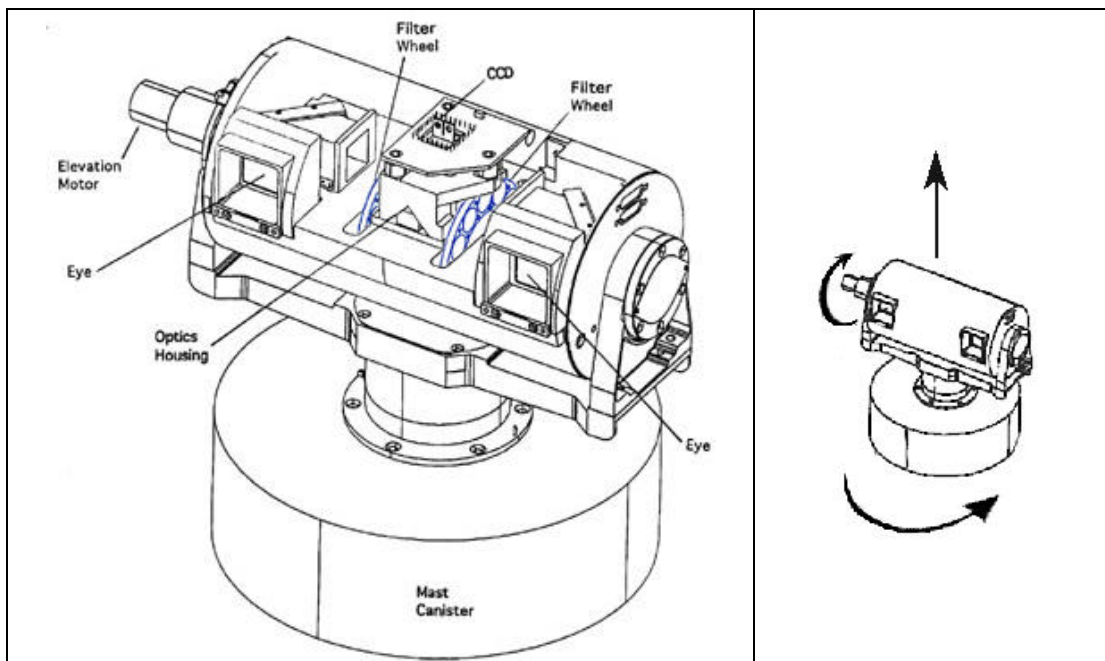


Figura B.2.3: Imager Stereo del Mars Pathfinder (IMP).

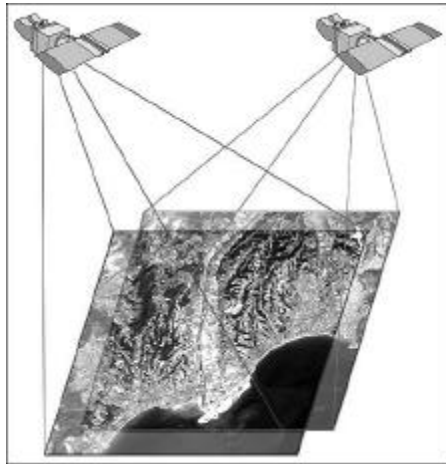


Figura B.3: Satélites SPOT: Système Probatoire d'Observation de la Terre.

## ANEXO C: ALGORITMOS IMPLEMENTADOS

### C.1 Algoritmo de Correlación Normalizada Acelerada

La correlación normalizada entre un punto de la imagen derecha  $R(x, y)$  y un punto en la imagen izquierda  $L(x+d, y)$  a una disparidad  $d$ , considerando una ventana o bloque de correlación de  $(2N+1) \cdot (2P+1)$  puntos se define como:

$$C(x, y, d) = \frac{\sum_{i=-N}^N \sum_{j=-P}^P (R(x+i, y+j) - \bar{R}(x, y)) (L(x+i+d, y+j) - \bar{L}(x+d, y))}{\sqrt{\sum_{i=-N}^N \sum_{j=-P}^P (R(x+i, y+j) - \bar{R}(x, y))^2} \cdot \sqrt{\sum_{i=-N}^N \sum_{j=-P}^P (L(x+i+d, y+j) - \bar{L}(x+d, y))^2}} \quad (\text{C.1})$$

donde

$$\bar{R}(x, y) = \frac{\sum_{i=-N}^N \sum_{j=-P}^P R(x+i, y+j)}{(2N+1) \cdot (2P+1)} \quad (\text{C.2})$$

$$\bar{L}(x, y) = \frac{\sum_{i=-N}^N \sum_{j=-P}^P L(x+i, y+j)}{(2N+1) \cdot (2P+1)}$$

$$x = N, \dots, (x_{max} - N - d_{max})$$

$$y = P, \dots, (y_{max} - P)$$

$$d = d_{min}, \dots, d_{max}$$

El cómputo de la expresión (C.1) puede simplificarse, descomponiendo el cálculo en diferentes partes:

$$C(x, y, d) = \frac{N(x, y, d)}{\mathbf{s}_R(x, y) \cdot \mathbf{s}_L(x+d, y)} \quad (\text{C.3})$$

$$N(x, y, d) = \sum_{i=-N}^N \sum_{j=-P}^P (R(x+i, y+j) - \bar{R}(x, y)) (L(x+i+d, y+j) - \bar{L}(x+d, y))$$

$$\mathbf{s}_R(x, y) = \sqrt{\sum_{i=-N}^N \sum_{j=-P}^P (R(x+i, y+j) - \bar{R}(x, y))^2}$$

$$\mathbf{s}_L(x+d, y) = \sqrt{\sum_{i=-N}^N \sum_{j=-P}^P (L(x+d+i, y+j) - \bar{L}(x+d, y))^2}$$

Los componentes del numerador y denominador a su vez pueden ser reformulados como:

$$N(x, y, d) = \sum_{i=-N}^N \sum_{j=-P}^P (R(x+i, y+j) \cdot L(x+i+d, y+j)) - (2N+1) \cdot (2P+1) \cdot \bar{R}(x, y) \cdot \bar{L}(x+d, y)$$

$$s_I(x, y) = \sqrt{\sum_{i=-N}^N \sum_{j=-P}^P (I(x+i, y+j)^2) - (2N+1) \cdot (2P+1) \cdot \bar{I}(x, y)^2}$$

donde  $I = R$  o  $L$  (imagen derecha o izquierda, respectivamente).

Al analizar los términos podrá notarse que tanto el numerador como el denominador requieren de  $(2N+1) \cdot (2P+1)$  multiplicaciones que tuvieron que utilizarse para el cálculo de la correlación en  $(x-1, y)$  y que en  $(x, y)$  se vuelven a utilizar (operaciones redundantes). También puede notarse que los cálculos de los promedios de intensidad se utilizan en todo el rango de disparidad.

Para evitar realizar cálculos redundantes, se aplica recursión sobre los índices  $i$  y  $j$  en el cálculo de los promedios, los componentes del denominador y el componente del numerador.

Inicialmente se calculan los promedios de la imagen derecha e izquierda y se almacenan para su uso posterior de la siguiente manera:

1. Suma de  $2P+1$  puntos en la fila  $y = P$ :

$$U_I(x, P) = \sum_{j=-P}^P I(x, P+j)$$

2. Promedio de  $2P+1$  puntos en la fila  $y+1$

$$V_I(x, y+1) = V_I(x, y) + U_I(x, y+2P+1) - U_I(x, y)$$

$$W_I(N, y) = \sum_{i=-P}^{i=P} V_I(N, y)$$

$$X_I(x+1, y) = X_I(x, y) + W_I(x+2N+1, y) - W_I(x, y)$$

$$\text{Promedio } \bar{I}(x, y) = \frac{X_I(x, y)}{(2N+1)(2P+1)}$$

Con la información del promedio para cada punto de ambas imágenes, se aplica el método recursivo de forma similar para el cálculo de los componentes del denominador (desviaciones estándares). Estos son también invariantes para el rango de disparidad, por lo que son almacenados:

$$\begin{aligned}
 E_I(x, y) &= I(x, y)^2 \\
 F_I(x, P) &= \sum_{j=-P}^P E_I(x, P + j) \\
 G_I(x, y + 1) &= G_I(x, y) + F_I(x, y + 2P + 1) - F_I(x, y) \\
 H_I(N, y) &= \sum_{i=-P}^{i=P} G_I(N, y) \\
 O_I(x + 1, y) &= O_I(x, y) + H_I(x + 2N + 1, y) - H_I(x, y) \\
 D_I(x, y) &= \sqrt{O_I(x, y) - (2N + 1)(2P + 1)\bar{I}(x, y)^2}
 \end{aligned}$$

Finalmente se calcula la componente del denominador:

$$\begin{aligned}
 Q(x, y, d) &= I_R(x, y)I_L(x + d, y) \\
 R(x, P, d) &= \sum_{j=-P}^P Q(x, P + j, d) \\
 R(x, y + 1, d) &= R(x, y, d) + Q(x, y + 2P + 1, d) - Q(x, y, d) \\
 S(N, y, d) &= \sum_{i=-P}^{i=P} R(N, y, d) \\
 S(x + 1, y, d) &= S(x, y, d) + R(x + 2N + 1, y, d) - R(x, y, d) \\
 N(x, y, d) &= S(x, y, d) - (2N + 1)(2P + 1)\bar{I}_R(x, y)\bar{I}_L(x + d, y)
 \end{aligned}$$

Con este resultado y las componentes del denominador almacenadas se calcula  $C$ . El método recursivo se representa gráficamente en la figura C.1.

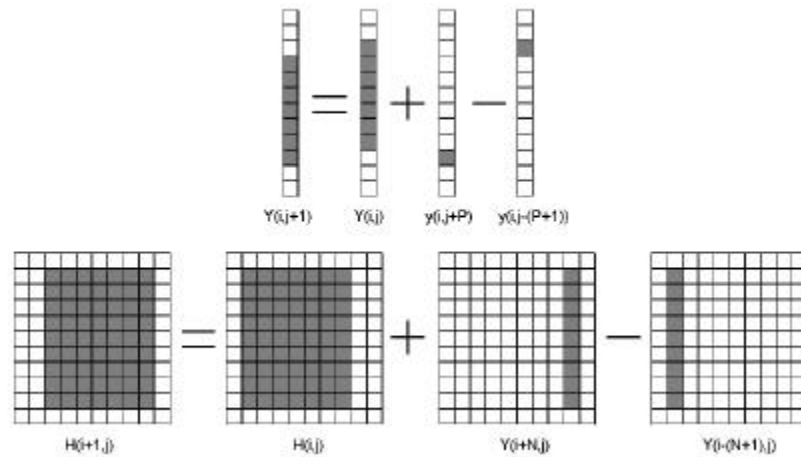


Figura C.1: Simplificación del cálculo de  $C$ .

El uso de las relaciones descritas permiten evitar cualquier redundancia en el cómputo de  $C$  y hacen el tiempo de procesamiento independiente del tamaño de la ventana de correlación.

Si se utiliza una ventana de correlación de ancho  $2N + 1$ , alto  $2P + 1$  y un rango de disparidad  $D = d_{max} - d_{min}$ , se logra reducir la complejidad del algoritmo de  $O(X \cdot Y \cdot D)$  a  $O((X+Y) \cdot D)$ .

## C.2 Cálculo de la Confiabilidad

Como se explico en el punto B de la sección 3.3.2, la confiabilidad se calcula empleando:

$$W(x, y) = 1 - \frac{C_2(x, y)}{C_1(x, y)}$$

Donde  $C_1$  y  $C_2$  se corresponden al máximo de la curva de correlación y al segundo peak más grande después del máximo, como se ilustra en la figura C.2. Los valores  $C_1$  y  $C_2$  se pueden obtener de manera eficiente, analizando la curva de correlación en forma simultánea al proceso de búsqueda de correspondencias, con los algoritmos implementados, y cuyos *pseudo-códigos* se presentan a continuación.

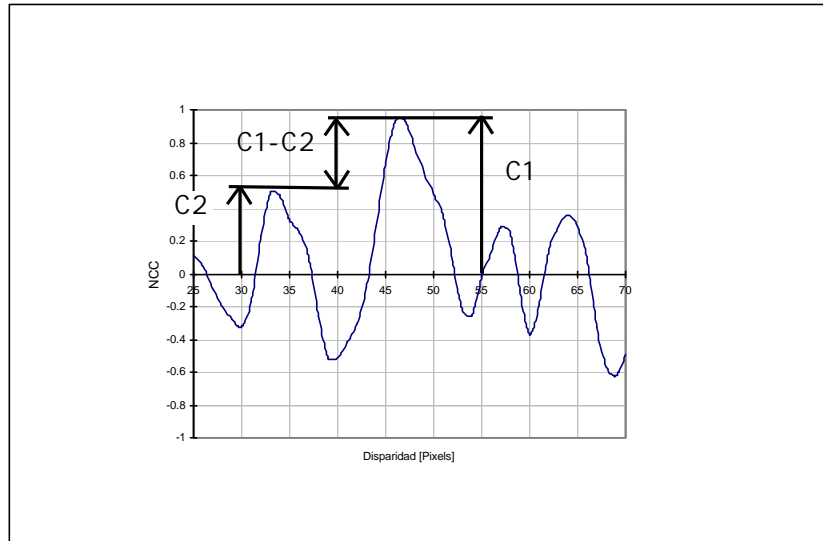


Figura C.2: Coeficiente de Confianza.

**Se define el tipo:**

**matriz\_de\_listas** Estructura matricial de  $N \times N$  (dimensión de la imagen de referencia), cuyos elementos son listas con una cantidad de elementos igual a **Largo\_Lista**.

**Se definen las variables:**

*Nota: Todas las variables son de tipo float, exceptuando donde se indique otro tipo.*

**Lista\_de\_Peaks** Tipo matriz\_de\_listas.

**Lista\_de\_Disparidades** Tipo matriz\_de\_listas.

**pos** (int) Posición en el mapa de disparidades o la imagen de referencia.

**valor\_peak** Valor del *peak*. Puede corresponder a un máximo local o global.

**corr** Valor de la Correlación calculada según las expresiones de la primera sección del Anexo C.

**img\_corr\_prev** Matriz de almacenamiento de las correlaciones de la etapa anterior  $k-1$ .



<b>img_corr_max</b>	Matriz de almacenamiento de las correlaciones máximas.
<b>img_corr_prev_max</b>	Matriz de almacenamiento de las correlaciones previas a encontrar el máximo.
<b>img_disp_RL</b> (int)	Matriz de almacenamiento de las disparidades correspondientes a cada máximo de <b>img_corr_max</b> .
<b>img_corr_next</b>	Matriz de almacenamiento de las correlaciones encontradas en para la disparidad en <b>img_disp_RL+1</b> (justo después de que se encontró el máximo de la correlación).
<b>disp_min_loc</b> (int)	Disparidad Mínima Local (inicialmente es igual a la disparidad mínima para todo el mapa de disparidades).
<b>disp_max_loc</b> (int)	Disparidad Máxima Local (inicialmente es igual a la disparidad máxima para todo el mapa de disparidades).

### Se definen las funciones:

```

Init_Vars (...)
/*Inicialización de Variables*/
for (x=0, y=0; x<256&&y<ymax; x++, y++)
{Lista_de_Peaks(x,y) = [];
Lista_de_Disparidades(x,y) = []; //Correspondientes a cada Peak
Valor_de_Peak = obtiene_correlación (N, P, /*Tamaño de Venatana*/
                                     x, y, /*Posición*/
                                     disp_min);

img_disp_RL(x,y)=disp_min;
}
Largo_Lista = 3; //Número de elementos de la lista de Peaks)

```

```

Actualiza_Lista_de_Peaks (Lista_de_Peaks,
                          Lista_de_Disparidades, // Correspondientes a cada peak
                          Valor_de_Peak,
                          disp, // Disparidad_de_Peak
                          pos, // Posición_del_Pixel al cual se le calcula NCC.
                          Largo_Lista, // Número de elementos de la lista de Peaks
                          )
{
/*Busca la Posición de la Lista donde insertar el nuevo Valor_de_Peak. En la lista se
almacenan los Peaks de mayor a menor. */
while (Lista_de_Peaks[j][Pos] > Valor_de_Peak && j < (Largo_Lista - 1))
    j++;

/*Desplaza elementos menores a Valor_de_Peak a posiciones de la j-1 a la última de la
lista. Se bota el elemento más bajo.*/
for (k=(Nelements-1);k>j;k--)
    {img_in[k][pos]=img_in[k-1][pos];
    img_in_disp[k][pos]=img_in_disp[k-1][pos];}

/*Se verifica que Valor_de_Peak sea mayor que el elemento que había en la posición j.
De ser así éste se reemplaza. Esta verificación puede parecer redundante dado que j
se encontro como la posición para la cual Valor_de_Peak es mayor que el valor
existenten en la posición j de la lista, sin embargo es necesaria cuando el la posición de
Valor_de_Peak corresponde a j = Largo_Lista */

if (img_in[j][pos]<peak)
    {img_in[j][pos]=peak;
    img_in_disp[k][pos]=disp;}
}

```

```

Búsqueda de Correlación Máxima Y Detección del Peak (...)
/*for todas las disparidades en el rango de operación do*/
for (disp=disp_min; disp<=disp_max; disp++)
{ /*for todos los elementos de la imagen de referencia do*/
  for (x=0, y=0; x<256&&y<ymax; x++, y++)
  { corr = obtiene_correlación (N, P,      /*Tamaño de Venatana*/
                               x, y,    /*Posición*/
                               disp);

  /*Si la disparidad disp se encuentra en el rango local.*/
  if ((disp_min_loc[pos]<=disp)&&(disp<disp_max_loc[pos]))
  { /*Si la correlación actual corr es mayor que la máxima que había corr_max.*/

    if (corr>img_corr_max[pos])
    { disp_RL[pos]=disp;
      img_corr_max[pos]=corr;
      img_corr_prev_max[pos]=img_corr_prev[pos];
    } else
      if (img_disp_RL[pos]==disp-1)
        img_corr_next[pos]=corr;

    /*Detecta el peak como un cambio en el sentido de crecimiento de la curva. Si
    corr es un peak, entonces actualiza la lista de peaks de la curva de
    correlación*/
    if (corr>img_corr_prev[pos])
      img_corr_peak_disp[pos]=disp;
    else if (img_corr_peak_disp[pos]==disp-1)
      { /*Correlación previa era un Peak*/
        Actualiza_Lista_de_Peaks (Lista_de_Peaks, img_corr_prev[pos],
                                Lista_de_Disparidades,
                                img_corr_peak_disp[pos], pos, Largo_Lista);
      }
    if (disp==disp_max&&corr>img_corr_prev[pos])
      Actualiza_Lista_de_Peaks (img_corr_peak_L, corr,
                                img_corr_peak_disp_L, disp, pos, NP);
    /*Fin de la Detección de Peaks*/
  }
}
}

```

### C.3 Cálculo de la Resolución Sub-Pixel mediante Interpolación Cuadrática

Una alternativa para encontrar la posición del máximo en una curva es la aproximación de la curva a una parábola de la forma:

$$y(x) = a \cdot x^2 + b \cdot x + c$$

Nos interesa encontrar el punto  $\hat{x}$  para el cual  $y(x)$  es máximo, por lo tanto derivando la expresión anterior con respecto a  $x$  se tiene que el máximo ocurre para:

$$y' = 2a \cdot x + b = 0 \rightarrow \hat{x} = \frac{-b}{2 \cdot a}$$

La expresión anterior revela que basta con conocer los coeficientes  $a$  y  $b$  para calcular  $\hat{x}$ . Para encontrar los 3 coeficientes desconocidos se requieren 3 puntos por los cuales pasará la parábola (por 2 puntos pasan varias parábolas que no necesariamente se aproximan mejor a la curva). Dados los puntos que se muestran en la figura C.3, se puede establecer en forma matricial el sistema de ecuaciones que satisfacen la función de la parábola:

$$\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Resolviendo el sistema de ecuaciones anterior dados los 3 puntos por los cuales pasa la parábola, se obtienen  $a$  y  $b$  para encontrar  $\hat{x}$  correspondiente a la posición del máximo interpolado. Los coeficientes  $a$  y  $b$  pueden expresarse como:

$$b = \frac{(x_1^2 - x_2^2)(y_2 - y_3) - (x_2^2 - x_3^2)(y_1 - y_2)}{(x_1^2 - x_2^2)(x_2 - x_3) - (x_2^2 - x_3^2)(x_1 - x_2)}$$

$$a = \frac{(y_2 - y_3) - b \cdot (x_2 - x_3)}{(x_2^2 - x_3^2)}$$

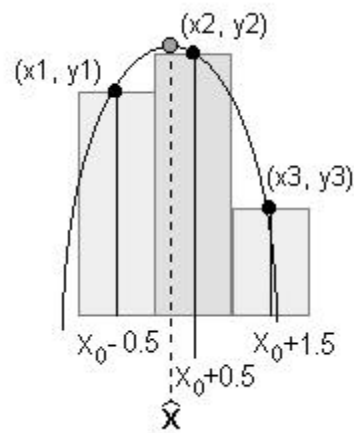


Figura C.3: Interpolación cuadrática

Utilizando para  $(x_1, y_1)$ ,  $(x_2, y_2)$ ,  $(x_3, y_3)$  los valores,

$$x1=disp-0.5; \quad y1=img\_corr\_prev\_max[pos];$$

$$x2=disp+0.5; \quad y2=corr;$$

$$x3=disp+1.5; \quad y3=img\_corr\_next[pos];$$

calculados a través del procedimiento descrito en la sección 2 del anexo C, se puede encontrar la disparidad sub-píxel **img\_disp** del elemento en la posición **pos** del mapa de disparidades como:

$$img\_disp[pos]=interp\_quad\_fast(x1,y1,x2,y2,x3,y3);$$

donde **interp\_quad\_fast** es la función que calcula los coeficientes a y b y devuelve la posición del máximo  $(-b/2a)$  como un valor del tipo float.

#### C.4 Cálculo del Rango de Búsqueda Ajustado Localmente

Supóngase que a partir de un análisis inicial de la escena con un rango de búsqueda amplio  $[d_{0min} - d_{0max}]$ , se genera un mapa de disparidades que contiene el valor de disparidad asignado a cada pixel de la imagen de referencia, por ejemplo la derecha en un sistema binocular. El mapa de disparidades inicial  $D(x, y)$  contendrá una cierta cantidad de disparidades erróneas debido a que se dan más ambigüedades al buscar en un rango extenso. Por los análisis realizados sabemos que el error promedio en la disparidad  $Dd$  normalmente, no supera los 1.5 pixels. Por esta razón podemos realizar una segunda búsqueda de correspondencia en torno al valor de disparidad encontrado inicialmente  $\pm Dd$ .

En las zonas que han sido definidas como no válidas por la etapa de validación no existe información previa, en torno a la cual buscar la disparidad. En este caso el rango de búsqueda se define con la ayuda de la figura C.4 como:

$$\begin{aligned}d_1(x, y)_{max} &= d_0(x, y)_{max} + \Delta d_{max} \\d_1(x, y)_{min} &= d_0(x, y)_{min} - \Delta d_{max}\end{aligned}$$

Las expresiones anteriores establecen el criterio para la determinación del *rango de disparidades ajustado localmente* en las zonas no determinadas. Suponiendo que no existen cambios demasiado abruptos en las disparidades dentro de estas zonas, una elección apropiada del rango tomaría como límites, los valores de disparidad más próximos en el sentido del barrido horizontal  $\pm Dd$ . El mínimo de los valores más cercanos menos  $Dd$  (-1.5 pixels) sería el límite inferior del rango. De manera análoga, el límite superior del intervalo de búsqueda se fijaría como la disparidad máxima más cercana más el error promedio de las disparidades  $Dd$ .

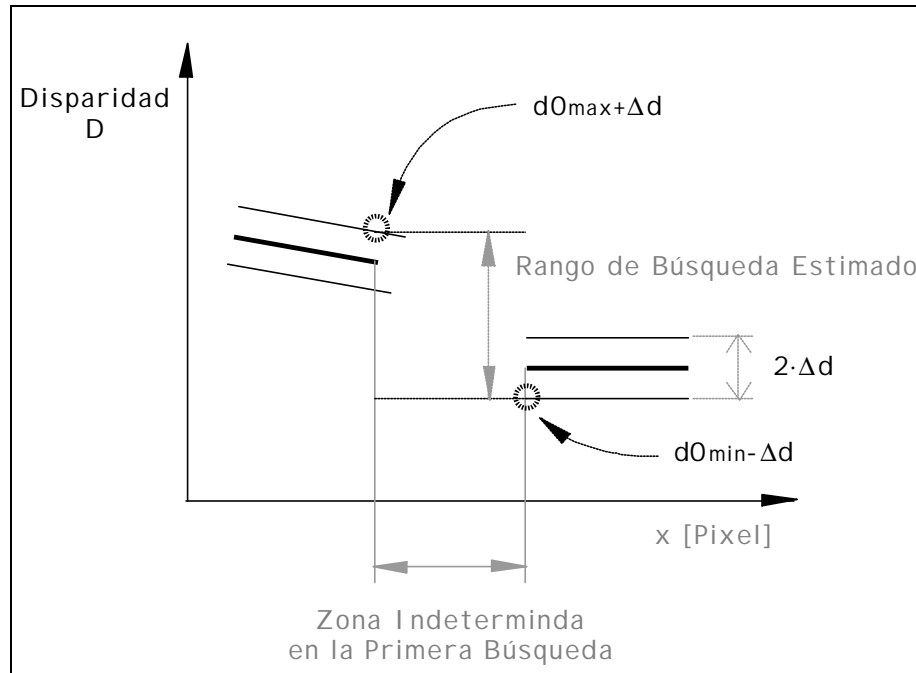


Figura C.4: Determinación del rango de búsqueda en zonas indeterminadas.

## GLOSARIO

Definición informal de algunos términos con la intención de refrescar la memoria.

- ASIC** Application Specific Integrated Circuit.
- Baseline** Distancia entre los centros ópticos de las cámaras.
- Azimuth** Giro Horizontal. Angulo que forma un plano vertical que corta la esfera con otro de referencia.
- Disparidad** Distancia entre las proyecciones de un punto en el espacio 3D sobre los planos de la imagen.
- Mapas de Disparidad:** Imagen en la que la intensidad de los pixels guarda cierta relación con el grado de disparidad.
- fps** Cuadros por segundo (*frames per second*).
- Foreshortening** Es el efecto por el cual las cuadrículas de una superficie enmallada parecen tener un área igual o menor a la real. Las cuadrículas aparentan tener un área cada vez más pequeña cuando sus normales son perpendiculares a la línea de vision.
- Imagen** En el caso discreto, matriz de pixels o pels (picture elements) que contienen un valor que representa una intensidad de luminosidad o un color en el caso de imágenes indexadas (aquellas que hacen referencia a una tabla (paleta) de colores).
- Plano de la Imagen:** Plano donde se forma la imagen. En el caso de la cámaras el CCD o el negativo. La retina en el ojo.
- Matching** Búsqueda de Correspondencias.
- MV** Machine Vision es el proceso de extracción de la información proporcionada por los sensores visuales para que las máquinas puedan



tomar decisiones “inteligentes”. Algunos sostienen que MV es parte de la percepción mediante visión artificial (*Artificial Visual Perception*).

<b>Pan</b>	Movimiento de una cámara en el sentido azimutal.
<b>Pitch</b>	Angulo de elevación de la cámara.
<b>RF</b>	Radio-Frecuencia.
<b>Roll</b>	Angulo de rotación de la cámara en torno a su eje óptico.
<b>Scan Line</b>	Línea de Barrido Horizontal.
<b>SSD</b>	Sum of Squared Differences o también llamada distancia Euclidiana entre dos vectores (vectores de pixels en procesamiento de imágenes) utilizado como estimador de la similitud en el proceso de correspondencia estéreo.
<b>Yaw</b>	Angulo de convergencia de la cámara (movimiento azimutal).

## INDICE GENERAL

### *Listado por Palabras*

#### **TRAILER NOT DISPOSABLE TRASH - LO QUE ESTA ENTRE LINEAS NO VA!**

En resultados discutir: Selección de NP.

Mapas de Disparidades vs. Tamaño de NP.

Curvas de Corr. vs. NP.

Curvas de Corr. vs. Largo del Intervalo (Sin y Con Patrones periódicos).

Curvas de Corr. vs. NP+nivel de textura.

Textura Baja -> NP grandes. Textura Alta -> NP bajos.

Patrones Horizontales.

Mapas de Correlación y Threshold.

Mapas de Confianza.

Un gran cantidad de los esfuerzos de la investigación sobre visión artificial se han destinado al estudio de 3D de los objetos mediante el análisis de imágenes. El análisis computacional de imágenes estereoscópicas para recuperar la información de profundidad se inicia a mediados de los 70's. El principio básico en el que se basa éste método pasivo es la triangulación. Aunque muchas técnicas activas de medición de distancias también se basan en triangulación, la naturaleza del problema es distinta ya que el triángulo para recuperar la profundidad está predefinido por tres puntos (la fuente de luz, el punto iluminado en la escena, y la imagen de ese punto). Por lo tanto, en los métodos activos el problema de correspondencia ha sido ya solucionado al incorporar una fuente de iluminación artificial.

La estereopsis se considera como un método pasivo porque la triangulación debe realizarse con ayuda de la luz del ambiente únicamente. Por lo tanto la correspondencia debe realizarse entre características físicas de algunos elementos en el espacio captados en dos o más imágenes desde perspectivas distintas. Aunque se ha tratado determinar la distancia a través de una única imagen a partir del análisis de las sombras (*shape from shading*) los resultados que se han obtenido sólo son satisfactorios en casos en los que muchas restricciones a la escena se han aplicado (forma de los objetos e iluminación), y no son prácticos ya que dependen fuertemente del conocimiento exacto de la fuente de iluminación (posición y orientación).

El problema de estimación pasiva de distancias es importante cuando existen limitaciones para utilizar iluminación artificial como ocurre frecuentemente en la práctica. Algunas de las aplicaciones basadas en la estereoscopía para la medición de profundidad incluyen la cartografía automatizada, navegación aérea, vehículos exploradores (rovers) autónomos, robótica, automatización industrial y stereomicroscopía.

## ANEXO D: TRANSFORMACIONES GEOMÉTRICAS TRIDIMENSIONALES

El manejo y visualización de imágenes tridimensionales requiere el empleo de geometría y transformaciones en coordenadas tridimensionales. Estas transformaciones están constituidas por la composición de las transformaciones básicas de translación, puesta en escala y rotación [Plas87].

### Translación

Un objeto es desplazado cierta distancia y dirección a partir de su posición original. La dirección y el desplazamiento de la translación están definidos por un vector

$$\mathbf{V} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k} \quad (\text{C.1})$$

Esto se puede representar en forma matricial en coordenadas homogéneas como

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & c \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (\text{C.2})$$

### Puesta en escala

El proceso de puesta en escala modifica las dimensiones de un objeto. El factor de escala  $s$  determina si la escala es una amplificación,  $s > 1$ , o una reducción,  $s < 1$ .

La puesta en escala con respecto al origen, donde dicho origen permanece fijo, se efectúa por la transformación

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (\text{C.3})$$

## Rotación

La rotación en tres dimensiones es más compleja que la de dos dimensiones. En dos dimensiones, una rotación está definida por un ángulo y un centro de rotación. Las rotaciones tridimensionales requieren la prescripción de un ángulo y un eje de rotación. Las rotaciones *canónicas* están definidas cuando se elige uno de los ejes coordenados positivos  $x$ ,  $y$  o  $z$  como eje de rotación. Así se tiene para cada eje

Eje  $x$ :

$$R_{q,i} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \mathbf{q} & -\text{sen} \mathbf{q} \\ 0 & \text{sen} \mathbf{q} & \cos \mathbf{q} \end{bmatrix} \quad (\text{C.4})$$

Eje  $y$ :

$$R_{q,j} = \begin{bmatrix} \cos \mathbf{q} & 0 & \text{sen} \mathbf{q} \\ 0 & 1 & 0 \\ -\text{sen} \mathbf{q} & 0 & \cos \mathbf{q} \end{bmatrix} \quad (\text{C.5})$$

Eje  $z$ :

$$R_{q,k} = \begin{bmatrix} \cos \mathbf{q} & -\text{sen} \mathbf{q} & 0 \\ \text{sen} \mathbf{q} & \cos \mathbf{q} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (\text{C.6})$$

El caso general de rotación en relación con un eje  $L$  puede construirse a partir de estas rotaciones canónicas mediante la multiplicación de matrices, resultando

$$\begin{bmatrix} \cos \mathbf{q}_y \cos \mathbf{q}_z & \text{sen} \mathbf{q}_z \text{sen} \mathbf{q}_y \cos \mathbf{q}_z - \text{sen} \mathbf{q}_z \cos \mathbf{q}_x & \cos \mathbf{q}_z \text{sen} \mathbf{q}_y \cos \mathbf{q}_x + \text{sen} \mathbf{q}_x \text{sen} \mathbf{q}_z \\ \cos \mathbf{q}_y \text{sen} \mathbf{q}_z & \text{sen} \mathbf{q}_x \text{sen} \mathbf{q}_z \text{sen} \mathbf{q}_y + \cos \mathbf{q}_x \cos \mathbf{q}_z & \cos \mathbf{q}_x \text{sen} \mathbf{q}_z \text{sen} \mathbf{q}_y - \text{sen} \mathbf{q}_x \cos \mathbf{q}_z \\ -\text{sen} \mathbf{q}_{yz} & \text{sen} \mathbf{q}_x \cos \mathbf{q}_y & \cos \mathbf{q}_x \cos \mathbf{q}_y \end{bmatrix} \quad (\text{C.7})$$

Se debe tener en consideración que el orden de multiplicación implica rotaciones distintas.

## ANEXO E: VISUALIZACIÓN DE SUPERFICIES TRIDIMENSIONALES

### Proyección en perspectiva

Una transformación perspectiva se establece al asignar un centro de proyección y un plano de vista para una escena tridimensional. El *plano de vista* está determinado por su *punto de referencia de vista*  $R$  y la *normal al plano de vista*  $N$ . El *punto objeto*  $P$  está localizado en las coordenadas  $(x, y, z)$ . El problema se reduce a encontrar las coordenadas del *punto imagen*  $P(x', y', z')$ .

Para el caso en que el plano de vista es el plano  $xy$ , y el centro de proyección se toma como el punto  $C(0, 0, -d)$  sobre el eje negativo  $z$ , se tiene

$$\begin{aligned}x' &= \frac{d \cdot x}{z + d} \\y' &= \frac{d \cdot y}{z + d} \\z' &= 0\end{aligned}\tag{D.1}$$

Para el caso en que  $d \rightarrow \infty$  se elimina el efecto de perspectiva y se obtiene una proyección paralela, como por ejemplo la representación isométrica.

### Modelos de superficies

Para poder obtener una descripción visual de las superficies tridimensionales en una proyección bidimensional, se pueden emplear un conjunto de primitivas o formas geométricas como líneas y polígonos.

La visualización más básica consiste en crear un modelo de “marco de alambre”, que al tener suficientes líneas enmalladas permite dar la ilusión de redondez y uniformidad, como se muestra en la figura D.1.

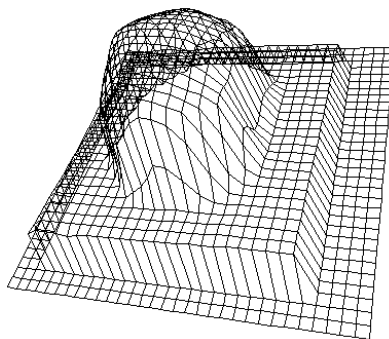


Figura D.1: Modelo de marco de alambre del par estéreo de la figura 3.5.

Esta representación tiene la ventaja de ser fácil de implementar y de tener tiempos de ejecución para el despliegue reducidos.

Sin embargo, para poder obtener una representación más realista de la escena a representar, se deben utilizar técnicas de remoción de superficies ocultas, sombreado y textura. En este caso se utilizan como primitivas polígonos o superficies polinomiales conocidas.

En una imagen con objetos opacos, aquellos que están más cerca del ojo y de la línea de visión de otros objetos los bloquearán a la vista. Las superficies ocultas deben eliminarse a fin de presentar una imagen realista en la pantalla. Para ello se utiliza el algoritmo de “Buffer Z”, que compara la información de profundidad de cada polígono proyectado con los demás que están sobre el mismo punto de proyección, decidiendo cuál de los dos se va finalmente a desplegar.

Para lograr un efecto aún mayor de realismo, se puede aplicar sombreado a cada polígono. El modelo que se utiliza generalmente para escenas simples con un sólo objeto de superficies *lambertianas* (que reflejan la luz hacia todas direcciones con la misma intensidad) es el de reflexión *difusa*, *especular* y *ambiental* [Irwi95]. En el caso de la reflexión *difusa*, la luz es reflejada uniformemente desde la superficie del objeto. En la reflexión *especular* una alta proporción de luz incidente es reflejada en un limitado rango angular, dándole al objeto un aspecto brillante. La reflexión *ambiental*, por último, es una aproximación a las reflexiones múltiples que existen en una escena que producen una iluminación general, para cualquier ángulo.

Para explicitar este modelo, se definen los vectores normalizados  $\mathbf{N}$ , normal a la superficie y el vector de luz  $\mathbf{L}$ , que apunta a la fuente luminosa,  $\mathbf{R}$  vector de reflexión y  $\mathbf{V}$  el vector del observador, como se muestra en la figura D.2.

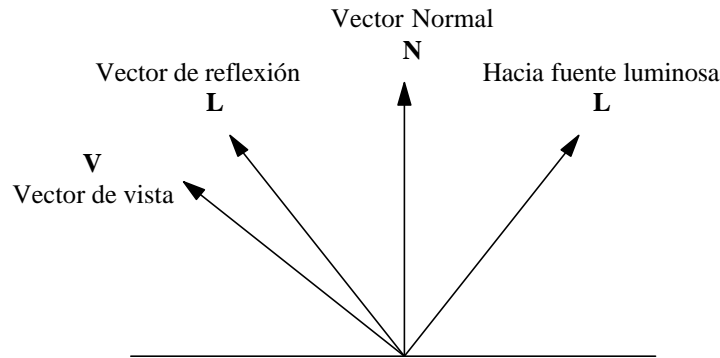


Figura D.2: Vectores de iluminación.

En este caso se utiliza una fuente de luz en el infinito, simulando al sol, lo que hace que  $\mathbf{L}$  sea paralelo para cualquier punto de la superficie.

Si se define  $I$  como la intensidad de reflexión en un punto de la superficie, se puede modelar la reflexión *ambiental* como

$$I = k_d I_a \quad (\text{D.2})$$

El término de *reflexión difuso* se puede definir por el producto punto de  $\mathbf{N}$  y  $\mathbf{L}$  como

$$I = k_d I_p (\mathbf{N} \cdot \mathbf{L}) \quad (\text{D.3})$$

El término de *reflexión especular* se puede definir por el producto punto de  $\mathbf{R}$  y  $\mathbf{V}$  como

$$I = k_e I_p (\mathbf{R} \cdot \mathbf{V})^n \quad (\text{D.4})$$

Luego, sumando los términos (D.2), (D.3) y (D.4), se tiene una intensidad de sombreado para cada punto de la proyección



$$I = k_d I_p + I_p [k_d (\mathbf{N} \cdot \mathbf{L}) + k_e (\mathbf{R} \cdot \mathbf{V})^n] \quad (\text{D.5})$$

El resultado de aplicar sombreado se aprecia en la figura D.3.

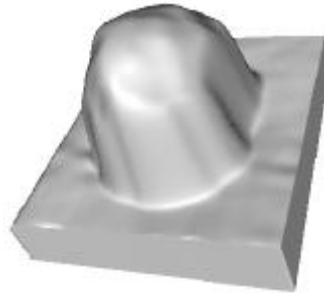


Figura D.3: Modelo sombreado de la disparidad del par estéreo de la figura 3.5.

Finalmente, se puede aplicar la textura original de la imagen bidimensional a la proyección tridimensional, de tal forma de dar mayor realismo a la representación. Para ello se puede ponderar la intensidad del sombreado con la intensidad original de la imagen, como se muestra en la figura D.4.



Figura D.4: Modelo sombreado y con texturas de la disparidad del par estéreo de la figura 3.5.