UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS FÍSICAS
Departamento de Arquitectura de Computadores y Automática

Sistema de reconocimiento y localización de objetos cuasi-esféricos por telemetría láser. Aplicación a la detección automática de frutos para el robot Agribot

TESIS DOCTORAL
1998

Antonio Ramón Jiménez Ruiz

Instituto de Automática Industrial (IAI)
Consejo Superior de Investigaciones Científicas (CSIC)

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS FÍSICAS
Departamento de Arquitectura de Computadores y Automática

Sistema de reconocimiento y localización de objetos cuasi-esféricos por telemetría láser. Aplicación a la detección automática de frutos para el robot Agribot

Memoria presentada por
D. Antonio Ramón Jiménez Ruiz
bajo la dirección del
Dr. D. Ramón Ceres Ruiz
para optar al grado de Doctor en Ciencias Físicas

Instituto de Automática Industrial (IAI)
Consejo Superior de Investigaciones Científicas (CSIC)

A mis padres

A Mª Ángeles

Índice General

Agradecimientos iii

Resumen de la tesis v

1 Automatización agrícola: La detección de frutos. El robot asistido AGRIBOT 1

1.1 La automatización agrícola 2

1.1.1 Factores socioeconómicos y tecnológicos en la automatización agrícola 3

1.1.2 Automatización en el sector agroalimentario 4

1.1.3 Automatización en la recolección de frutas delicadas 8

1.2 Percepción automática en los procesos de recolección 11

1.2.1 Variabilidad del entorno agrícola 11

1.2.2 Revisión de los principales métodos de detección de frutos propuestos para tareas de recolección agrícolas 12

1.2.2.1 Propuesta de Schertz y Brown 12

1.2.2.2 Universidad de Virginia: Parrish y Goksel 13

1.2.2.3 El proyecto MAGALI: D'Esnon y Rabatel 14

1.2.2.4 Universidad de Florida y centro USDA: Harrell y Slaughter 16

1.2.2.5 Universidad de Purdue: Whittaker 17

1.2.2.6 A.I.D: Levi, Falla y Pappalardo 18

1.2.2.7 Centro Sunkist y Universidad de California: Sites y Delwiche 19

1.2.2.8 El proyecto AUFO: Kassay 20

1.2.2.9 El proyecto CITRUS: Juste, Sevila, Plá, Moltó 21

1.2.2.10 Univ. de Purdue y C. Volcani: Cardenas, Dobrousin, Benady, Miles 22

1.2.2.11 CIRAA: Buemi 25

1.2.3 Análisis de las soluciones propuestas: Limitaciones 26

1.2.3.1 Captación de la imagen 26

1.2.3.2 Métodos de análisis aplicados 28
1.2.3.3 Resultados 29
1.3 La estrategia de recolección asistida: El robot AGRIBOT 29
1.3.1 Principio de operación asistido 30
1.3.2 Descripción del robot AGRIBOT 30
1.3.3 Resultados y discusión 35

2 Estrategia de detección mixta. Estudio, configuración y caracterización de un sensor de distancia y reflectancia 37
2.1 Estrategia de detección mixta: Detección/localización automática y asistida 38
2.1.1 Principio de operación mixto 38
2.1.2 Captación de distancia y reflectancia: Análisis de formas, propiedades ópticas y distribución espacial. 39
2.1.3 Especificaciones de las imágenes de distancia y reflectancia 40
2.2 Técnicas para captar distancia y reflectancia 42
2.2.1 Técnicas basadas en cámaras ópticas 42
2.2.1.1 Estéreo visión 47
2.2.1.2 Visión dinámica 48
2.2.1.3 Enfoque 49
2.2.1.4 Gradiente de textura 50
2.2.1.5 Gradientes de iluminación 51
2.2.1.6 Luz estructurada 51
2.2.2 Técnicas ultrasónicas 52
2.2.3 Palpación táctil 55
2.2.4 Técnicas basadas en telemetría láser 58
2.2.4.1 Telemetría láser mediante triangulación 61
2.2.4.2 Telemetría láser por diferencia de fases 62
2.2.4.3 Telemetría láser por tiempo de vuelo 64
2.2.4.4 Telemetría láser por frecuencia modulada 65
2.3 Análisis, selección y configuración del sistema de medida 66
2.3.1 Principio de medida: Telemetría láser por diferencia de fases. . . . 66
2.3.2 Telémetro láser: Selección y análisis tecnológico 68
2.3.2.1 Selección: Acuity Range 4000-LIR 68
2.3.2.2 Análisis tecnológico: Índice de eficiencia tecnológica . . . 70
2.3.3 Configuración física del sistema de exploración 74
2.3.3.1 Componentes del sistema de deflexión 74
2.3.3.2 Aspectos de seguridad ante la radiación láser 76
2.3.3.3 Adquisición de las imágenes 81

2.3.3.4 Limitaciones encontradas 82
2.4 Caracterización del telémetro láser 84
2.4.1 Modelado de los factores que afectan a la reflectancia registrada . 84
2.4.1.1 Planteamiento inicial del modelo 85
2.4.1.2 Modelo general: Superficies no ideales 88
2.4.1.3 Modelo aproximado: Superficies difusoras perfectas . . . 91
2.4.1.4 Reflectividad y fracción difusa: Propiedades características de una superficie 93
2.4.2 Modelado de la repetitividad de la distancia 94
2.4.2.1 Repetitividad en régimen estático 95
2.4.2.2 Repetitividad en situaciones dinámicas 98
2.5 Conclusiones 101

3 Restauración del mapa de distancias: Filtrado adaptativo y calidad de restauración 105
3.1 Técnicas de preprocesamiento de imágenes 106
3.1.1 Métodos de realce 107
3.1.1.1 Transformaciones del histograma 107
3.1.1.2 Acentuado de bordes 108
3.1.1.3 Coloreado 109
3.1.2 Métodos de restauración 110
3.1.2.1 Filtros lineales 110
3.1.2.2 Filtros no lineales 112
3.1.2.3 Suavizado adaptativo 116
3.1.2.4 Suavizado adaptativo iterativo 120
3.2 Evaluación de la calidad de restauración 121
3.2.1 Definición de una nueva métrica de calidad de restauración: *GRI* . 123
3.2.2 Evaluación de técnicas de restauración mediante la métrica *GRI* . 126
3.3 Nuevo método de restauración adaptativa por ajustes de planos multiresolución con fidelidad-3σ 128
3.3.1 Definición de la técnica de restauración 3σ-*MPF* 129
3.3.2 Evaluación comparativa de la técnica de restauración 3σ-*MPF* por la métrica *GRI* 132
3.4 Conclusiones 140

4 Método de reconocimiento de objetos esféricos: Estrategia modular de acumulación de indicios mediante múltiples primitivas 145
4.1 Técnicas principales de análisis de imágenes 146
4.1.1 Segmentación 147

4.1.1.1 Técnicas para la detección de bordes 147
4.1.1.2 Segmentación por agregación y detección de contornos . . 151
4.1.1.3 Segmentación por umbral 153
4.1.1.4 Segmentación basada en características de regiones 154
4.1.1.5 Segmentación por movimiento 156
4.1.2 Descripción o extracción de características 156
4.1.2.1 Descriptores de contorno 157
4.1.2.2 Descriptores de regiones. 159
4.1.2.3 Descripción de estructuras tridimensionales 162
4.1.3 Reconocimiento o clasificación 164
4.1.3.1 Teoría de decisiones 164
4.1.3.2 Métodos estructurales 171
4.2 Estrategia de reconocimiento propuesta 173
4.2.1 Consideraciones iniciales sobre métodos de reconocimiento 173
4.2.2 Estrategia de reconocimiento basada en la extracción de primitivas y en la acumulación de indicios 175
4.3 Definición y generación de primitivas 178
4.3.1 Primitivas puntuales contorno 181
4.3.1.1 Características 185
4.3.1.2 Funciones discriminantes 187
4.3.1.3 Clasificación de los puntos imagen 191
4.3.2 Primitivas puntuales corona 193
4.3.3 Primitivas regionales convexas 199
4.3.4 Primitivas regionales reflectividad 201
4.4 Estimación de parámetros e indicios 204
4.4.1 Estimación con primitivas puntuales 205
4.4.2 Estimación con primitivas regionales 209
4.5 Generación de hipótesis finales y verificación 211
4.6 Conclusiones 212

5 Evaluación conjunta del sistema y los algoritmos propuestos en condiciones simuladas y reales 215
5.1 Introducción a la evaluación de algoritmos 216
5.1.1 Consideraciones iniciales en la evaluación 216
5.1.2 Conceptos fundamentales en la evaluación de algoritmos 217
5.2 Evaluación de los algoritmos propuestos mediante simulación 219
5.2.1 Complejidad algorítmica 219
5.2.2 Sensibilidad ante degradaciones 220

5.2.3 Influencia de la restauración sobre los resultados 231
5.3 Evaluación de los algoritmos propuestos utilizando datos empíricos 233
5.3.1 Evaluación de la capacidad de detección con imágenes de laboratorio 234
5.3.2 Pruebas de detección en una plantación natural 239
5.3.2.1 Discriminación por reflectividad 242
5.3.2.2 Influencia del Sol 245
5.3.2.3 Influencia del viento 246
5.4 Conclusiones 247

Conclusiones, aportaciones y trabajos futuros **253**

A Hojas técnicas de los componentes del sistema telemétrico de exploración **257**

B Medidas experimentales para modelar el telémetro láser **269**

C Muestra de imágenes de laboratorio distancia-reflectancia utilizadas en la evaluación empírica **275**

Agradecimientos

En primer lugar quisiera agradecer a mi director de tesis, el Dr. D. Ramón Ceres Ruiz, la oportunidad brindada para realizar el trabajo de investigación ligado al proyecto Agribot que ahora culmina con la finalización de esta tesis. Destacando la gran ayuda prestada con sus consejos, su manera de enfocar los problemas y los esfuerzos en perfeccionar el trabajo realizado. Junto con Leopoldo y José Miguel el trabajo se hizo más fácil y se creó un entorno de trabajo agradable, se proporcionaron todos los recursos materiales necesarios siendo esto muy importante para trabajar con eficacia.

Gracias también a mi hermana María Antonia, que asumiendo muy bien el papel de mayor de los hermanos, ha estado siempre pendiente de mí, preocupándose de mis problemas y tratando de guiarme por el mejor camino. Quiero así mismo recordar a Teo, Eduardo, Vicente y José Manuel que estuvieron trabajando junto a mí al principio, y de los cuales aprendí las primeras cosas de ultrasonidos. Cómo no, agradecer a mis compañeros de tesis José Luis y Enrique, al primero por los consejos y ánimos dados para seguir adelante, y al segundo por los "rollos" filosófico-técnicos que me concedió, con los que yo me distraía al cambiar de tema, y él se desahogaba contando sus logros e incertidumbres. También recordar al resto de compañeros del departamento, unos aún aquí y otros que ya se fueron, sin los cuales las cosas no hubiesen sido iguales.

Quiero expresar mi agradecimiento a todo el personal del IAI, que bajo la dirección del Dr. D. Antonio Cordero, han hecho posible que las diferentes tareas que implicaba el trabajo realizado se hayan desarrollado de una forma distribuida. En especial me gustaría agradecer a Angel Rebollo la cantidad de "cables" que me ha echado (en todos los sentidos), y a Román Cordero por su capacidad para distribuir la tarea entre los talleres de mecanizado, electrónica y delineación, agilizando el trabajo hasta donde era posible.

Finalmente, agradecer al MEC-CICYT la beca de posgraduado concedida y las oportunidades en forma de salidas de perfeccionamiento al extranjero, las cuales han resultado ser muy positivas tanto para mi formación académica como personal. En primer lugar, agradecer al Prof. H. Van Brussel, Dr. D. Reynaerts y al Dr. Adhi S. Soembajijo las facilidades prestadas en todo momento para realizar los primeros trabajos con imágenes en la Universidad de Lovaina. Igualmente, la estancia en la Universidad Estatal de Michigan con el profesor A.K. Jain y S. Shirai fueron de indudable valor, así como el trabajo desarrollado en la Universidad de Surrey bajo la supervisión de J. Illingworth y J. Kittler donde aprendí más aspectos del procesamiento y análisis de imágenes.

Resumen de la tesis

Objetivos

El trabajo expuesto en la presente memoria se ha desarrollado en el Instituto de Automática Industrial del *CSIC*, dentro del proyecto denominado *Robot Ligero Orientado a la Recolección Asistida de Frutos (AGRIBOT)*, quedando ligado al proyecto continuación del anterior, *Optimización de la Estructura y Sistemas Localizador y Prensor para el Robot Recolector de Frutos, AGRIBOT II.* Ambos han sido financiados por la Comisión Interministerial de Ciencia y Tecnología en el área de Tecnologías Avanzadas de la Producción (*CICYT-TAP93-0583 y TAP96-0398*).

El objetivo de estos dos proyectos se centra en la investigación de nuevas estrategias para tratar de automatizar en un cierto grado las tareas de recolección de frutos delicados en árbol. Este planteamiento ha requerido principalmente el estudio y la construcción de tres sistemas complementarios como son el brazo robótico, el dispositivo de prensión del fruto y el sistema de localización de frutos. Sobre este último aspecto, trata nuestro trabajo de tesis.

El *objetivo de la presente tesis* consiste en el estudio y realización de un sistema de detección y localización automático de frutos en árboles mediante telemetría láser. Con este trabajo se pretende profundizar en una metodología alternativa a las estudiadas por la mayoría de los investigadores anteriores basadas principalmente en visión artificial mediante cámaras. El problema esencial que se trata de resolver es detectar el mayor número posible de frutas con la asignación de sus correspondientes coordenadas espaciales y con la mínima tasa de falsas detecciones; todo ello teniendo en consideración la complejidad de los entornos agrícolas.

En este sentido, la tesis se centra en el estudio de nuevas técnicas de captación, procesamiento y análisis de imágenes, utilizando la información dual de distancia y reflectancia captada con un sistema sensorial basado en telemetría láser, con el fin de facilitar la obtención de información discriminante y simplificar la complejidad de los algoritmos. Estas técnicas deberán ser aplicadas en la elaboración de un prototipo de sistema de reconocimiento automático de frutos, que será planteado con mayor generalidad como un dispositivo de reconocimiento, localización y caracterización de objetos esféricos en entornos no estructurados.

Metodología del trabajo y estructura de la memoria

La metodología seguida para llevar a cabo el presente trabajo de investigación se corresponde con las etapas y aspectos que se presentan a continuación.

En el capítulo 1, comenzamos presentando las diversas técnicas de automatización, ya incorporadas o en fase de investigación, para procesos agroalimentarios. En especial, revisamos las diferentes investigaciones dedicadas a la automatización de la recolección de fruta delicada en árboles leñosos, y en particular los procesos de percepción artificial desarrollados para dotar a los robots de la capacidad de detectar y localizar los frutos a ser recolectados. Posteriormente, presentamos una descripción del proyecto *AGRIBOT*, y de su estrategia asistida o semi-automática con la que primeramente se concibe el proyecto. Esta estrategia supone la intervención de un operario para realizar las tareas de guiado del robot, reconocimiento y señalización del fruto. Se realiza un análisis crítico de los resultados obtenidos en esta primera fase de desarrollo del proyecto, y se plantea la necesidad de incorporar un sistema automático de detección y localización que apoye al método asistido.

En el segundo capítulo, proponemos una *estrategia de detección mixta* que integra el método automático con el asistido ya existente en el robot Agribot. Para incorporar el modo automático de detección de fruta, en vista de las limitaciones de los métodos desarrollados por otros autores basados en cámaras CCD, se propone fundamentar el reconocimiento en el análisis de formas, propiedades ópticas y distribuciones espaciales de los objetos. Para ello se requiere utilizar dispositivos de medida que suministren información de distancia y reflectancia, con lo cual se revisan diferentes técnicas candidatas, encontrándose que la telemetría láser es la más adecuada para cumplir las especificaciones marcadas. Basándonos en un tipo de telémetro que opera según el principio de diferencia de fases, configuramos un sistema de deflexión de haz que mediante barridos sucesivos permite obtener parejas de imágenes distancia-reflectancia. Finalmente, se realiza una caracterización del telémetro láser, presentando un modelo matemático que relaciona la información de reflectancia generada por el sensor, con otros factores que intervienen en el proceso de medida. Este modelo va a ser la base para realizar la integración de la información presente en las imágenes de distancia y reflectancia, permitiendo obtener propiedades más discriminantes como se verá en el capítulo 4 que trata del análisis de imágenes. Así mismo, se presenta otro modelo que relaciona la desviación estándar que caracteriza la repetitividad de la medida en distancias, con el nivel de señal o reflectancia captada por el sensor, lo cual es útil a la hora de estimar el ruido presente en la imagen de distancias. Dicha estimación se utilizará en la restauración de las imágenes que se presentará en el capítulo 3.

Efectivamente, las imágenes de distancia están contaminadas fundamentalmente de ruido con distribución gausiana y este factor es sin duda perjudicial a la hora de intentar extraer información de ellas. En el capítulo 3 estudiamos como minimizar este problema, comenzando con una revisión de las técnicas más habituales para la restauración y filtrado de imágenes. Ante la dificultad que existe a la hora de determinar qué técnica de restauración es la más adecuada, presentamos una nueva métrica, *GRI*, que mide la calidad de una restauración. Dicha métrica considera tanto la fidelidad en la preservación de bordes como el grado de suavidad proporcionado, y permite de esta forma disponer de un criterio objetivo de clasificación de técnicas de restauración. Se hace una

comparación de las técnicas tradicionales de filtrado no iterativas, encontrándose que no son totalmente satisfactorias para nuestros objetivos. Por este motivo presentamos una nueva técnica adaptativa de filtrado, denominada *3σ-MPF*, que permite superar en calidad de restauración a las técnicas anteriores de acuerdo con las evaluaciones realizadas mediante la métrica *GRI*.

Una vez que las imágenes distancia-reflectancia están adecuadamente restauradas, estamos en condiciones de aplicar métodos de análisis sobre ellas con el propósito de detectar la presencia de frutos, localizarlos y caracterizarlos. En el capítulo 4, inicialmente mostramos una revisión de las principales técnicas de análisis utilizadas en el campo de la visión por computador. A continuación presentamos la estrategia de análisis propuesta, la cual se divide en tres etapas. La primera consiste en una fase de extracción de cuatro tipos de primitivas, de las cuales tres (contorno, corona y convexidad) se caracterizan por tener una alta probabilidad de pertenecer a un objeto esférico y la restante (reflectividad) identifica regiones con propiedades ópticas iguales a las del objeto buscado. La segunda fase consiste en una estimación de los parámetros de la esfera a partir de los indicios generados por cada una de las primitivas. Finalmente, en la última fase se generan las hipótesis definitivas después de agrupar hipótesis parciales compatibles y de aplicar una etapa de eliminación de casos incoherentes. La principal novedad del método está en la extracción de primitivas, las cuales fueron especialmente definidas para captar indicios o pistas que den evidencia de la presencia de objetos esféricos. Este sistema es modular en el sentido de que se puede utilizar el número y los tipos de primitivas que se estimen adecuadas, siempre y cuando, mediante ellas se continúe captando la información discriminante que permita realizar el reconocimiento. Así mismo, el modelo del telémetro láser desarrollado en el capítulo 2, es utilizado durante el proceso de análisis para integrar la información proveniente de las imágenes de distancia y reflectancia. Los algoritmos de reconocimiento y localización que presentamos son capaces de generar como salida las coordenadas tridimensionales donde se encuentra cada objeto detectado, así como el radio de la esfera y un valor medio de reflectividad correspondiente a la superficie visible de ese objeto.

En el capítulo 5 se presenta una evaluación conjunta de los algoritmos propuestos en los dos capítulos anteriores y ciertos aspectos que afectan al sistema al operar en condiciones naturales. Se realizan dos tipos de evaluaciones, una simulada y otra experimental. En el primer caso los criterios elegidos para medir la calidad de los resultados generados por los algoritmos son los errores en la estimación de los parámetros de la esfera y el número de indicios generados que indican la capacidad de detección de esferas. El estudio se hace variando tanto parámetros internos del algoritmo como las características de las imágenes a analizar. En la evaluación experimental se utilizan imágenes correspondientes a escenas de árboles frutales tanto artificiales como naturales, y lo que se evalúa es la tasa obtenida de detecciones correctas y falsas. En el estudio con diferentes escenas de laboratorio se muestra que en torno a un 80% de la fruta visible es detectada y la probabilidad de que se den detecciones falsas es prácticamente nula. Las pruebas de campo realizadas indican que los frutos son discernibles del fondo por reflectividad y permiten detectar dos factores degradantes que influyen en la calidad de las imágenes captadas; estos son el viento y la iluminación solar.

Finalmente, se presentan las conclusiones finales, las aportaciones realizadas y las posibles líneas futuras de investigación en este campo.

Capítulo 1

Automatización agrícola: La detección de frutos. El robot asistido AGRIBOT

Resumen. *En este capítulo presentamos las principales tendencias de automatización en labores agrícolas. En especial estudiamos los trabajos de automatización en tareas de recolección de frutas y hortalizas, presentando las mayores dificultades encontradas: guiado, detección y captura. Se profundiza en los trabajos para la detección y localización de frutos, dando una amplia revisión de los desarrollos previos más significativos y haciendo un análisis crítico de las estrategias aplicadas tanto en la captación como en el análisis de la información.*

Finalmente presentamos la estrategia de recolección asistida contemplada en un primer enfoque del proyecto Agribot (I.A.I.), la cual pretende hacer viable la recolección semiautomática, dividiendo las tareas entre el hombre y el robot. Los resultados obtenidos en este trabajo son presentados y discutidos, planteando finalmente la necesidad de incorporar un sistema adicional de apoyo en la detección y localización automática de la fruta.

En la prehistoria los hombres vivían de la caza, la pesca y de la recolección de algunos productos. Se alimentaban con plantas y frutos comestibles que encontraban en los bosques. No vivían en lugares fijos, sino que recorrían el país en busca de alimentos. Cuando se instalaron en un lugar fijo, comenzaron a trabajar la tierra que les rodeaba y a cultivar en ella plantas comestibles. Al principio, estos cultivos les bastaban para satisfacer sus propias necesidades y constituían un suplemento a la alimentación que obtenían mediante la caza o la pesca. Más tarde comenzaron a repartirse las diferentes tareas, correspondiendo a unos la caza o la pesca y a otros el cultivo de las plantas que deberían servir de alimento a toda la comunidad. Al descubrir que algunos animales salvajes se podían criar en cautividad, comenzó la ganadería. Poco a poco, los asentamientos fueron aumentando de tamaño y organizándose en la misma medida los trabajos.

Posteriormente, las explotaciones agrícolas se hicieron a mayor escala. Se fabricaron herramientas sencillas, como la azada, para escardar y el arado para labrar la tierra. Al principio estas herramientas eran de piedra y de madera, pero, posteriormente, se hicieron de metal. Inicialmente, eran los propios hombres quienes tiraban de los arados, más tarde se dieron cuenta de que este trabajo podían realizarlo ciertos animales. La introducción de esta rudimentaria mecanización permitió el labrado de mayores parcelas de terreno y la obtención de más copiosas cosechas. En el transcurso de los siglos, mejoraron los métodos de cultivo, se comenzó a utilizar insecticidas y abonos químicos, y se inventó toda una serie de maquinaria agrícola de naturaleza fundamentalmente mecánica (tractores, cosechadoras, trilladoras, ...), que facilitó y aumentó en muy alto grado el rendimiento de las tareas agrícolas.

En la actualidad, las exigencias crecientes de calidad y de variedad de oferta a precios cada vez más reducidos, están exigiendo el empleo de diversas tecnologías para mejorar la producción. La automatización está siendo uno de los principales caminos emprendidos y está permitiendo la realización de labores complejas que anteriormente eran propias del hombre, pudiéndose atribuir muchos de estos logros a la utilización de computadoras y sensores que permiten realizar sistemas automáticos suficientemente versátiles como para adaptarse a los cambios en el medio.

1.1 La automatización agrícola

Por automatización entendemos "la acción de sustituir en un proceso el operador humano por dispositivos mecánicos o electrónicos" (Real Academia Española de la Lengua). De esta forma labores que eran peligrosas, tediosas o que estaban limitadas por la capacidad humana, mediante la automatización pueden ser elaboradas incluso más eficientemente, sin causar riesgos al operario y permitiendo dedicar sus capacidades a otras labores de menor aporte energético y mayor contribución intelectual, en las que su intervención es actualmente imprescindible. Los procesos automatizados están evolucionando actualmente desde la realización de operaciones secuenciales y repetitivas en ambientes normalmente estructurados, a tareas cada vez más complejas y cambiantes en las que es preciso tomar decisiones y adaptar dinámicamente los planes de actuación en función de las condiciones del entorno.

Fundamentalmente el sector industrial es el que ha experimentado un mayor auge en

la incorporación de nuevas tecnologías que permiten automatizar un gran número de procesos anteriormente realizados por el hombre. En el sector agrario, sin embargo, la incorporación de sistemas automáticos ha sido más limitada. A pesar de ello, aunque quizás a un ritmo lento, cada vez son más las tareas agropecuarias que se ven beneficiadas por estas nuevas tecnologías.

1.1.1 Factores socioeconómicos y tecnológicos en la automatización agrícola

Actualmente, en todos los sectores y en especial en el sector agroalimentario, las exigencias cada vez más rigurosas en cuanto a producción, calidad y presentación de los productos, está haciendo que se requiera la incorporación de tecnologías avanzadas para mejorar la producción. Este hecho unido al factor de que la mano de obra en el sector primario ha sufrido un desplazamiento hacia la industria y al sector servicios [29], hace que dicha mano de obra sea cada vez más escasa, siendo además poco atractiva para los trabajadores debido al carácter temporal y a la dureza de este tipo de labores. Por estos motivos, la incorporación de máquinas con mayor o menor grado de automatización permite aumentar la producción, diversidad y calidad de los productos, y logra reducir costes debido a la menor mano de obra necesaria.

A pesar de los factores ventajosos que proporciona la automatización, como ya se ha dicho, el desarrollo producido en el sector agrícola ha sido escaso. Las razones fundamentales que lo justifican son, por un lado, aspectos problemáticos de tipo socioeconómico, y por otro, motivos técnicos.

Los *motivos socioeconómicos* están directamente relacionados con la atomización de la explotaciones, el carácter estacional de los cultivos, el bajo valor unitario del producto, la tradicional escasez de inversión y la falta de preparación de la mano de obra actual para aceptar nuevas alternativas de ayuda a la producción. La atomización de las explotaciones hace que cualquier incorporación de nueva tecnología suponga una relativa fuerte inversión que no es capaz de costear el propietario. Este hecho unido al carácter estacional de la mayoría de los productos, que implica que una sembradora, fumigadora o sistema de recolección solo va a ser utilizado durante unos días u horas a lo largo del año, hace que no se estimule la inversión y se sigan utilizando métodos tradicionales.

Los *motivos técnicos* relacionados con la escasa implantación de tecnología en sectores agroalimentarios son, la falta de uniformidad de los productos y de los propios entornos naturales de producción, en los cuales se requiere una operación a la intemperie sufriendo unas condiciones orográficas y meteorológicas muy variables y en algunos casos adversas. Esta falta de estructuración del entorno o ausencia de uniformidad, hace que los sistemas automáticos deban estar dotados de una cierta "inteligencia" de tal forma que puedan adaptarse y operar con diferentes comportamientos al percibir estos cambios. Esta adaptabilidad requerida de los sistemas automáticos, supone un desafío tecnológico lo que precisamente constituye un área de gran actividad en la investigación actual; por este motivo, las automatizaciones en este sector ya implantadas han sido aquellas que suponen una menor incertidumbre y que se caracterizan por presentar una mayor uniformidad y ambientes más controlados.

1.1.2 Automatización en el sector agroalimentario

A pesar de las dificultades para la incorporación de sistemas automáticos en la industria agroalimentaria, existen suficientes procesos automatizados, o bien en vías de automatización, como para que merezcan una breve descripción. Vamos, por tanto, a presentar una relación de los aspectos relacionados con la automatización en labores agrícolas, comenzando con las operaciones iniciales de preparación del terreno y sembrado, y finalizando con el empaquetado previo al consumo final [29].

- *Preparación de terrenos.* En este apartado se incluyen todas las labores previas al cultivo incluyendo la eliminación de hierbas, el preabonado y el nivelado de terrenos. Todas estas tareas se realizan utilizando tractores comerciales a los que se les agregan dispositivos de accionamiento y sensores específicos para realizar las diversas labores (fig. 1.1). Los principales trabajos de automatización en esta etapa están relacionados con el autoguiado de los tractores [93, 151, 71, 16, 200]. Uno de los primeros trabajos en este sentido fue el prototipo del NIAE (National Institute of Agricultural Engineering - Reino Unido) donde se incorpora un sistema de posicionamiento global de hiperfrecuencias que toma la posición a partir de reflexiones en balizas pasivas fijas. Para el guiado local utiliza sensores ultrasónicos que detectan y siguen los surcos en las plantaciones. El nivelado de terrenos es otro aspecto importante en aquellas plantaciones donde se requiere un riego en manta intentando conseguir una capa uniforme de agua. En este sentido se utiliza un emisor láser acoplado al tractor emitiendo haces nivelados. El desplazamiento vertical medido sobre unas balizas fijas, permite determinar el desnivel, posibilitando la acción correctora de unas palas niveladoras. Entre los trabajos para la detección y eliminación de hierbas, recientemente se ha presentado un sistema que detecta las hierbas por visión artificial y posteriormente las elimina aplicándolas descargas de alta tensión mediante un brazo robotizado [175]. Otras estrategias para la eliminación de malas hierbas con un ahorro importante en herbicidas, consiste en la detección precisa de estas mediante visión para, posteriormente, realizar un tratamiento localizado [11].

- *Cultivo.* Las máquinas de sembrado de grano son los dispositivos más desarrollados, constando de un tractor al que se acopla un remolque con el grano y el mecanismo de distribución de éste. La técnica más común es la que utiliza un sistema de dispersión centrífugo que distribuye el grano uniformemente a medida que el tractor avanza. Las primeras soluciones presentadas se basan en técnicas mecánicas, sin embargo la reciente incorporación de la electrónica ha permitido realizar sistemas de distribución de grano con diferentes configuraciones, como la distribución en rombo que presenta algunas ventajas en cierto tipo de cultivos. Durante el proceso de crecimiento de las plantas la tendencia actual consiste en configurar las plantaciones de tal forma que se faciliten las posteriores tareas automáticas de recolección al aumentar la visibilidad de los frutos. Este cambio de fisonomía de las plantaciones se produce en algunos casos aplicando diversas técnicas de poda [112] y en otros mediante el guiado fibrilar de las plantas [207, 121]. Para dotar a las plantas leñosas de nuevas propiedades se recurre a diversos tipos de injertos, los cuales se empiezan a realizar de forma automática siendo el robot ROSAL [19] y los trabajos de Hwang [87] destacados ejemplos representativos.

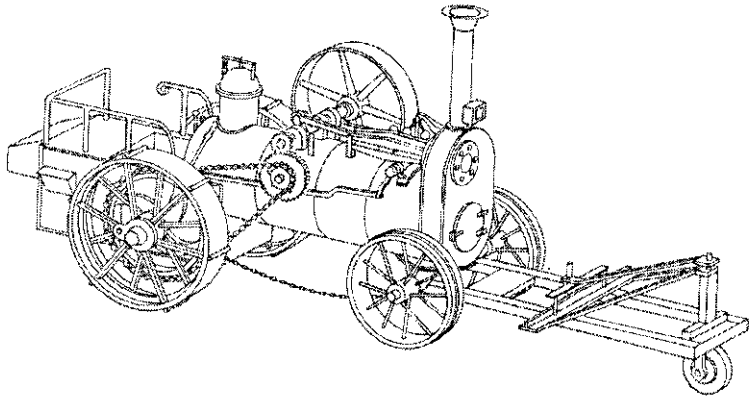


Figura 1.1: Uno de los primeros tractores a vapor que entró en servicio en 1860, que requería la intervención de tres personas para su manejo.

- *Fertirrigación.* Los primeros sistemas automáticos de riego realizados eran en lazo abierto suministrando una cantidad de agua fija previamente programada. Posteriormente mediante la utilización de sensores (termómetros, pluviómetros, anemómetros, ...) es posible realizar un control del grado de humedad del terreno y realizar un riego mucho más flexible adaptándose a las condiciones meteorológicas. El abonado automático se realiza disolviendo mediante bombas volumétricas sales minerales en el agua utilizada para el riego. Se utilizan sondas para medir el estado del suelo (acidez, humedad, salinidad, ...) de tal forma que se puede evaluar de una forma precisa las necesidades en agua y sales minerales requeridas por la tierra [206, 52]. El tratamiento local sobre la plantación utilizando robot móviles, permite aplicar de forma precisa las sustancias requeridas y a la vez conseguir ahorros del producto cercanos al 90% [136].

- *Recolección.* En este proceso cabe diferenciar entre dos grupos de cultivos que hacen que el proceso de recolección sea diferente: los *granos* (cereales, maíz, zahína,...) y las *frutas/hortalizas.* En el primer caso se corta y se recoge toda la planta, dejándose para una etapa posterior la separación del grano y la paja. La recolección de este tipo de productos está resuelta de forma satisfactoria mediante el uso de *cosechadoras* las cuales podrán ser completamente automatizadas mediante técnicas de autoguiado basadas en GPS (Global Positioning System) o análisis visual automático de la zona de operación mediante cámaras. En el caso de frutas y hortalizas, solamente se recolecta el fruto, y la planta se conserva para posteriores cosechas. Si la planta es de tipo leñoso y el fruto no es delicado (almendras, aceitunas, ...) es posible aplicar vibradores de tronco o de rama desprendiendo el fruto del árbol y dejándolo caer sobre una red. Sin embargo, en el caso de fruta delicada (manzanas, naranjas, tomates, ...) esta técnica no se puede utilizar y se requiere una recolección individualizada que actualmente se realiza de forma manual (fig. 1.2). Existen diversos trabajos de investigación en este campo en los

Figura 1.2: Método y equipo actual para la recolección de naranjas.

cuales se contempla la utilización de robots manipuladores, los cuales deben realizar tareas de navegación, detección, localización, agarre, corte y depositado de cada una de las frutas [148, 207, 121, 160, 45, 189, 213, 130, 187, 117, 112, 10, 24, 75].

En la siguiente sección veremos con mayor detalle aspectos relacionados con la recolección de fruta delicada de forma individualizada, puesto que este tema está directamente relacionado con el objetivo del presente trabajo; y en los siguientes capítulos nos centraremos en las etapas de reconocimiento y localización de los frutos en modo automático que es el objetivo final de la presente tesis.

- *Inspección y clasificación.* En esta etapa lo que se pretende es analizar por unidades la calidad de la producción, y en base a ella realizar una clasificación de los productos en diferentes grupos con semejantes parámetros de calidad, entre los cuales se encuentra también el grupo con unidades defectuosas. Para evaluar la calidad de un producto es necesario medir diferentes parámetros entre los cuales están: el color, tamaño, forma, firmeza, textura, peso, sabor, aroma y presencia de defectos tanto superficiales como internos. Para determinar estos parámetros existen un amplio rango de métodos no destructivos los cuales miden propiedades físicas como el peso, la densidad, rigidez y respuesta acústica, y propiedades electro-ópticas mediante visión artificial, rayos X e impedancias eléctricas.

  El proceso de inspección y clasificación tiene un grado de automatización alto existiendo por tanto bastantes sistemas autónomos ilustrativos [199]. Entre ellos podemos citar el proyecto SHIVA el cual contempla la realización de un sistema robotizado para la inspección, manipulación y empaquetado de frutas y hortalizas.

El aspecto innovativo en el proyecto está en la utilización de nuevos sensores que permiten obtener información de propiedades tanto internas como externas del producto. Otros trabajos incluyen la automatización en la determinación de la firmeza y madurez de frutas mediante técnicas mecánicas [161, 18], la detección de defectos superficiales [106, 57, 58, 144, 129, 42, 41, 205], la clasificación según formas y tamaños utilizando sensores tridimensionales láser de alta resolución [78, 79, 95, 61] o utilizando técnicas de visión [144, 22, 208, 196, 23, 137], la estimación del grado de madurez utilizando el color superficial de la fruta [144, 32], la clasificación basada simplemente en el color [147, 195, 4] y la detección de la presencia de pedúnculos como criterio de calidad o con el fin de cortarlos [176, 144, 217].

- *Maduración y conservación.* Un aspecto de gran importancia, especialmente en el caso de frutas y verduras, es el suministro del producto para el consumo en la mejor etapa de madurez, y a la vez la conservación de dichos productos durante largos periodos de tiempo a pesar de que sean productos estacionales y tienden a degradarse rápidamente. A parte de las soluciones tradicionales de conservación (mantenimiento en medios salinos, azucarados o en aceite), desde hace unos años es norma habitual la conservación al frío, los encerados para evitar deshidrataciones y los envasados al vacío de los productos previamente esterilizados [111]. Como mejora de las cámaras frigoríficas, se han introducido las *cámaras de atmósfera controlada* donde diversos parámetros que afectan al proceso de conservación de las frutas son ajustados, como por ejemplo la temperatura, humedad, luz, etileno, anhídrido carbónico y el oxígeno. La utilización de carretillas autoguiadas en grandes cámaras frigoríficas para el almacenamiento de los productos mediante sistemas paletizados, es una incorporación que también es ya una realidad. En este caso, no solo se evita el trabajo de personas en estas condiciones adversas, sino que se mejora la gestión de los productos al estar todo supervisado por un procesador central que controla existencias, ubicación y tiempos de almacenamiento de los productos.

- *Empaquetado y paletizado.* Esta etapa final suele consistir de varios subprocesos como son la dosificación, el envasado, etiquetado y paletizado. En el primer caso, las grandes piezas de fruta, hortalizas o carnes son cortadas en pequeños trozos para su posterior envasado. Una vez que estos productos sólidos están troceados se suelen cargar manualmente en bandejas, siendo dichas bandejas posteriormente pesadas y etiquetadas de forma automática. Es cada vez más común realizar posteriormente una inspección de la calidad del envasado detectando etiquetas erróneas, envases no correctamente cerrados o presencia de objetos extraños. Finalmente, la tendencia actual es la de colocar los envases conteniendo los productos en palets de tal forma que se agilice su almacenamiento, transporte y distribución.

Como hemos podido ver existe ya un cierto grado de automatización en los procesos agroalimentarios, sin embargo existen unas etapas que están más desarrolladas que otras. Con un alto nivel de automatización están los procesos de dosificación y envasado de líquidos, y la conservación y maduración en ambientes artificiales. A un nivel en proceso acelerado de implantación están los procesos de inspección y clasificación, de fertirrigación, el empaquetado y la paletización. Sin embargo existen otros procesos con

un bajo nivel de automatización; nos estamos refiriendo a ciertas tareas de recolección en campo, especialmente en la recogida de productos delicados donde se han desarrollado varios trabajos de investigación, pero que no han llegado a implantarse debido a los problemas técnicos encontrados al intentar dotar al robot de comportamientos totalmente autónomos.

1.1.3 Automatización en la recolección de frutas delicadas

La automatización de la recolección de frutos en plantas leñosas, es una labor compleja, habiéndose podido automatizar, y solo de una forma parcial, la recolección de frutos poco delicados como aceitunas, nueces y almendras utilizando métodos masivos e indiscriminados como vibradores de tronco o rama, peines, succionadores de aire o productos químicos para atacar el cáliz y facilitar el desprendimiento de la fruta [37]. Sin embargo la fruta delicada no puede ser recogida por estos métodos tan agresivos, especialmente si el consumo al que está destinada es fruta de mesa. Al caer la fruta del árbol se producen roces con la ramas y finalmente sufren daños al impactar con el suelo, esto hace que la calidad externa e interna del fruto se deteriore, impidiendo su comercialización. Adicionalmente, en algunas variedades la recolección se realiza cuando el fruto de la siguiente temporada está germinando, esto hace que en el proceso de recolección indiscriminado no solamente se desprendan los frutos sino también la flor, provocando reducciones de la producción de hasta el 25% [37]. Solamente es aceptable utilizar métodos agresivos en variedades delicadas en los casos en que el destino final vaya a ser la producción de zumos, concentrados, mermeladas o confituras [28]. Por tanto, la recolección manual actual de productos delicados para un consumo fresco, solamente puede ser sustituida por otro tipo de recolección individualizada: la *recolección robotizada.*

La utilización de robots en la recolección individualizada de productos delicados como el tomate, naranja, limón, melocotón y manzana, por citar las más representativas, trata de emular el trabajo realizado por una persona recolectora. Esto supone que el robot a diseñar debe ser capaz de realizar diversas tareas: desplazamiento y guiado por la plantación, detección y localización de la fruta, aproximación de un órgano de captura, agarre del fruto, desprendimiento del árbol y finalmente su depositado en un recipiente. Para dotar a un robot de estas habilidades es necesario utilizar tecnología ligada a sensores de localización y navegación, manipuladores y pinzas para el agarre utilizando accionamientos neumáticos o eléctricos, y sistemas de procesamiento para análisis de la información, control y toma de decisiones. Además, existen plantaciones en las que la producción no siempre madura al mismo tiempo, con lo cual la recolección, a parte de tenerse que hacer de forma individualizada, se debe realizar de forma selectiva recolectando solo aquellos frutos que están en estado idóneo para su consumo. Este hecho provoca que los sistemas sensoriales del robot deban ser numerosos y altamente especializados para captar las diferentes características según las cuales se debe guiar para realizar la toma de decisiones.

En la literatura podemos encontrar diversos robots o sistemas experimentales desarrollados con el objetivo de automatizar la recolección. Cabe citar los trabajos desarrollados en las universidades de Virginia [160], Florida [188, 189, 76], California [187] y en el centro italiano A.I.D. [130, 17] con diferentes investigaciones y desarrollos para

la recolección de manzanas, naranjas, tomates y melocotones. Otro tipo de recolecciones como la de melones [213, 27, 46, 10], sandías, uvas, calabazas, repollo, bayas, pepinos o champiñones también han sido investigadas [163, 191, 207, 148, 88]. El proyecto *Magali* [45, 170] se dedicó a la realización de un robot para la recolección de manzanas, que consta de un manipulador esférico accionado hidráulicamente y de un vehículo autopropulsado que permite el guiado automático en el campo utilizando cuatro sensores ultrasónicos. Igualmente, para la recolección de manzanas fue diseñado el robot *Aufo* [117] que utilizaba seis brazos con solo dos ejes horizontales cada uno, pudiendo realizar movimientos limitados a un plano vertical. Mediante desplazamientos angulares se barre todo el árbol y la fruta se detecta por triangulación utilizando visión esteroscópica. El proyecto hispano-francés *Citrus* [112, 165, 113] está dedicado a la recolección de naranjas, incluyendo un estudio agronómico, el desarrollo de un sistema de localización automático del fruto, y el diseño y control de un brazo que inicialmente utilizaba un sistema de coordenadas cilíndricas y que en la versión final se sustituyó por un sistema esférico similar al utilizado en el robot *Magali*.

Las principales dificultades encontradas en las soluciones presentadas para recolección automática están en el guiado del robot en campo, la detección automática de los frutos y el agarre/desprendimiento de cada fruta.

- *Guiado.* Las técnicas de autoguiado en labores agrícolas utilizan elementos sensoriales para dotar, normalmente a un tractor, de la capacidad de navegación autónoma. La información requerida para realizar esta tarea es la posición absoluta de la plataforma móvil, aunque en otros basta utilizar información local siguiendo algún surco o navegando a lo largo de los pasillos marcados por las hileras de los productos [71, 151, 16, 200, 59]. Las dificultades mayores se encuentran en las labores en campo, por la falta de uniformidad y por la dificultad que entraña el dotar de la suficiente flexibilidad al sistema para adaptarse a posibles situaciones imprevistas. En este sentido los trabajos de autoguiado en invernaderos son más sencillos ya que la estructuración es mucho mayor, las plantaciones son más homogéneas y las condiciones de iluminación son más fácilmente controlables. La colocación de las balizas es más simple debido a que el campo de actuación del vehículo está perfectamente definido. Además existe un especial interés en el autoguiado en invernaderos ya que las labores de fumigación y pulverización son más nocivas para una persona en este tipo de entornos cerrados [157, 24, 134]. En cuanto a las técnicas de autoguiado en campo, se suelen utilizar técnicas para el posicionamiento grosero del tractor (balizas fijas con emisor abordo [77], cámaras fijas y tractor posicionado por triangulación [156], GPS), y para el desplazamiento preciso se utiliza otro tipo de sensores que detecten el modo de navegación local como la orientación o la proximidad a la zona de operación (seguidores de surco ópticos y ultrasónicos [77], giróscopos, sensores de dirección geomagnéticos y detección por visión artificial de las hileras de la plantación [20, 200]).

- *Agarre y desprendimiento.* En este caso los sistemas utilizados para el corte de la fruta se pueden agrupar en dos clases: 1) los basados en la torsión del pedúnculo y 2) los que realizan un corte en la base del pedúnculo mediante cuchillas o sierras (fig. 1.3). De nuevo nos encontramos con el problema de evitar cualquier tipo de daño al fruto, para lo cual las pinzas deben ser lo menos invasivas posibles y el corte

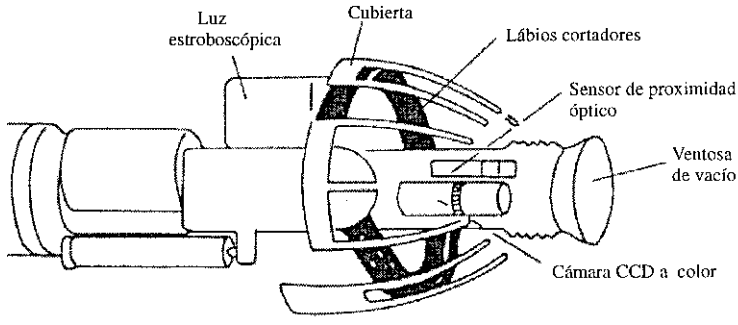


Figura 1.3: Pinza diseñada para el robot de recolección de naranjas CPR de la Universidad de Florida [188, 189, 76, 182].

producido se debe hacer próximo al cáliz, puesto que si queda muy largo se pueden producir daños entre las frutas en el almacenamiento, y a la vez hay que evitar la caída del cáliz para evitar posibles infecciones [145]. Los principales motivos que dificultan la recogida son la interferencia de hojas y ramas, y el diámetro variable del fruto que a veces es excesivamente grande o pequeño para el diseño particular de pinza. La efectividad en la recogida o relación de frutos útiles frente a los que se intentaron recoger está entre el 48% y el 64% para las pruebas realizadas en la captura de naranjas del proyecto *Citrus*, con un porcentaje de 10-15% de frutos caídos.

- *Detección.* La detección del fruto es otro aspecto muy crítico. Básicamente la mayoría de los sistemas ensayados para realizar la discriminación emplean la diferencia de color entre el fruto y fondo [160, 45, 170, 188, 189, 76, 187, 117, 112, 27, 46, 24], o en otros casos analizan la forma convexa en las imágenes de intensidades obtenidas utilizando cámaras CCD en color o en blanco y negro [213, 130, 165, 10]. Las mayores dificultades encontradas están relacionadas con la iluminación y las sombras en la escena, el hecho de que el color no sea siempre un factor discriminante, la oclusión de la fruta por parte de las hojas, ramas u otras frutas, y la presencia de objetos o fuentes luminosas visibles a través de los árboles. En cuanto a la oclusión de frutos, existe un estudio para el caso de naranjos que indica que únicamente el 40-50% de la fruta es visible desde el exterior y se recomienda la poda mecánica en formas cónicas para conseguir aumentar la fructificación exterior, pudiéndose facilitar de esta forma la detección automática de frutos al conseguirse visibilidades del orden del 75%. Igualmente la visibilidad puede ser aumentada mediante una dirección de recolección descendente proporcionando aumentos en visibilidad en torno al 7-8% frente a la dirección ascendente [111].

  Debido a que la detección de frutos en ambientes no estructurados como los agrícolas es el tema central del presente trabajo de tesis, vamos a dedicar la siguiente sección a analizar este problema más detenidamente realizando una amplia revisión de los trabajos que ya han sido propuestos en este campo.

1.2 Percepción automática en los procesos de recolección

1.2.1 Variabilidad del entorno agrícola

El entorno agrícola se caracteriza por una considerable variabilidad de sus productos, y en relación con el medio, de las condiciones meteorológicas y orográficas. Los productos agrícolas son cambiantes en forma, tamaño, color, textura y dureza, incluso perteneciendo al mismo tipo y a la misma variedad de plantación. En este sentido podemos encontrarnos árboles de diferentes tamaños y formas que contienen frutos en distintas etapas de maduración y por tanto en diferentes estados. Las características de los frutos también varían en función de las condiciones del terreno, de la densidad de árboles en la plantación, etc.. La densidad y distribución de frutos en el árbol depende de su orientación, obteniéndose mayor densidad de frutos en las caras del árbol orientadas al Sol [113]. A la variabilidad de los productos agrícolas se unen los problemas derivados de las condiciones meteorológicas adversas. La temperatura exterior puede variar en función del lugar de operación, de la época del año, de la hora y de la presencia o ausencia de nubes. Junto a la temperatura aparecen otras variables muy importantes como son la lluvia, la niebla, la humedad, el viento y el polvo.

Esta variabilidad del entorno afecta de una forma directa a las condiciones de visibilidad de los frutos en el árbol, que es el factor fundamental para un sistema de visión automático. La iluminación del árbol es un factor muy cambiante y que depende de condiciones externas difícilmente controlables, no siendo un problema perfectamente resuelto incluso utilizando luz artificial. La oclusión de los frutos por parte de otros frutos u hojas del árbol es, junto al problema de la iluminación, uno de los factores más problemáticos para la detección automática de objetos en entornos agrícolas. El factor de oclusión es tal que existe un cierto porcentaje de frutas que no son visibles ni parcialmente (50-60% [111]). La naturaleza no cerrada del árbol hace que se puedan ver objetos a través de las hojas del árbol lo cual añade más dificultades para la interpretación de las imágenes. El hecho de que los objetos puedan variar de tamaño y forma redunda en la necesidad de realizar sistemas de detección que sean suficientemente robustos como para adaptarse a estos cambios. Debido a que los frutos se pueden presentar con diferentes tonalidades de color, los métodos de detección no se deberían basar fundamentalmente en el color para realizar las clasificaciones.

Otros entornos más estructurados (p. ej. pruebas en laboratorio, líneas industriales de inspección y clasificación) reducen y simplifican el problema de reconocimiento al tener un mayor conocimiento acerca del proceso. Las fuentes de iluminación son controladas, el universo de objetos está más definido y se conocen incluso las zonas posibles de localización y orientación de los objetos. En el entorno agrícola existe un menor conocimiento de las condiciones en las que se puede presentar un objeto, y en consecuencia una dificultad en su modelado. Debido a que el reconocimiento consiste en comparar algo con un modelo, si no somos capaces de generar este modelo adecuadamente los resultados serán pobres. Por tanto, el proceso de reconocimiento en entornos no estructurados es más complejo y se necesita un diseño especial con el fin de dotar al sistema de una aceptable robustez.

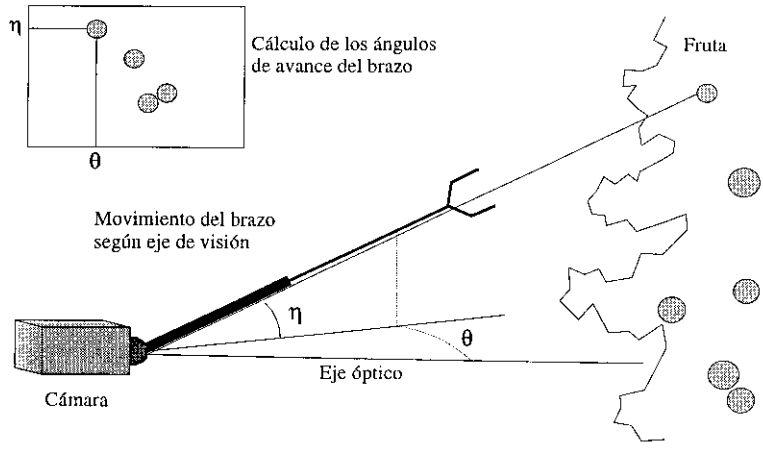


Figura 1.4: Principio básico de localización tridimensional del fruto: Primero se obtiene las coordenadas angulares en base a la posición bidimensional del fruto en la imagen, y finalmente, se mueve el brazo a lo largo del eje de visión del fruto, hasta que se produce un contacto.

1.2.2 Revisión de los principales métodos de detección de frutos propuestos para tareas de recolección agrícolas

Algunas de las tareas imprescindibles en el desarrollo de sistemas de recolección selectiva de frutos es el reconocimiento, localización y la determinación del tamaño y madurez de cada fruta de forma individualizada. Esta información es necesaria para poder guiar un brazo recolector hacia aquellos frutos considerados en condiciones óptimas de ser recolectados. En los siguientes apartados se hace una revisión de las diferentes soluciones aportadas para tratar de solucionar el problema planteado. Esta revisión se presenta organizada por grupos de trabajo y además está ordenada cronológicamente de tal forma que se pueda ver la evolución de la labor investigadora en este área. Aunque no se pretende ser exhaustivos, el estudio es suficientemente amplio como para proporcionar una visión bastante completa de las principales técnicas y sistemas desarrollados para la detección y localización de frutos en entornos agrícolas.

1.2.2.1 Propuesta de Schertz y Brown

La primera referencia que aparece en la literatura, que considera la detección y localización de frutos mediante técnicas automáticas, data del año 1968. En este trabajo Schertz y Brown sugirieron que la detección de frutas podría ser realizada mediante el uso de información fotométrica [184]. La técnica sugerida se basa en el uso de la diferencia de reflectividad luminosa entre las hojas del árbol y la superficie de los frutos, tanto en la zona visible del espectro electromagnético como en el infrarrojo cercano.

La sugerencia de Schertz y Brown es aplicada en la inspección agrícola utilizando la luz reflejada en una banda espectral centrada en 660 nm, pudiéndose distinguir entre

diferentes tipos de naranjas de tipo Valencia [56]. El objetivo de la clasificación era distinguir entre frutos de color naranja, naranja claro o de color verdoso, quedando dicho objetivo satisfecho en líneas de inspección bajo condiciones de iluminación estructurada.

1.2.2.2 Universidad de Virginia: Parrish y Goksel

El primer sistema de visión artificial para la detección y localización de manzanas en entornos agrícolas se desarrolla en la Universidad de Virginia [160]. Este sistema de visión no forma parte de un robot agrícola sino que es el componente principal de un trabajo de experimentación en laboratorio para analizar los principales problemas en la recolección automática agrícola. El modo de actuación del sistema de detección y localización está basado en el principio propuesto por Schertz y Brown [184] donde primeramente se detecta el fruto en base a una imagen de intensidades y posteriormente se guía al mecanismo de captura del fruto a lo largo del eje de visualización hasta que se produce un contacto, obteniéndose de esta forma la localización del fruto (fig. 1.4). En estos experimentos se utilizó un árbol artificial al cual se acoplaron frutos artificiales en diferentes posiciones para garantizar una distribución lo más próxima a la de un caso real. No se utilizó ningún sistema de captura del fruto por considerarse fuera del objetivo de la investigación. El sensor utilizado es una cámara blanco y negro que está posicionada de manera solidaria respecto a un brazo rudimentario de tres grados de libertad y coordenadas cilíndricas. La cámara tiene acoplada un filtro óptico de color rojo para resaltar los colores rojizos frente a los verdosos que corresponden a las hojas del árbol. De esta forma se obtiene una imagen de intensidades o niveles de gris la cual es procesada y analizada en tres etapas diferenciadas:

1. En la primera etapa se realiza una segmentación aplicando un umbral y se obtiene una imagen binaria donde los niveles de gris de la imagen superiores al umbral se representan por un "1" lógico e indican las zonas correspondientes a superficies de frutos. Los puntos en la imagen por debajo del umbral se representan por un "0" lógico indicando que pertenecen al fondo de la imagen, no teniendo interés y por tanto siendo ignorados.

2. Se realiza un suavizado de la imagen binaria aplicando operadores morfológicos sobre ella y de esta forma eliminando segmentos pequeños y aislados debidos a ruidos.

3. Finalmente, por cada uno de los segmentos, se calcula la diferencia de posición en el eje horizontal entre el punto más a la izquierda y el situado más a la derecha. Igualmente se calcula la diferencia entre los extremos verticales. La relación entre estos valores extremos en el eje vertical y horizontal dan una idea del tamaño y del grado de redondez del segmento. Esta relación debe superar un cierto umbral para que el segmento pueda ser considerado un candidato a posible fruto. De cada candidato a fruto se calcula la densidad de puntos con "1" lógico en una ventana circular de tamaño y posición indicados por los valores extremos. Si esta densidad supera un determinado umbral finalmente el segmento es clasificado como una manzana.

1.2.2.3 El proyecto MAGALI: D'Esnon y Rabatel

El proyecto francés MAGALI se propone la realización de un robot de recolección de manzanas [45]. La filosofía general de funcionamiento del proceso de localización y captura es la misma que sugirió Schertz y la que desarrolló Parrish, es decir, realizar el movimiento de aproximación al fruto siguiendo la línea recta marcada por el eje de visualización del fruto. El movimiento de aproximación finaliza al detectarse la presencia del fruto mediante un sensor de contacto. Existen dos fases de desarrollo diferenciadas dentro de este proyecto, tanto en el aspecto mecánico como en el desarrollo del sistema de visión artificial.

En la primera versión del sistema de visión se utiliza una cámara de color mediante la cual se captura una imagen parcial de la escena. La imagen en color obtenida es procesada mediante un sistema de procesamiento analógico de señal que selecciona los puntos que tienen un determinado color, obteniendo de esta forma una imagen segmentada. Un posterior procesamiento de la imagen binaria determina el centro de los segmentos bajo consideración. Pero esta solución aportada es poco robusta y necesita el uso de una cortina opaca colocada detrás de la escena para conseguir un fondo oscuro y de esta forma no cometer tantas detecciones erróneas.

En la segunda versión [170, 45] se utilizan tres cámaras de color y tres filtros ópticos centrados en las siguientes longitudes de onda: 950 nm (infrarrojo), 650 nm (rojo) y 550 (verde). La selección de las longitudes de onda de estos filtros se realizó en base a un estudio fotométrico del espectro en la banda de los 500 nm a los 1500 nm, aunque no se consideraron longitudes de onda superiores a 1100 nm debido a que caen fuera del ancho de banda de la cámara CCD. Se utilizaron tres tipos de manzanas cada una de ellas con diferente color: Golden (amarillo/verde), Starky (roja) y Granny Smith (verde). Se observó que en torno a 950 nm la reflectividad de las tres variedades de frutas estudiadas y la de las hojas era muy parecida (fig. 1.5 y 1.6). De esta forma se eligió la imagen filtrada a 950 nm como referencia para que el proceso de reconocimiento fuese lo más insensible posible a las variaciones de luminosidad. La banda infrarroja que va desde los 750 nm a los 900 nm se caracterizaba por permitir distinguir fácilmente cualquier tipo de fruta de las hojas debido a que la reflectividad de la fruta es mayor que la de las hojas en esta zona espectral. En la banda visible de 500 a 700 nm la fruta tiene una reflectividad mayor en promedio, sin embargo algunos tipos de fruta se confunden con algunas hojas. Los autores eligieron las frecuencias de los filtros interferenciales en la zona visible (650 y 550 nm), aspecto que parece poco comprensible debido a la existencia de la banda de 750 a 900 que permitiría distinguir fácilmente entre las dos clases (fruta y hoja). Posteriormente se calculaba el coeficiente de relación entre las imágenes de 650 y 550 nm respecto a la de 950 nm. Este cómputo se realizaba mediante un sistema electrónico analógico obteniéndose una imagen segmentada binaria. Un análisis posterior calcula la posición del centro geométrico de los segmentos en la imagen binaria. Utilizando una técnica similar a esta segunda versión del MAGALI, Kawamura también se ha estudiado la detección y recolección de tomates [119].

El sistema de visión desarrollado para el MAGALI es capaz de reconocer incluso manzanas de color verde, pero la tasa de aciertos es baja (aprox. 50%) y además los autores reconocen que se producen bastantes detecciones falsas, aunque no lo cuantifican, y que es sensible a las variaciones de iluminación. Así mismo el sistema no es flexible

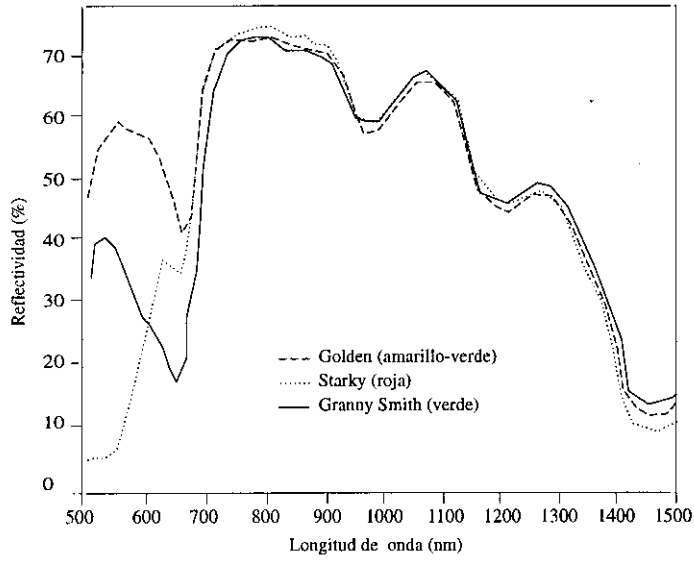


Figura 1.5: Reflectividad de varios tipos de manzanas (datos del CEMAGREF [170]).

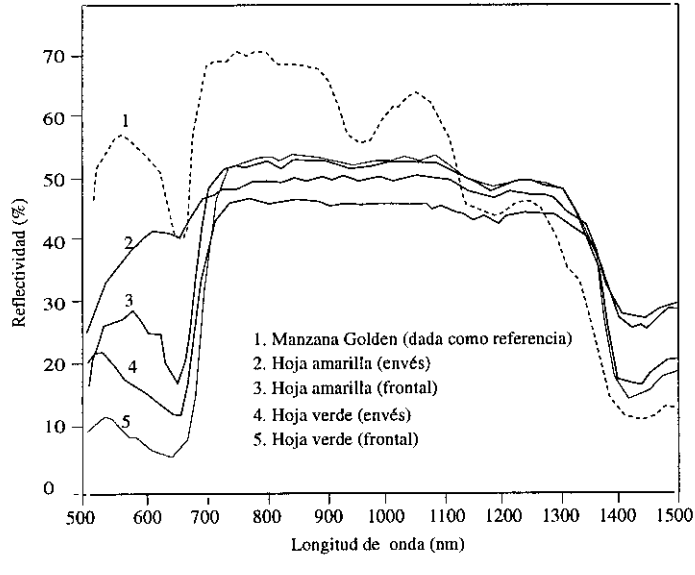


Figura 1.6: Reflectividad de las hojas de un manzano y de una variedad de manzana (datos del CEMAGREF [170]).

puesto que si se quiere utilizar para recolectar otro tipo de frutos se necesitaría realizar otro estudio espectral para poder extraer de nuevo las frecuencias óptimas de los filtros ópticos interferenciales.

1.2.2.4 Universidad de Florida y centro USDA: Harrell y Slaughter

Con el fin de dotar al robot recolector de naranjas CPR (Citrus Picking Robot) de la capacidad de reconocimiento automático de frutos se realizan varias labores de investigación mediante una colaboración entre la Universidad de Florida (Harrell) y el centro de investigación biológica USDA (Slaughter). Este robot tiene tres grados de libertad utilizando un sistema de coordenadas cilíndrico y por tanto realizando la aproximación al fruto de forma horizontal. La cámara de visión utilizada, así como otros dispositivos y sensores, van incorporados en el interior del brazo en las proximidades del órgano de captura (fig. 1.3), de esta forma la localización se hace centrando el fruto detectado respecto a la cámara y realizando un movimiento de aproximación horizontal a este, que finaliza cuando se detecta su presencia mediante un sensor ultrasónico.

Existen dos enfoques para la solución del problema de reconocimiento, aunque ambos están basados en el uso de una cámara de color. En el primer enfoque [188], se emplea una cámara de color con control de apertura e iluminación artificial pero sin ayuda de ningún filtro óptico. Las componentes de saturación y tono de color (hue) de cada uno de los puntos de la imagen se utilizan para realizar la segmentación mediante una clasificación en un espacio bidimensional de características. La clasificación se realiza mediante el uso de un clasificador lineal que actúa aislando una región rectangular en el espacio de características mediante el uso de un valor umbral máximo y mínimo para cada una de estas características. Mediante esta aproximación se consigue clasificar correctamente un 75% de los puntos, siendo el sistema muy sensible a las condiciones de iluminación y no siendo capaz de reconocer naranjas que no estén maduras. Se sugiere la ejecución del algoritmo de clasificación mediante una etapa electrónica analógica que aplique automáticamente los umbrales mediante el uso de comparadores; de esta forma se podría reducir el tiempo de procesamiento de 2.5 segundos por imagen (para un procesador Motorola 68020 a 12.5 MHz e imágenes de 384*485 puntos) a una cifra más reducida.

En el segundo trabajo presentado se utiliza básicamente la misma configuración anterior aunque ya no se utiliza luz artificial [189]. En este caso se trabaja con las componentes RGB (Rojo, Verde y Azul) de cada punto de la imagen obteniéndose un espacio tridimensional de características. De esta forma cada punto es clasificado como perteneciente a una fruta o al fondo mediante un clasificador bayesiano, para lo cual necesita de un proceso previo de aprendizaje sobre un conjunto de muestras de frutas. Los resultados presentados muestran que un 75% de los puntos pertenecientes a naranjas son clasificados correctamente (fig. 1.7). Este método sigue siendo solamente válido para naranjas maduras pues es la única forma de que el clasificador tenga capacidad de discriminación. Finalmente, como complemento al sistema activo de localización, se elaboraron algoritmos para la estimación del centro de los segmentos detectados, así como de los diámetros horizontales y verticales [76]. Este algoritmo consta de dos pasos, estando orientado el primero a la búsqueda de la región a analizar mediante un rastreo iterativo por la imagen en forma de espiral en saltos de 25 en 25 puntos. Una vez

Figura 1.7: (a) fotografía de un naranjo donde aparecen tres naranjas maduras, y de fondo, hojas y cielo. (b) segmentación basada en color de la imagen anterior [189].

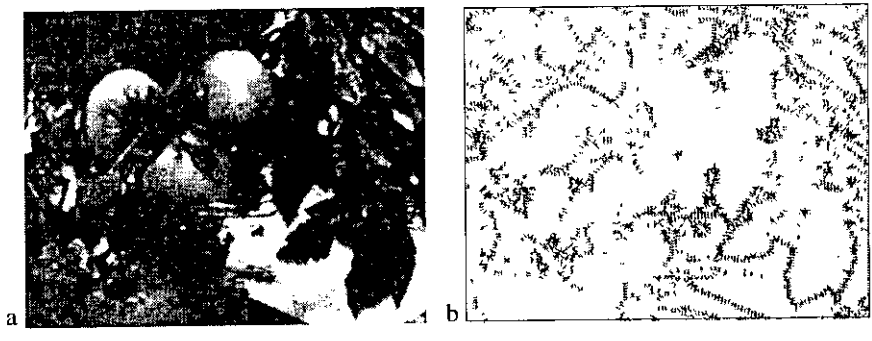


Figura 1.8: (a) imagen en B/N de una plantación de tomates. (b) bordes y direcciones del gradiente correspondientes a la imagen anterior [213].

detectado el segmento, entra en acción la etapa de cómputo del centro del segmento y de los diámetros horizontales y verticales que se basa en un proceso iterativo de trazo de líneas horizontales y verticales interiores al segmento hasta que el cruce entre ambas líneas se estabiliza convergiendo el algoritmo. El criterio utilizado para posicionar cada una de las líneas se basa en la posición del centro de la línea previamente trazada.

1.2.2.5 Universidad de Purdue: Whittaker

Whittaker presenta un sistema para reconocer y localizar tomates insensible al grado de madurez del fruto, y por tanto, independiente del color del tomate [213]. Para poder llevar a cabo sus objetivos considera que los sistemas de visión que analizan la imagen de forma local utilizando solo los niveles de intensidad sin tener en cuenta la distribución de los puntos en la imagen, no son adecuados para realizar un método de reconocimiento de objetos que sea independiente del color. Por tanto propone utilizar la forma de los contornos presentes en una imagen de intensidades para obtener indicios de posibles objetos esféricos.

Se utiliza una cámara blanco/negro que genera 256 niveles de gris. La imagen de grises obtenida es procesada aplicando un filtro de Sobel con lo cual se obtiene el vector gradiente correspondiente a cada punto de la imagen. Posteriormente se binariza la imagen teniendo en cuenta que la magnitud del vector debe superar un cierto umbral, siendo la elección de este umbral muy influyente en la calidad de la imagen resultante. La imagen obtenida contiene información sobre la localización de los contornos o bordes presentes en la imagen y también la dirección de dichos contornos (fig. 1.8). De esta forma se aplica la transformada circular de Hough en su versión optimizada en el número de votaciones al utilizar la dirección de los bordes. Para realizar la umbralización en la matriz de acumuladores se utiliza la distribución del histograma de esta matriz. Se consideró que la forma del histograma obtenido se ajustaba con precisión a una distribución exponencial con lo cual se presenta un método de selección del umbral basado en el porcentaje de área bajo la curva de la distribución exponencial y no se basa directamente en la selección aleatoria de una cantidad de votos como valor umbral. A pesar de la utilización de este método sigue haciendo falta una elección manual del umbral pero ahora en términos de porcentajes. Los valores de umbral ensayados son 97.5%, 99% y 99.9%, pero los mejores resultados observados se obtienen usando un valor del 99% el cual conduce a unos resultados de un 68% de detecciones correctas y un 42% en detecciones falsas.

El contorno semiesférico de las hojas (fig. 1.8b), las sombras, las zonas brillantes e incluso la curvatura de los rabos de los tomates son los mayores problemas que hacen que el sistema produzca muchas detecciones falsas. Así mismo los autores manifiestan que el algoritmo es costoso computacionalmente y no se podía aplicar en tiempo real en un procesador secuencial de los existentes en aquella época (8086 o equivalente). En cuanto a la localización de los tomates, esta no es total puesto que la dimensión de profundidad no es calculada y por tanto es necesario realizar un movimiento de aproximación del brazo a lo largo del eje de visión del fruto.

1.2.2.6 A.I.D: Levi, Falla y Pappalardo

En el centro italiano A.I.D. se realizó un investigación para comprobar la viabilidad de la recolección automática de naranjas [130]. En este sentido se desarrollo un robot muy sencillo de coordenadas cilíndricas con tres grados de libertad y con un dispositivo de captura de fruta. El sistema de visión utilizado emplea una cámara de color y se ayuda de iluminación artificial. El reconocimiento de los frutos se realiza analizando la distribución de los gradientes en la imagen, aspecto que caracteriza a este sistema (junto al trabajo anterior de Whittaker) en ser pionero en basar el reconocimiento de frutas en el análisis de la forma. Como en los anteriores trabajos, el sistema de visión determina la coordenada de profundidad por medio del movimiento del órgano terminal del brazo a lo largo del eje de visión al fruto. El algoritmo de procesamiento de la imagen en color se puede dividir en tres etapas claramente diferenciadas:

1. Preprocesamiento. Un filtro electrónico analógico es utilizado para realizar un aumento de contraste de la imagen. Posteriormente se digitaliza la imagen obteniendo una imagen de $512 \times 512 \times 6$ bits, donde los 6 bits de cada punto de la imagen codifican la proximidad en términos del tono de color (hue) a un valor

de referencia predeterminado que está próximo al color naranja. De esta manera obtenemos una imagen en tonos de grises donde se resaltan las zonas de interés.

2. Extracción de características mediante el cálculo del gradiente. Se aplica un filtro de Sobel obteniendo un vector indicando el gradiente correspondiente a cada punto de la imagen.

3. Reconocimiento. El reconocimiento se realiza mediante el uso de un modelo de un objeto esférico previamente almacenado. Para formar el modelo se crea un círculo sintético y a partir de él se calcula el gradiente obteniéndose un mapa de vectores. Para realizar el reconocimiento se mueve el modelo por toda la imagen paso a paso hasta obtener una coincidencia con una zona de la imagen.

Los resultados obtenidos son aceptables cuando se utiliza una luz artificial de 3200 K de temperatura de color, de esta forma, en torno a un 70% de las naranjas visibles son reconocidas aunque los resultados se degradan cuando los frutos tienden a estar verdes. Cuando se trabaja de noche con luz artificial los resultados son algo mejores debido a que es más fácil obtener la iluminación deseada.

Cabe reseñar que un método muy parecido al utilizado por Levi, pero realizando la segmentación por crecimiento de regiones y búsqueda de esferas, fue utilizado por Cox y otros para buscar objetos esféricos en fotografías aéreas [39].

1.2.2.7 Centro Sunkist y Universidad de California: Sites y Delwiche

Una investigación sobre la aplicación de la visión artificial en la detección de manzanas y melocotones en estado de madurez se llevó a cabo mediante la colaboración entre el centro canadiense SRC (Sunkist Research Center) y el Departamento de Ingeniería Agrícola de la Universidad de California [187]. En este caso el método está basado en los niveles de intensidad que se obtienen con una cámara B/N, un filtro óptico (630 a 670 nm) para incrementar el contraste entre las frutas y el fondo, y una iluminación artificial para trabajar tanto de noche como de día. El método de procesamiento se divide en cinco etapas:

1. Binarización. Basándose en la distribución del histograma de la imagen encontrada se determina elegir como valor umbral aquel que asigna un "0" lógico al 37% de los puntos y un "1" lógico al resto.

2. Mejora. Utilizando un filtro morfológico, que actúa sobre los ocho vecinos de cada punto, se realiza una erosión de la imagen.

3. Segmentación. Se realiza un etiquetado de cada uno de los puntos de la imagen basándose en criterios de conectividad ocho.

4. Extracción de características. Las características utilizadas y calculadas en cada uno de los segmentos obtenidos en la etapa anterior son: área, perímetro, compacidad, alargamiento y momentos invariantes.

5. Clasificación. Se utiliza un clasificador lineal.

Se realizaron una serie de pruebas y los resultados en cuanto a detecciones correctas fueron del 89% trabajando de noche para evitar problemas con el cielo y el Sol. En las pruebas realizadas trabajando de día, los resultados son del 84% y 20% en detecciones correctas y falsas, respectivamente. El sistema fue diseñado para detectar fruta madura, pero aún bajo esta restricción aparecen muchos problemas al trabajar de día debido al cielo, a las nubes cambiantes, a la luz solar frontal incidiendo sobre la cámara y a que las hojas directamente iluminadas por el Sol pueden aparecer más luminosas que una fruta con iluminación solar difusa a pesar de estar utilizando siempre la luz directa del sistema de iluminación artificial. Esto hace que de día se produzcan bastantes detecciones erróneas y se sugiera el uso de cubiertas para eliminar los fondos en la imagen con regiones de cielo o de Sol.

1.2.2.8 El proyecto AUFO: Kassay

Entre los años 1980 y 1989 se desarrolló, en el "Central Enterprise for the Organization of Agriculture and Food Industry" de Budapest, el robot AUFO-06 para la recolección de manzanas [117]. El robot está diseñado para operar con seis brazos, de dos grados de libertad cada uno, permitiendo un movimiento de trabajo en un plano vertical. El prototipo utilizado para realizar pruebas experimentales en 1991 constaba de un único brazo, una plataforma motorizada, el sistema hidráulico de movimiento, el órgano terminal de captura, la electrónica de control y el sistema de visión para el reconocimiento y localización de los frutos.

El sistema de visión consta de dos cámaras de color situadas a una cierta distancia y colocadas entre sí de tal forma que exista una cierta convergencia entre ellas para poder capturar imágenes de la misma escena y poder obtener la posición de los frutos mediante una técnica sencilla de estereo-visión. Primeramente se realiza una segmentación por umbralización en cada una de las dos imágenes de color, obteniéndose las imágenes binarias que contienen las regiones que corresponden a la superficie de las manzanas. Se calcula el centro geométrico de estos segmentos en ambas imágenes y a continuación por cada una de las parejas de segmentos se calcula la posición tridimensional. El método de cálculo de la posición 3-D se basa en el principio de triangulación, calculándose en primer lugar la posición de la posible fruta en las coordenadas X e Y mediante la proyección de los ejes de visión sobre el plano horizontal que contiene el eje óptico de ambas cámaras. A continuación se calculan las coordenadas z_1 y z_2, o alturas del fruto respecto a cada cámara. Para ello aplica un criterio de coincidencia calculando el ángulo en el plano vertical formado entre el eje de visión y el plano horizontal. Si ambas alturas difieren en menos de 40 mm entonces el fruto es considerado válido.

El método descrito presenta varios problemas, el primero de ellos aparece al considerar todas las posibles combinaciones de parejas entre segmentos de ambas imágenes, lo cual supone una búsqueda exhaustiva y además puede llevar a situaciones en las cuales aparecen posiciones virtuales de frutos que en realidad no existen. Este problema es compensado en parte por la comprobación descrita anteriormente entre las dos alturas obtenidas a partir de cada cámara, pero esta comprobación no sirve para solucionar todos las situaciones erróneas posibles. El hecho de trabajar con el método de estéreo visión en entornos donde hay tantas oclusiones y contrastes de iluminación, hace que no se encuentren siempre las correctas parejas y por tanto no se produzca la detección. Como

consecuencia el porcentaje de frutos detectados frente a los visibles por un operador humano es de solo un 41%. Los algoritmos de visión y de control de los movimientos de el brazo se ejecutaron sobre un Intel 8085, obteniéndose tiempos medios de procesamiento para el sistema de detección y de localización ligeramente superiores a 1 segundo, lo cual es aceptable teniendo en cuenta que el ciclo de trabajo del robot es de 14 segundos.

1.2.2.9 El proyecto CITRUS: Juste, Sevila, Plá, Moltó

El proyecto europeo CITRUS se centró en la realización de un estudio profundo de los diferentes aspectos que entran en juego en las tareas de recolección de cítricos y por tanto incluye un estudio agronómico encaminado al análisis del entorno de trabajo, de las variedades de fruta y de su manipulación [114]. Así mismo, también consideró el desarrollo de un sistema de visión, el diseño y control de un robot manipulador junto con los elementos de captura del fruto, y finalmente, las pruebas de campo de los prototipos realizados. En este proyecto Eureka participa el IVIA (Instituto Valenciano de Investigaciones Agrarias) y el CEMAGREF en Francia. En cuanto al sistema de visión desarrollado existen varias metodologías o soluciones aportadas, en los siguientes párrafos se describe cada una de ellas.

En el primer método se presenta un sistema que trabaja con una cámara en blanco y negro, utiliza un filtro rojo (630 nm) y dos focos de iluminación con el objetivo de obtener una iluminación uniforme y lo más independiente posible de las condiciones ambientales [112]. Se utiliza un algoritmo que aplica un umbral obteniéndose un 80% de aciertos en la detección de la fruta visible y a su vez un número muy alto de detecciones falsas.

El segundo método surge para mejorar los problemas presentados anteriormente. En este caso se utiliza una segunda cámara en blanco/negro con un filtro verde (560 nm) [112]. A partir de las dos imágenes obtenidas se calcula el coeficiente entre ellas para generar una imagen en niveles de gris que es independiente de los niveles de iluminación. Con esta imagen el algoritmo de binarización actúa de una forma más robusta y los resultados obtenidos son de aproximadamente un 80% de detecciones correctas y un 10% de detecciones falsas.

En la tercera aproximación se utiliza una cámara en color sin utilización de iluminación artificial [112]. De cada punto se consideran las componentes RGB y se realiza una clasificación mediante una función de decisión bayesiana en un método similar al usado por Slaughter y Harrell para el robot CPR de Florida [189, 76]. En este caso los resultados son del 90% y del 5% en detecciones correctas y falsas respectivamente.

Los tres métodos presentados anteriormente no aportan soluciones totalmente satisfactorias puesto que no son capaces de trabajar con fruta no madura y por tanto requieren frutos de color rojizo o naranja para poder realizar una discriminación correcta (fig. 1.9). En este sentido existe un trabajo posterior de Plá que trata de solventar los problemas indicados [165]. Se utiliza una cámara en blanco y negro e iluminación artificial para obtener una imagen de intensidades (fig. 1.10) que será analizada en una primera etapa en busca de superficies convexas mediante la convolución de la imagen con la Laplaciana del Gausiano. De esta forma obtenemos otra imagen que da una idea del grado de convexidad presente en un vecindario de cada punto de la imagen. Esta imagen es umbralizada haciendo una segmentación y quedándonos solo con segmentos que se caracterizan por ser convexos. La segunda etapa consiste en realizar un ajuste

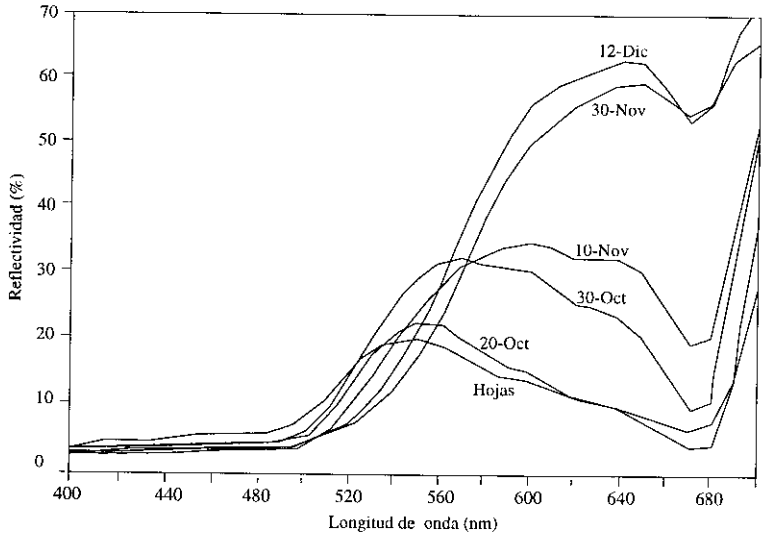


Figura 1.9: Variación de la reflectividad de las naranjas en función de la longitud de onda y de la etapa de desarrollo de la fruta (estudio realizado por Juste y Sevila [112]).

por mínimos cuadrados de una elipse sobre los puntos que superaron el umbral. Se usa una elipse y no una esfera debido al modelo de iluminación utilizado, y este ajuste solo se realiza en las direcciones horizontal y vertical de la imagen. El objetivo final de esta segunda etapa consiste en obtener un mayor nivel de certeza sobre la presencia de un objeto esférico. En función de si el ajuste fue satisfactorio o no, se reactualiza la imagen umbralizada de grados de convexidad sumando o no un cierto peso predefinido. Esta imagen de convexidades, finalmente, determina qué regiones corresponden a las frutas. Los resultados obtenidos mediante esta técnica son de un 75% en cuanto a detecciones correctas y un 8% en detecciones falsas. El método funciona tanto para naranjas maduras como verdes, pero sigue presentando falsas detecciones debidas a las zonas de cielo y de suelo presentes en la imagen (fig. 1.10). Utilizando un PC-386 el tiempo medio de procesamiento por cada fruta era de 20 segundos y el procesamiento por cada imagen de 512*512 puntos es de 3 a 4 minutos.

1.2.2.10 Univ. de Purdue y C. Volcani: Cardenas, Dobrousin, Benady, Miles

Un sistema de visión para la recolección de melones ha sido investigado bajo una estrecha colaboración entre la universidad de Purdue, EEUU, y el centro israelí Volcani. El caso de la detección de melones es muy similar a la detección de otro tipo de frutas que nacen en los árboles como son las naranjas, manzanas, etc.. Pero existen unas características peculiares que hacen de este tipo de recolección una tarea sensiblemente menos complicada que las anteriores. El hecho de que todas las frutas yazcan sobre el suelo hace que se restrinja significativamente el espacio de búsqueda y en cierta medida

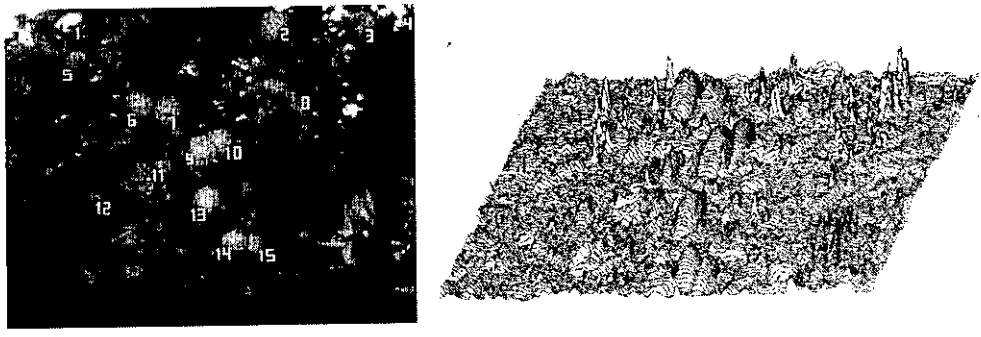


Figura 1.10: Imagen B/N de un naranjo y su representación tridimensional. Las frutas y las áreas de cielo se presentan mediante superficies convexas [165].

se estructura el entorno; de esta forma es posible aplicar ciertas heurísticas referentes a la ubicación espacial del melón cuando se dispone de información de distancia. Así mismo, debido al peso del melón y a que este está apoyado sobre el suelo, se pueden aplicar técnicas para mover las hojas que ocultan los frutos sin mover estos. De esta forma se puede reducir el factor de oclusión presente en este tipo de plantaciones.

En una primera aproximación a la resolución del problema de detección [27], se utiliza una cámara en blanco y negro para obtener una imagen de intensidades de la plantación de melones. La técnica de detección comprende dos etapas. La primera etapa identifica al melón, su posición bidimensional y su tamaño. Para hacer esto se realiza una mejora de la imagen, se aplica un umbral para segmentar la imagen, seguido de una extracción de características, una generación de hipótesis de posibles localizaciones de melón y la posterior verificación de esas hipótesis analizando la forma y la textura en determinados vecindarios entorno a la posición bajo hipótesis. De esta forma se obtienen finalmente los candidatos definitivos. La segunda etapa realiza una evaluación basada en el conocimiento del entorno, en el tamaño de los melones y en las restricciones que se deben cumplir entre los candidatos para que no haya múltiples ocurrencias y detecciones falsas. Estas reglas asignan los candidatos de la primera etapa a una de las siguientes clases: melón, ruido o redundante. Cuando no se usa esta segunda etapa un 89% de los melones son detectados pero se produce una enorme cantidad de falsas detecciones. Usando ambas etapas los resultados en cuanto a detecciones correcta y falsas son de un 84% y un 10% respectivamente, con lo cual se eliminan muchas falsas detecciones a costa de perder alguna localización correcta.

Otro sistema de visión para la recolección de melones es presentado por Dobrousin (fig. 1.11) y se caracteriza porque se divide en dos subsistemas: subsistema de visión lejana y de visión cercana [46]. El subsistema de visión lejana usa una cámara en blanco y negro para localizar las coordenadas X-Y del melón. Varias imágenes de la misma escena son capturadas pero con diferentes configuraciones en la posición de las hojas que cubren los melones. Esto se consigue mediante el uso de un ventilador de aire que mueve las hojas de la plantación, siendo su objetivo la eliminación de la oclusión de los melones por parte de las hojas. Todas las imágenes son filtradas, segmentadas aplicando umbrales

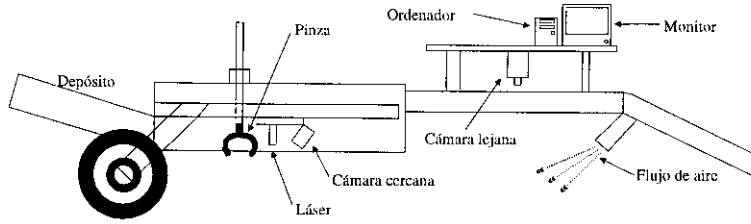


Figura 1.11: Esquema del remolque de tractor utilizado para la recolección de melones. Se puede apreciar el sistema de visión lejano, el cercano, la ubicación de la pinza y del sistema de soplado para mover las hojas de la plantación.

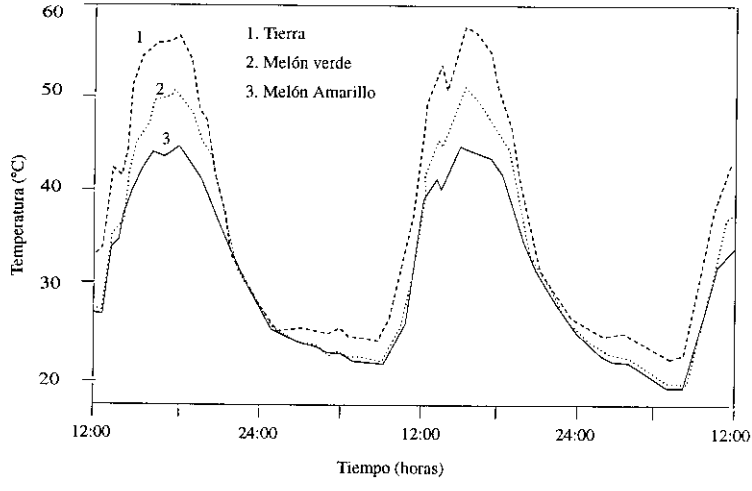


Figura 1.12: Variación de la temperatura en una plantación de melones en función de las horas del día [46].

elegidos analizando sus histogramas y finalmente se aplica un operador morfológico de erosión. Las imágenes binarias obtenidas son integradas realizando una operación lógica "O". Posteriormente, sobre la única imagen binaria resultante, se realiza una extracción de características incluyendo forma, área y tamaño. Finalmente una clasificación basada en reglas es utilizada para obtener las frutas válidas, aproximadamente un 80% de los melones son detectados. Estas rutinas fueron integradas en un sistema de procesamiento en cascada (pipeline) en tiempo real. Los principales problemas encontrados se deben a la sensibilidad en la selección del umbral de segmentación, a la iluminación variable, a las sombras y a las hojas presentes. Los autores proponen el uso de imágenes infrarrojas para detectar la diferencia de temperatura que debería existir entre las hojas, el suelo y los melones, presentando unos resultados en los que se muestra que por la tarde la temperatura de los melones es inferior a la de la tierra visible a través de las hojas (fig. 1.12).

El subsistema de visión cercano es presentado en otra publicación por Benady y Miles. Pretende obtener la altura Z a la que se encuentra el centro del melón, que es el único parámetro que no calculó el subsistema lejano, y refinar las posiciones en el plano X,Y [10]. En este sistema se usa una proyección lineal de luz láser para iluminar la escena. Este plano de luz cuando interactua con la superficie del melón es visualizado como una línea curva al ser registrado por una cámara que forma un determinado ángulo con el plano de luz láser. La deformación de la línea recta inicial indica la altura del objeto usando una técnica de triangulación. Estos perfiles lineales deformados son capturados a intervalos espaciales regulares a medida que avanza el sistema y son analizados utilizando la transformada circular de Hough para obtener la matriz de votos indicando los candidatos a ser el centro de un melón. Para obtener los candidatos más probables, la distribución de votos alrededor de cada punto es utilizada en vez de un valor absoluto de votos. Para incrementar la eficiencia del algoritmo se utilizan algunas reglas específicas del dominio. El conocimiento incluido en el sistema utiliza el tamaño, la forma del fruto y la posición esperada del suelo. Estas son las reglas utilizadas:

- *Altura absoluta.* Sabiendo donde está el suelo y el radio de los melones, podemos saber de antemano si un punto de la imagen pertenece al suelo o a las hojas en función de su altura absoluta.

- *Pertenencia.* Los puntos que emitieron voto a un fruto, mediante la transformada de Hough, pertenecen a la superficie del fruto.

- *Convexidad.* El centro del melón siempre está debajo de la superficie visible.

- *Apoyo.* El melón debe estar siempre apoyado sobre el terreno.

- *Solapamiento.* Dos frutos no pueden ocupar el mismo volumen simultáneamente.

- *Interioridad.* No deben existir puntos dentro del volumen de un melón.

Mediante este sistema todas las frutas que eran visualmente discernibles fueron detectadas correctamente y no se produjeron detecciones falsas.

1.2.2.11 CIRAA: Buemi

El robot italiano Agrobot se desarrolló en el CIRAA (Centro Interuniversitario de Robótica Agrícola y Ambiental). Está diseñado para realizar operaciones en invernadero, siendo la recolección una de las tareas para la que está diseñado [24]. El sistema de visión usado en este proyecto está basado en una cámara de color la cual suministra las componentes HSI de cada punto en la imagen. El tono de color y la saturación son utilizados para segmentar la imagen mediante umbrales. La información tridimensional es obtenida mediante visión esteroscópica al usar dos vistas de la misma imagen. Alrededor del 90% de los tomates maduros fueron detectados y los errores más frecuentes fueron debidos a oclusiones.

1.2.3 Análisis de las soluciones propuestas: Limitaciones

Los sistemas de detección automáticos analizados anteriormente comprenden normalmente tres fases. Una primera fase de *captación* de la imagen del entorno a reconocer, una segunda para la *mejora* de la imagen, y finalmente una tercera fase dedicada al *análisis* de la información para realizar el reconocimiento y obtener información de alto nivel. Las soluciones aportadas para el reconocimiento y localización de frutos presentadas en la sección anterior se centran en la primera y tercera fase, dejando la etapa de restauración en un apartado secundario. En la tabla 1.1 se resumen las diferentes soluciones que se han presentado. En ella se indican los sensores utilizados para la captación, el método de análisis empleado, la capacidad para la detección de fruta no madura, y los porcentajes de clasificación correcta y errónea.

Podemos ver que las metodologías de captación, se centran en la adquisición de la intensidad y del color de la escena. Solamente en un trabajo [10] se adquiere otro tipo de magnitud: la distancia. En cuanto a las técnicas de reconocimiento utilizadas, la mayoría utiliza estrategias basadas en propiedades locales a cada punto de la imagen (niveles de gris o color) y en el resto de los casos se utilizan propiedades correspondientes a vecindarios o pequeñas regiones de la imagen, caracterizando dichas regiones fundamentalmente por su forma. Los resultados obtenidos los valoramos en términos de los porcentajes de acierto/error y de la capacidad de detectar fruta en diferentes estados de madurez. En la mayoría de los casos solo es posible trabajar con fruta madura, los porcentajes de clasificación correcta no superan el 90% y es una regla general la aparición de detecciones falsas.

1.2.3.1 Captación de la imagen

En los trabajos revisados, el proceso de captación genera tres tipos de imágenes, cada una de las cuales representa diferentes propiedades del entorno: *intensidad de luz reflejada*, *respuesta espectral* y *distancia*. La captación de la intensidad, se realiza utilizando cámaras B/N con o sin iluminación artificial. La captación de características espectrales en determinadas bandas del espectro, se obtienen empleando cámaras en color que suministran las componentes RGB o HSI, y también, utilizando cámaras en B/N con filtros ópticos centrados en determinadas frecuencias espectrales. Ambos tipos de imágenes se caracterizan por:

- *Oclusión.* La oclusión de los frutos por parte de las hojas y otros frutos.
- *Sombras.* Las sombras que surgen por la iluminación solar y el alto contraste presente no se ven completamente solucionados mediante el uso de iluminación artificial. La fruta al Sol es diez veces más brillante que las hojas a la sombra y las hojas al Sol son cuatro veces más brillantes que una fruta a la sombra [184] y este problema no se soluciona completamente con iluminación artificial [187].
- *No-profundidad.* La información de profundidades no es conocida, lo cual es necesario para poder realizar la localización del fruto. La falta de esta información hace necesaria la utilización de la técnica de aproximación del órgano de captura a lo largo de la línea de visión del fruto. También se han utilizado técnicas estereoscópicas siempre muy comprometidas en este tipo de entornos.

Grupo investigación	Referencias	Tipo de fruta[1]	Sensores y accesorios[2] (tipo imagen)	Método de análisis[3]	Detecta fruta verde	Detecciones correctas-erróneas[4]
U. Virginia	[160]	Manz	B/N+F (Espectral)	Puntual (Umbr+ExtC+ClaR)	No	N.D.
MAGALI	[45]	Manz	Color (Espectral)	Puntual (Umbr)	No	N.D.
	[45, 170]	Manz	3 Color+ 3 F (Espectral)	Puntual (Razón+Umbr)	Sí	50%-alto%
U. Florida y USDA	[188]	Narj	Color+ L (Espectral)	Puntual (Hue&Sat+ClaL)	No	75%-N.D.
	[189, 76]	Narj	Color (Espectral)	Puntual (RGB+ ClaB)	No	75%-N.D.
U. Purdue	[213]	Toma	B/N (Intensidad)	Forma (Contorno+CHT)	Sí	68%-42%
A.I.D.	[130]	Narj	Color+ F+ L (Espectral)	Forma (Gradiente+Compar.)	No	70%-N.D.
Sunkist y U. Calif.	[187]	Manz y Melc	B/N+ F+ L (Espectral)	Puntual (Umbr+ExtC+ClaL)	No	84%-20%
AUFO	[117]	Manz	2 Color (Espectral)	Puntual (Umbr+Estereo)	No	41%-N.D.
CITRUS	[112]	Narj	B/N+ F+ 2L (Espectral)	Puntual (Umbr)	No	80%-alto%
	[112]	Narj	2 B/N+ 2F+ 2L (Espectral)	Puntual (Razón+Umbr)	No	80%-10%
	[112]	Narj	Color (Espectral)	Puntual (RGB+ ClaB)	No	90%-5%
	[165]	Narj	B/N+ L (Intensidad)	Forma (Convx+ Umbr&Ajuste)	Sí	75%-8%
U. Purdue y Volcani	[27]	Meln	B/N (Intensidad)	Puntual (Umbr+ ExtC+ ClaR)	No	84%-10%
	[46]	Meln	B/N+ Aire (Intensidad)	Puntual (Umbr+ ExtC+ ClaR)	No	80%-N.D.
	[10]	Meln	Laser&B/N+ Aire (Distancia)	Forma (Perfil+ CHT+ ClaR)	Sí	100%-0%
CIRAA	[24]	Toma	Color (Espectral)	Puntual (Hue&Sat+ Umbr+ Estéreo)	No	90%-N.D

[1] Manz=Manzanas, Narj=Naranjas, Toma=Tomates, Meln=Melones, Melc=Melocotones.
[2] B/N= cámara blanco y negro, Color= Cámara en color, F=Filtro óptico, L=Luz artificial, Aire=Sopladores de Aire para mover las hojas.
[3] Umbr=Segmentación aplicando Umbrales, ExtC=Extracción de Características, Compar=Comparación de plantillas (Template Matching), ClaL=Clasificador Lineal, ClaB= Clasificador Bayesiano, ClaR=Clasificador basado en reglas, RGB=Espacio de características Red-Green-Blue, Hue&Sat=Espacio de características Tono-Saturación (Hue-Saturation), CHT=Transformada Circular de Hough, Gradiente=Imagen de Gradientes locales, Convx=Imagen de Convexidades, Perfil=Imagen de perfiles.
[4] N.D.=No Declarado.

Tabla 1.1: Resumen de los sistemas de visión más importantes para la detección y localización de productos agrícolas para labores de recolección.

- *Regiones confusas.* La presencia de regiones como áreas visibles de suelo, la aparición de cielo en la imagen o del Sol a través de las hojas y otros posibles objetos presentes en el entorno, pueden generar áreas confusas en la imagen. Por ello se suele recomendar trabajar de noche o con cubiertas opacas detrás del árbol para evitar estos problemas y obtener mejores resultados.

La captación de información espectral, hace que en determinadas aplicaciones se facilite mucho la segmentación de los frutos con respecto al fondo de hojas. Sin embargo los problemas anteriores siguen presentes y la utilización del color hace que la técnica no sea válida para fruta con color similar a las hojas.

La captación de la distancia en la imagen hace que la mayoría de los problemas que acabamos de presentar desaparezcan. Principalmente los relativos a las sombras y a la presencia de regiones confusas por visualización del cielo o del Sol a través de las hojas. Además se obtienen de una forma directa las coordenadas de localización del fruto. Sin embargo este tipo de imágenes no resuelve el problema del ocultamiento. La única referencia que utiliza una captación de la distancia se aplica en el sistema de visión cercano para la recolección de melones [10]. En este caso, los problemas de oclusión se resuelven aplicando una turbina de aire, con lo que se obtiene unos resultados muy buenos.

1.2.3.2 Métodos de análisis aplicados

En cuanto al proceso de reconocimiento, las técnicas revisadas en la sección anterior muestran dos tipos de aproximaciones fundamentales: las basadas en *propiedades puntuales* y las que se basan en la *forma superficial.*

Las técnicas basadas en *propiedades puntuales* utilizan el valor o los valores asociados a cada punto de la imagen para decidir si dicho punto pertenece a un fruto o al fondo. Estos valores suelen ser el nivel de intensidad o nivel de gris y las componentes de color de estos puntos (RGB, HSI o alguna componente espectral aislada) [160, 45, 170, 188, 189, 76, 187, 117, 112, 27, 46, 24]. Estos métodos de reconocimiento suelen ser sencillos y rápidos en términos de cómputo con lo cual son muy atractivos para ser ejecutados en aplicaciones que trabajen en tiempo real. Sin embargo, se caracterizan por ser poco flexibles ya que una vez desarrollados para una determinada aplicación no sirven para otras, debiéndose reajustar los parámetros internos del sistema de reconocimiento. También suelen ser poco robustos puesto que son muy sensibles a las variaciones de iluminación necesitándose reajustes de los valores del umbral. Estos ajustes son muy críticos y se suelen hacer basándose en la experiencia o mediante métodos automáticos que no siempre son adecuados [45].

Los sistemas de reconocimiento basados en la *forma* tienen en cuenta el grado de convexidad de las superficies o analizan el grado de circularidad de los contornos y perfiles de los objetos [213, 130, 165, 10]. Estos métodos son más generales al poderse aplicar para detectar frutos independientemente de su color. Sin embargo, la utilización de estas técnicas sobre imágenes de intensidad o sobre imágenes espectrales no es lo mas correcto. Se puede hacer una interpretación errónea de determinadas áreas de la imagen por presentar características de esfericidad que en realidad corresponden a otros fenómenos. Por ejemplo, en la figura 1.8b los contornos de las hojas presentan

características de esfericidad y esto provoca interpretaciones erróneas cuando se analiza la imagen por métodos de reconocimiento de formas [213]. Análogamente, en la figura 1.10 vemos como las áreas correspondientes al cielo tienen formas convexas ocasionando conflictos con las regiones igualmente convexas de los frutos. Los trabajos que utilizan técnicas de análisis de formas sobre imágenes de intensidad o espectrales [213, 130, 165], se caracterizan todos ellos por obtener tasas de detección errónea apreciables.

1.2.3.3 Resultados

Hemos visto en los métodos utilizados tres variantes de *captación* de imágenes: intensidad, características espectrales y distancia. Las técnicas de *análisis* a su vez se basan en propiedades puntuales (p.ej. color, intensidad) o en la forma. Los diferentes trabajos son el resultado de combinar estos enfoques, obteniéndose, según la pareja captación/análisis utilizada, las siguientes categorías :

- *Intensidad/puntual* [27, 46].
- *Intensidad/forma* [213, 165].
- *Espectro/puntual* [160, 45, 188, 189, 187, 117, 112, 24].
- *Espectro/forma* [130].
- *Distancia/forma* [10].

Salvo en el caso especial de la recolección de melones presentado por Benady y Miles [10], los resultados obtenidos se pueden resumir así: porcentajes de reconocimiento correctos limitados, que están en torno al 80-85%, detección únicamente de fruta madura (salvo en los casos de análisis por forma), tasas de detección erróneas superiores al 5-10%. Estos resultados no son malos teniendo en cuenta los porcentajes de detecciones correctas, sin embargo el hecho de que existan frecuentes detecciones falsas provocaría la activación de ciclos de recolección con resultados no productivos, ocasionando una disminución del rendimiento del robot recolector. Además si consideramos el grado de adaptabilidad a otras condiciones o variedades de fruta, los métodos se caracterizan por no ser flexibles ni generalizables. En definitiva son aún poco eficientes y sobre todo poco robustos.

1.3 La estrategia de recolección asistida: El robot AGRIBOT

La tendencia actual que han seguido los trabajos de automatización en la recolección, ha sido la de realizar una automatización total. Este objetivo ha supuesto que se presenten diversos problemas prácticos que no permiten que estos sistemas sean lo suficientemente operativos como para que puedan ser industrializados y utilizados en la recolección. A diferencia de esta tendencia el proyecto español AGRIBOT que se desarrolla en el Instituto de Automática Industrial, IAI, del Consejo Superior de Investigaciones Científicas, bajo la financiación de la Comisión Interministerial de Ciencia y Tecnología (CICYT-TAP93-0583), plantea en su primera configuración el estudio y la realización de un robot recolector de frutos frescos en árbol mediante una estrategia *asistida* [30].

1.3.1 Principio de operación asistido

La estrategia *asistida* supone un cambio de concepción frente a las soluciones previas y pretende introducir un modo de operación semi-automático donde el hombre y la máquina cooperen de una forma armónica distribuyéndose las tareas en función de las mejores cualidades de cada uno. De esta forma las tareas, con un alto componente intelectual, como el guiado por el campo y la detección de los frutos le correspondería al operador, ya que estas tareas son sencillas de realizar y casi no suponen ningún esfuerzo físico a una persona. Por el contrario las tareas de localización, aproximación al árbol, corte y almacenamiento de la fruta son encomendadas al robot debido a que para este son tareas muy sencillas de realizar (como ocurre en el caso de la localización una vez que el operario ha señalado la fruta), o bien, debido a que son labores que requieren un esfuerzo físico y sin embargo al robot no le supone ningún problema (como en el caso de la aproximación, corte y almacenamiento). Cabe destacar que existe otra tarea que es la de aprehensión o agarre de la fruta que supone un esfuerzo para el operario y para el robot no es una tarea sencilla. En este caso la tarea ha sido encomendada al robot puesto que en caso contrario no tendría sentido la presencia de un manipulador y evolucionaríamos a una solución totalmente manual.

La estrategia asistida supone un planteamiento realista, por la que se intenta eliminar las principales dificultades encontradas, utilizando la cooperación humana en tareas triviales para este. De esta forma se pretende hacer viable la recolección individualizada de frutos, aunque la estrategia semiautomática suponga siempre la presencia de un operario. A pesar de no conseguir una automatización total, el enfoque es igualmente interesante puesto que puede suponer una mejora en la producción, en la calidad de los productos y una disminución de costes al reducir significativamente la mano de obra.

1.3.2 Descripción del robot AGRIBOT

La estructura del robot recolector propuesta en el proyecto Agribot se presenta en la figura 1.13. Puede verse que el robot consta de una plataforma móvil que se mueve a lo largo de los pasillos formados por cada dos hileras de árboles. Sobre la plataforma, en una cabina de mando, va situado un operador y a ambos lados se disponen dos brazos recolectores. Sobre la cabina del operario está colocado un sistema de localización basado en un telémetro láser que será utilizado para que el operador, desde su puesto de control, mediante un joystick apunte individualmente a cada fruto y de esta forma podamos calcular la posición tridimensional de cada fruto. Mientras se van adquiriendo estas coordenadas el sistema de procesamiento controlando los brazos recolectores va distribuyendo la carga de trabajo de tal forma que se optimicen las trayectorias recorridas y por tanto el tiempo empleado en los ciclos de recogida.

La utilización de dos brazos recolectores pretende disminuir el tiempo medio de recogida de la fruta, el cual para ser competitivo debe estar en torno a los dos segundos por fruta. Un estudio de productividad en la recogida de naranjas de tipo Valencia [167] indica que la producción de un recolector humano en una hora es de 15 $arrobas/hombre \cdot hora$, incluyendo los tiempos de transporte hasta la estación de pesado, lo cual supone un ciclo de 4 segundos/pieza considerando que una arroba contiene entre 60 y 65 piezas. Esto quiere decir que si obtenemos una velocidad de recogida por cada brazo de 3

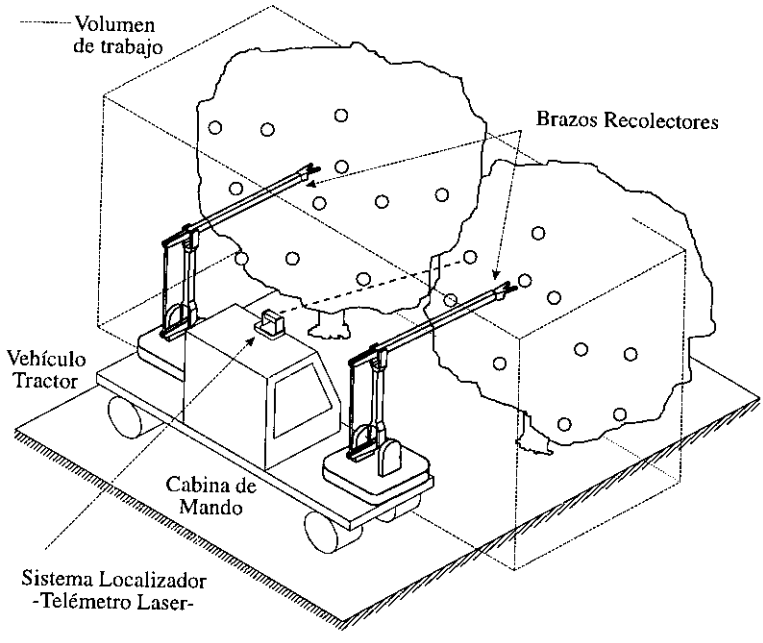


Figura 1.13: Estructura general del robot Agribot [167].

segundos/pieza, tenemos un ciclo de 1.5 segundos/pieza entre los dos brazos y por consiguiente un incremento en la recogida equivalente al trabajo de 2.6 hombres.

En el esquema de la figura 1.14 se muestran todos los componentes físicos de los que consta el robot Agribot necesarios para realizar el proceso de control y actuación para el cual fue diseñado. Así mismo podemos apreciar el flujo de la información y los comandos que se transmiten a lo largo de las líneas de conexión existentes entre dichos componentes, siendo el operario el agente que desencadena las acciones de recolección al ir señalando secuencialmente cada uno de los frutos. A continuación vamos a presentar una breve descripción de cada uno de los elementos físicos de que consta el robot Agribot:

- *Manipulador.* Tras un estudio geométrico, cinemático y dinámico, analizando diferentes tipos de manipuladores se diseño y construyó un brazo con estructura angular de paralelogramo [168, 167] (fig. 1.15). Esta elección se hizo así fundamentalmente porque esta solución presentaba una mejor distribución de masas y una atenuación de los efectos no lineales que aparecen en su control. El manipulador fue construido con cuatro grados de libertad, cada uno de los cuales se corresponde al equivalente humano para la cintura, hombro, codo y muñeca. Cada uno de estos ejes está accionado por motores de corriente continua de 80, 2x250 y 90 w respectivamente, con codificadores ópticos y reducciones adecuadas para conseguir las precisiones de posicionamiento deseadas. Los cuerpos principales del brazo son dos perfiles estándar de aluminio, lo cual permite que este tenga un peso muy ligero (16.8 Kg), con unas longitudes de 90 y 110 cm lo que supone una

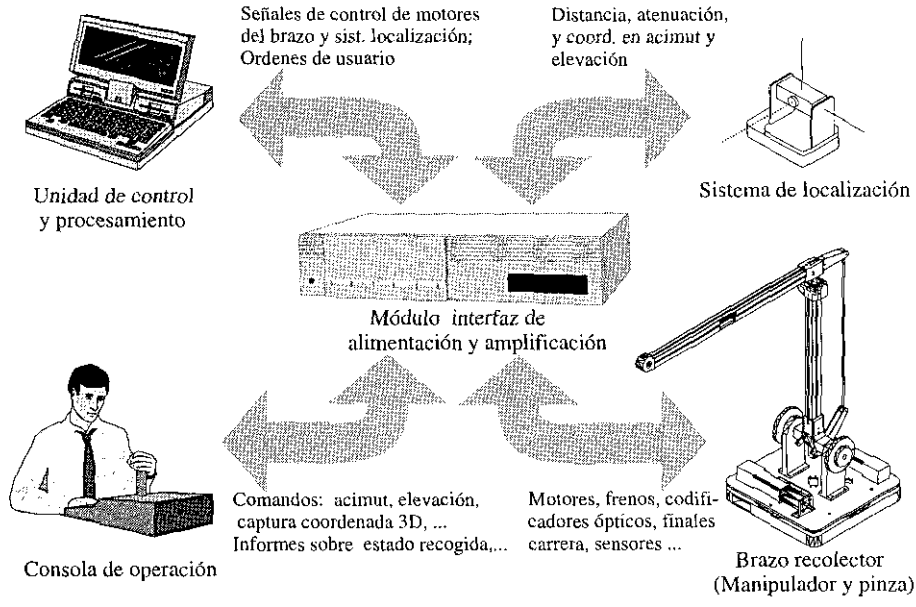


Figura 1.14: Componentes, interconexionado y flujo de información del robot Agribot.

envergadura total de más de dos metros.

- *Pinza.* El primer prototipo de pinza diseñado y construido en el IAI [209], es un dispositivo mecánico que utiliza sensores y actuadores electro-ópticos y neumáticos. Las tareas que debe realizar la pinza son variadas; por un lado debe actuar en la aproximación final al fruto para situarse en disposición de capturar la fruta, para ello, mediante el movimiento controlado del manipulador y utilizando un dispositivo pasivo de autocentrado en forma de V se consigue posicionar el pedúnculo del fruto en el vértice de la V donde se encuentran dos sensores de infrarrojos que indican la presencia del pedúnculo de la fruta (fig. 1.16a). Una vez detectada esta situación, se acciona una ventosa junto con un venturi y un sistema de compresión para realizar una aspiración sobre la superficie de la fruta para conseguir fijarla. La correcta captura de la fruta es detectada utilizando un sensor de presión que permite discernir cuando se produce una disminución de la presión atmosférica en el interior del conducto de aspiración. Una vez confirmada esta situación se activa una sierra circular para cortar el rabo de la fruta. Después del corte, mediante gravedad, un balancín que soporta la ventosa pasa de la posición horizontal inicial a un vertical donde la fruta queda enfrentada a la boca de una manga flexible que está acoplada a la estructura del brazo (fig. 1.16b). Cuando la aspiración cesa la fruta entra en la manga y cae por gravedad al depósito de almacenamiento situado en el propio vehículo.

- *Sistema de localización.*

  El componente principal de este sistema es un telémetro láser para la medida

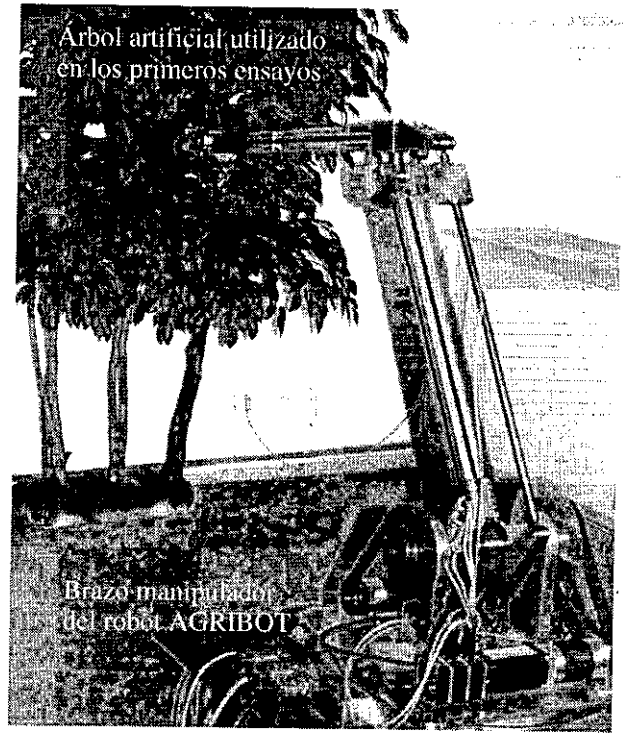


Figura 1.15: Manipulador del robot Agribot.

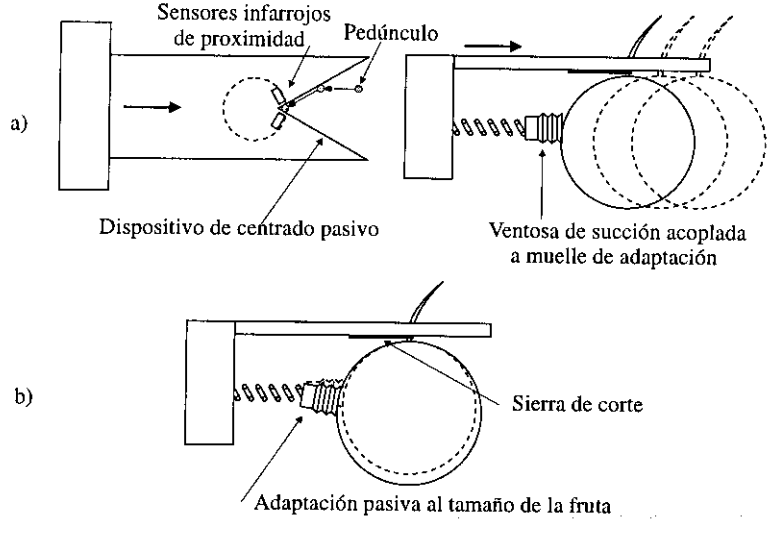


Figura 1.16: Primer prototipo de pinza del robot Agribot: (a) proceso de aproximación de la pinza y detección del pedúnculo, (b) operaciones de agarre y corte.

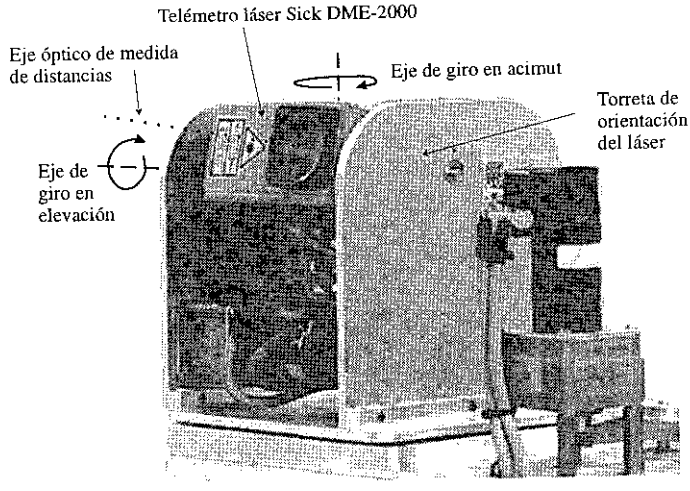


Figura 1.17: Sistema para la localización asistida de frutos del robot Agribot.

de distancias (modelo DME-2000 de la firma Sick), presentando como principales características, un rango neto de medida de 2 metros, precisión de 10 mm, frecuencia de medida 30 Hz, potencia 3.4 mW, longitud de onda 670 nm y clase II. Este telémetro va montado sobre una torreta, desarrollada igualmente en el IAI [166], que permite moverlo con dos grados de libertad: acimut y elevación, mediante dos motores DC de 15 W, pudiéndose conocer la posición actual de cada eje en cada momento mediante el uso de codificadores ópticos (fig. 1.17). De esta forma disponemos de un sistema de medida en coordenadas esféricas, que es utilizado para obtener la posición espacial del fruto cuando el haz láser visible incide sobre la superficie de este.

- *Unidad de control y procesamiento.* El componente central de esta unidad es un ordenador personal industrial con microprocesador Intel Pentium 150 MHz, en cuyas ranuras de expansión incorpora tarjetas de adquisición de señales tanto analógicas como digitales para la lectura de señales generadas por diversos sensores (interruptores, pulsadores, sensores de vacío y fotoeléctricos, potenciómetros de joystick, lecturas analógicas del telémetro, etc ...), para la activación de diversos dispositivos (sierra, electroválvulas de succión, frenos de motores, ...), y unidades controladoras de motores de corriente continua para el control descentralizado de seis motores. Apoyándose en este soporte físico, los algoritmos desarrollados sobre un entorno de programación multitarea (RTKernel) gestionan en tiempo real la generación de trayectorias para el desplazamiento de los brazos, el proceso de aproximación, agarre, corte y depositado de la pinza, a la vez que atiende las acciones del operador de señalización de los frutos por medio de una palanca de mando o joystick.

- *Consola de operación.* Esta consola es la interfase de comunicación hombre-máquina en la cual el componente principal es la palanca de mando con su

correspondiente pulsador para señalar el instante en que las coordenadas esféricas actuales corresponden a la posición espacial de un fruto. Adicionalmente esta consola incluye diversos pulsadores, selectores y visualizadores para marcar los inicios y paradas del proceso de recolección, definir orígenes de coordenadas, cambiar modos de trabajo y tener una información del estado actual de operación.

- *Alimentaciones y etapas de potencia.* En estas se generan las diferentes tensiones continuas (48,24,+/-15,5 voltios) necesarias para alimentar los motores y los dispositivos electrónicos. Además se incluyen etapas de potencia necesarias para convertir las señales de control de baja potencia generadas en las tarjetas controladoras de motores del PC, en señales con la potencia necesaria para mover cada uno de los motores.

1.3.3 Resultados y discusión

De los dos brazos manipuladores previstos en el diseño inicial se ha construido uno, que permite obtener una velocidad lineal máxima del extremo del brazo de 2.8 m/s en el caso de una extensión del brazo completa, consiguiéndose un alcance de 2,2 metros desde el eje de la cintura del brazo. El tiempo medio alcanzado para un movimiento típico de aproximación al árbol es de 1.7 segundos, mientras que el tiempo empleado en el proceso de agarre, corte y depositado ronda los 3 segundos, lo cual, unido al tiempo de señalización, permite realizar ciclos de captura a una cadencia de una fruta cada 6 segundos, que es significativamente superior al tiempo de 3 segundos marcado como objetivo en cada brazo.

El sistema de localización presenta un buen comportamiento alcanzando precisiones de 5 mm para objetos con buena reflectividad, degradándose la precisión paulatinamente a medida que aquella disminuye, y alcanzándose 11 mm de precisión para reflectividades del 20-30%. En cuanto a la señalización del fruto utilizando el haz luminoso del telémetro láser Sick-DME-2000, la detección del punto luminoso de color rojo por parte del operario, es factible en ambientes internos con iluminaciones tenues, pero resulta casi imposible su visualización en ambientes con fuerte iluminación solar. Este hecho hace que sea necesario aumentar la potencia de emisión del haz señalizador y a su vez disminuir la longitud de onda del láser utilizado para trabajar en una banda donde el ojo humano sea más sensible (555 nm).

La tarea de señalización manual de cada uno de los frutos supone un alto esfuerzo de concentración del operario que en muchas ocasiones no puede seguir señalando porque el brazo robótico interfiere la línea de visualización o porque el proceso de captura provoca un cierto movimiento en las ramas y frutos a los que se pretende señalar. Teniendo además en cuenta que el ciclo pretendido es de 1.5 frutas por segundo, esto implica que el operario debe realizar el proceso de señalización y pulsación a esta misma cadencia, lo cual supone un gran estrés y crea la necesidad de realizar descansos o cambios de turno de trabajo frecuentemente. Es por ello que sería conveniente facilitar de algún modo el proceso de apuntamiento de la fruta para hacer que la intervención del operario no sea tan intensa, para lo cual un sistema de reconocimiento automático que captase los frutos más evidentes sería lo más aconsejable.

Capítulo 2

Estrategia de detección mixta. Estudio, configuración y caracterización de un sensor de distancia y reflectancia

Resumen. *En este capítulo proponemos una nueva estrategia para la detección y localización de frutos: "Estrategia de detección mixta", que integra un método automático con el modo asistido ya existente en el robot Agribot. Para dotar al robot de un modo automático suficientemente robusto, se propone basar el reconocimiento en el análisis de formas, propiedades ópticas y distribución espacial de los objetos, para lo cual se requiere la generación de imágenes de distancia y reflectancia. Se fijan las especificaciones requeridas de éstas imágenes y se realiza una revisión de las diferentes técnicas que permiten captar ambos parámetros. La telemetría láser operando según el principio de diferencia de fases es considerada la técnica más adecuada. Seleccionamos el tipo de telémetro láser que mejor se ajusta a nuestras especificaciones y realizamos un análisis tecnológico donde se plasma la problemática actual para conseguir telémetros que satisfagan nuestros requisitos con amplitud. Presentamos la configuración física de un sistema de barrido por deflexión de haz, que utilizando el telémetro láser puntual previamente seleccionado, nos permite obtener imágenes de distancia y reflectancia.*

Finalmente, realizamos la caracterización del telémetro láser, presentando un modelo matemático que relaciona la información de reflectancia suministrada por el sensor con diversos factores que intervienen en el proceso de medida (distancia, reflectividad, ángulo de incidencia, etc...). Análogamente, modelamos la repetitividad de la medida de distancia la cual depende fundamentalmente de la reflectancia y del periodo de muestreo. Como se verá en posteriores capítulos, estos dos modelos constituyen las bases para: 1) integrar la doble información distancia/reflectancia, 2) deducir la reflectividad de una superficie, y 3) estimar el ruido del mapa de distancia para su posterior restauración.

En el capítulo anterior hemos visto que la estrategia de recolección asistida que se presenta en el proyecto AGRIBOT, permite reducir la complejidad de las tareas asignadas al robot recolector mediante la cooperación del operario con la máquina en las etapas de guiado y detección de las frutas. Para la localización de la fruta, el operario utiliza un sistema de señalización láser movido por una palanca de mando mediante el cual comunica al sistema de control del robot las coordenadas de las frutas que ha reconocido. Teniendo en cuenta que la velocidad de ciclo pretendida para el robot Agribot es de 1.5 segundos/fruto [167], nos encontramos con que el proceso de señalización es una tarea intensa y el operario difícilmente va a ser capaz de realizar la señalización con esta cadencia, especialmente si se pretende que este trabaje de forma continuada durante una jornada laboral. Por estos motivos nos planteamos la forma de dotar al robot de un sistema automático adicional de detección y localización de frutos que permita al operario reducir el tiempo dedicado a la señalización manual de las frutas. Esta solución que proponemos la hemos denominado *estrategia de detección y localización mixta.*

2.1 Estrategia de detección mixta: Detección/localización automática y asistida

2.1.1 Principio de operación mixto

Este principio de operación pretende integrar la estrategia de detección y localización *asistida*, planteada en el capítulo anterior para el robot Agribot (sección 1.3), con una estrategia *automática*. En el caso asistido, el operador detecta la fruta reconociéndola visualmente y posteriormente el robot determina la localización de la fruta por las coordenadas esféricas del sistema de localización, el cual fue previamente posicionado por el operario moviendo una palanca de mando. En un sistema totalmente automático de reconocimiento, tanto la detección como la localización se realiza sin ninguna intervención del usuario. Por separado, tanto la estrategia asistida como la automática, tienen su propio sentido. Nosotros queremos utilizar ambos principios e integrarlos, de tal forma que se complementen mútuamente para paliar las desventajas intrínsecas de cada método.

La principal desventaja de un método automático es que no reconoce una cierta proporción de las frutas, especialmente cuando el porcentaje de oclusión del fruto es elevado ($> 70\%$). La desventaja del proceso asistido es la labor requerida en la señalización de cada una de las frutas que resulta muy laboriosa e impide obtener las velocidades de señalización deseadas. Si integramos ambos métodos realizando primeramente un reconocimiento *automático*, con lo cual detectaríamos y localizaríamos la mayor parte de las frutas, y posteriormente aplicásemos el método *asistido* sobre las frutas que no ha sido capaz de detectar el sistema automático, tenemos por un lado, que las frutas no detectadas por el sistema automático son escasas y por tanto hay que realizar pocas señalizaciones manuales en modo asistido, con lo cual la desventaja del modo asistido desaparece; por otro lado, las frutas más complicadas de reconocer que no fueron detectadas por el sistema automático, finalmente son detectadas al aplicar el modo asistido, con lo cual la desventaja del proceso automático también desaparece.

Para llevar a la práctica la estrategia mixta de percepción disponemos en la actualidad del modo de localización asistido, sin embargo no tenemos ningún sistema de reconocimiento

automático, y los revisados en el primer capítulo (sección 1.2) no son suficientemente robustos y suelen presentar detecciones falsas. Por ello tenemos la necesidad de integrar al modo asistido un sistema automático de reconocimiento y localización que detecte el mayor porcentaje de frutos, y especialmente que no genere detecciones falsas. Como ya adelantamos en el apartado inicial "Objetivos y metodología del trabajo", este es precisamente el objetivo central de la tesis.

2.1.2 Captación de distancia y reflectancia: Análisis de formas, propiedades ópticas y distribución espacial.

Con el objetivo de desarrollar el sistema automático de detección de frutos correspondiente a la estrategia de percepción mixta propuesta para el Agribot, primeramente vamos a analizar cual debe ser la técnica de captación más apropiada para el problema planteado. Para ello hemos analizado los resultados obtenidos en los trabajos previos de detección de fruta y en otros procesos afines de reconocimiento, si bien antes vamos a formalizar algunos conceptos con los que vamos a trabajar en esta tesis.

En la mayoría de estos trabajos, los sensores utilizados son cámaras CCD en color o en blanco y negro, con lo cual obtienen imágenes denominadas de "intensidad". Este término es una simplificación de la magnitud física que en realidad se capta que es la intensidad radiante ó potencia radiante por estereoradian $[w/sr]$. En otros trabajos también se puede oír hablar de términos como reflectancia, reflectancia registrada o reflectancia aparente, todos ellos refiriéndose al mismo concepto de una imagen de intensidades que se caracteriza por ser muy similar a aquellas captadas con nuestros ojos. Existe otro concepto que está ligado a las propiedades intrínsecas de una superficie, que relaciona el flujo radiante que incide sobre ella y el flujo total que refleja, este parámetro se conoce como coeficiente de reflexión, reflectividad o reflectancia, siendo todos ellos sinónimos, y se representa por la letra griega ρ. Como se puede ver, entre el primer concepto y el segundo existe una coincidencia de términos al hablar de reflectancia. A partir de este momento y para que no haya confusión nos referiremos a estos dos conceptos mediante estos términos:

- *Reflectancia ($\Re$).* Utilizado para designar la intensidad radiante o la reflectancia registrada mediante un sensor al medir una región de un entorno iluminado.
- *Reflectividad (ρ).* Utilizado para designar la propiedad de una superficie que relaciona el flujo reflejado con el flujo incidente y que depende de la longitud de onda de la luz utilizada.

Una vez definido lo que entendemos por reflectancia, podemos decir que alguna de las limitaciones de este tipo de imágenes, tal como se vio en el capítulo 1 (sección 1.2.3), son los contrastes producidos por la iluminación natural que crea fuertes sombras, la aparición de regiones confusas debidas al Sol o al cielo visible a través del árbol, y la ausencia de información de profundidad. Los valores puntuales de reflectancia en la imagen son parámetros bastante variables y por tanto no constituyen bases discriminantes estables. Un buen método de reconocimiento debe basarse en la caracterización del objeto mediante propiedades estables y que no cambien al presentarse dicho objeto en

diferentes entornos o condiciones. Por ello, un método de detección de frutos debe utilizar características intrínsecas al propio fruto y depender lo mínimo posible de otros factores externos como las condiciones de iluminación o ambientales. Debido a que los *mapas de distancia* son independientes de la iluminación y la reflectividad, y que las sombras o las marcas no aparecen, el proceso de reconocimiento de objetos basándose en la *forma* debería ser menos difícil en imágenes de distancia que en imágenes de reflectancia [100]. No por ello la información de reflectancia hay que desestimarla, ya que si la iluminación está controlada, permite captar propiedades ópticas de la superficie de los objetos posibilitando su caracterización. De esta forma se aportan factores discriminantes adicionales a los generados mediante un análisis de forma en mapas de distancia.

La *distancia* a la que se encuentran los objetos (frutos), es un factor esencial para obtener la localización tridimensional de estos y para que no se requiera utilizar la técnica de aproximación ciega del brazo manipulador a lo largo de la línea de visualización del fruto, que ralentiza el ciclo de trabajo del robot. Adicionalmente la información de distancia permite realizar análisis basados en la distribución espacial de los objetos, es decir, se puede limitar el espacio de búsqueda en la *imagen de distancias* a aquellos volúmenes donde se prevea la aparición de los objetos de interés, o como es nuestro caso, al volumen de trabajo del robot manipulador.

Por los motivos expuestos proponemos realizar el sistema de detección y localización automático mediante un análisis basado en la *forma*, en las *propiedades ópticas* de una superficie, y en la *distribución espacial* de los objetos, para lo cual se deberá utilizar una técnica de captación que genere imágenes incluyendo información de distancia y reflectancia.

2.1.3 Especificaciones de las imágenes de distancia y reflectancia

En este apartado vamos a definir las características que debe satisfacer la técnica de captación de imágenes de distancia y reflectancia, para que un proceso de análisis sea viable. Teniendo presente que nuestro objetivo se centra en la detección de frutos tales como naranjas, manzanas o melocotones, esto implica tamaños medios del objeto a detectar entre 60-130 mm y una distancia de posicionamiento respecto al sensor entre 1 y 4 metros. Uniendo estos factores a otros requisitos básicos de resolución espacial, resolución en profundidades y a otras consideraciones necesarias para aplicar las técnicas de análisis en tiempo real, obtenemos que la técnica de captación de profundidad/reflectancia debe cumplir las siguientes especificaciones:

- *Distancia absoluta.* Para determinar la localización espacial del objeto, es necesario conocer la distancia absoluta del sensor a la superficie a medir. Esta información además de facilitar la localización, permite eliminar ciertas ambigüedades y limitar el volumen de análisis de las imágenes.

- *Reflectancia.* La reflectancia deberá ser medida para posibilitar el análisis basado en propiedades ópticas y para complementar la información de distancia. La reflectancia deberá ser lo más estable e independiente de factores variables externos como la iluminación o los contrastes producidos por sombras, para lo cual la iluminación deberá estar controlada.

- *No contacto.* El método de medida debe ser sin contacto. En este sentido, son desaconsejables los métodos táctiles que requieren contacto y una búsqueda secuencial a lo largo del entorno, normalmente guiados por un brazo móvil, necesitando mucho tiempo para obtener la información tridimensional. Estos métodos además son invasivos, interactuando con el medio.

- *Robustez en la captación.* El principio de medida debe verse influenciado lo mínimo posible ante fenómenos perturbadores. Por ejemplo por variaciones de temperatura y de iluminación ambiental, por aparición de determinadas texturas, por presencia de discontinuidades en las superficies o por perturbaciones producidas por la aparición de objetos próximos al eje de medida. Este último caso se da en técnicas basadas en triangulación apareciendo regiones donde no es posible medir debido a que la señal utilizada queda interrumpida (fig. 2.11). El resultado es una falta de información en determinadas áreas que se conocen por *zonas de oclusión* y *sombras*, las cuales son debidas al método de medida utilizado.

- *Rango de medida en distancia.* El rango de medida del sensor debe alcanzar los cuatro metros para cubrir el volumen de trabajo definido por el conjunto formado por el robot manipulador y el árbol (fig. 1.13).

- *Precisión en posición 3-D.* La *precisión* es una medida de la máxima desviación de una medida frente al valor real, y se suele expresar en términos del error cuadrático medio como $\pm 2\sigma$. La precisión en la localización de un punto en el espacio viene dada por la precisión en cada una de las coordenadas del sistema de referencia utilizado. Por ejemplo, si el resultado se da en coordenadas esféricas, intervienen la precisión angular y la precisión en la determinación de la distancia. Para el caso del robot Agribot, se considera que el sistema puede admitir errores máximos de ± 10 mm en cada uno de los ejes cartesianos.

- *Resolución en distancia y repetitividad.* La *resolución* es una medida que indica el cambio más pequeño en distancia que el sensor es capaz de detectar. Con el objeto de extraer información fiable en cuanto a la forma de los objetos, la resolución en distancia debe ser suficiente como para que la forma del objeto no quede degradada. En el caso del Agribot, para detectar cuerpos esféricos en torno a 35-40 mm de radio, que es el caso más típico, se estima que 1 mm de resolución en distancia es suficiente. Sin embargo, este parámetro por sí solo no es suficiente si no viene acompañado de una buena *repetitividad*, es decir, la medida repetida sobre un mismo punto debe dar una desviación estándar que se estima no debe superar 1 mm. Ambos factores, resolución y repetitividad, son muy importantes a la hora de captar la forma de las superficies en la imagen de una forma fiable y ajustada a la realidad.

- *Resolución espacial.* Para realizar un muestreo de la escena que permita captar la forma de las superficies, la resolución espacial o separación deseable entre puntos contiguos de la imagen estaría aproximadamente entre 2 y 4 mm.

- *Tiempo de medida.* El tiempo de captura debe ser lo suficientemente pequeño como para que considerando también el tiempo de procesamiento de la imagen, no se supere el ciclo de trabajo de la aplicación considerada. En el caso concreto del Agribot, considerando dos brazos recolectores trabajando, el ciclo de trabajo

estimado es de 1.5 segundos por fruto. Si asumimos que un tercio del ciclo de trabajo puede estar dedicado a tareas de detección y localización, y los otros dos tercios se dedican a la captura y manipulación, tenemos que el tiempo dedicado a la detección y localización es de 0.5 segundos/fruto. Según un estudio realizado en una plantación de naranjos, hemos podido comprobar que el promedio de frutos en una ventana de 50 × 50 cm es alrededor de 10 frutos. Esto implica que el tiempo que podemos dedicar a realizar una captura de una imagen de 50 × 50 cm es de 5 segundos, y si suponemos que necesitamos 3 mm de resolución espacial, esta imagen corresponde a una de 167 × 167 puntos con lo cual hay que captar 28.000 puntos en 5 segundos, lo que implica una velocidad de muestreo de 5.6 kHz.

Con estas especificaciones se han dado datos cuantitativos referentes a los requisitos necesarios para obtener imágenes de distancia/reflectancia con la suficiente fidelidad como para poder plantear una posterior etapa de reconocimiento y localización de frutos basándose en la *forma*, en las *propiedades ópticas* superficiales y en la *distribución espacial* de los frutos. A pesar de que las especificaciones han sido dadas atendiendo a un problema concreto, el planteamiento tiene la suficiente generalidad como para permitir cubrir un gran número de problemas y aplicaciones de reconocimiento, con excepción de aquellas que trabajan con objetos grandes a largas distancias, o con objetos pequeños a cortas distancias.

2.2 Técnicas para captar distancia y reflectancia

En la sección anterior hemos planteado la necesidad de utilizar una técnica capaz de captar tanto información tridimensional como de reflectancia, por ser esta información muy completa para permitir aplicar posteriores análisis de reconocimiento automático de objetos. No es un aspecto nuevo el hecho de utilizar información tridimensional y de reflectancia para interpretar una escena. Los seres humanos utilizamos una serie de deducciones automáticas para obtener información tridimensional a partir de dos vistas de una escena o incluso a partir de una sola vista analizando gradientes en iluminación o experiencias acumuladas. En este sentido han surgido numerosos trabajos de investigación para extraer información tridimensional mediante diferentes estrategias, como la telemetría láser [50, 13, 67, 101, 110, 132, 179, 183, 186, 40], la estructuración de la luz [65, 66], la estéreo visión [55, 102, 146], ultrasonidos [133] ó mediante pieles táctiles [177], registrando en algunos casos tanto distancia como reflectancia [155]. A continuación vamos a revisar las técnicas más representativas de captación, tanto activas como pasivas, que nos van a permitir sentar las bases para posteriormente seleccionar el principio de medida más adecuado a nuestro problema.

2.2.1 Técnicas basadas en cámaras ópticas

Las cámaras son los principales dispositivos utilizados en los sistemas de visión artificial. Teniendo en cuenta el principio de captación utilizado, existen dos tipos de cámaras, las basadas en tubos de rayos catódicos (Vidicon) y las basadas en captadores de estado sólido (CCD y CID). Tanto en un caso como en otro, hablamos de sensores pasivos, puesto

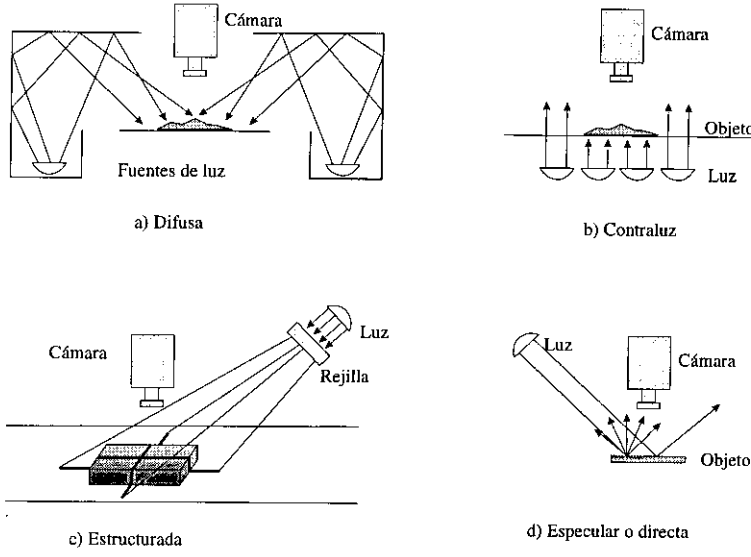


Figura 2.1: Técnicas fundamentales de iluminación para cámaras.

que utilizan la propia energía electromagnética reflejada por los objetos del entorno, para captar las imágenes. En este sentido, para este tipo de sensores, es de vital importancia la correcta iluminación de la escena a registrar. Por tanto, a continuación vamos a dedicar un apartado a las técnicas y fuentes de iluminación más empleadas para la captación adecuada de imágenes utilizando cámaras.

Técnicas y fuentes de iluminación asociadas. Tal como se ha indicado las cámaras son sensores pasivos y requieren por tanto una excitación del entorno mediante iluminación. La iluminación de una escena es un factor muy importante ligada a la complejidad de esta. Una iluminación arbitraria del entorno no suele ser aceptable puesto que puede producir imágenes de bajo contraste, reflexiones especulares, sombras y otros tipos de efectos no deseables. Un sistema de iluminación bien diseñado, debe iluminar la escena de tal forma que reduzca la complejidad de la imagen, resaltando las características importantes para la detección de objetos.

Existen fundamentalmente cuatro técnicas de iluminación de la escena. En la figura 2.1 se pueden apreciar las cuatro configuraciones que a continuación se describen:

- *Luz difusa.* La luz difusa se suele emplear cuando los objetos tienen superficies suaves y regulares, y nos interesa captar estas propiedades superficiales.

- *Contraluz.* La iluminación por contraluz produce imágenes binarias. Esta técnica es adecuada para aplicaciones en las que la silueta de los objetos es suficiente para reconocerlos.

- *Luz estructurada.* Estructurar la luz consiste en proyectar puntos, líneas o retículas de luz sobre la zona de operación. El patrón de luz proyectado sobre la zona de

trabajo es conocido, y las alteraciones de este patrón facilitan la detección de la presencia de un objeto. Analizando la deformación del patrón de luz, es posible determinar características tridimensionales del objeto.

- *Especular.* La técnica de iluminación especular o directa, se utiliza fundamentalmente para inspeccionar la superficie de los objetos. Los defectos de la superficie, tales como rugosidades o fisuras pueden detectarse utilizando un haz de luz direccional y midiendo la cantidad de luz dispersada. Si no hay defectos, se produce reflexión especular y se dispersa poca luz hacia la cámara, mientras que la presencia de defectos incrementa la cantidad de luz dispersada hacia la cámara, lo cual facilita su detección al presentarse en forma de regiones brillantes en la imagen.

Una vez apuntadas las técnicas de iluminación es necesario distinguir la fuente de iluminación más adecuada para cada situación. La fuente de luz por excelencia es la luz natural o luz solar, la cual puede ser apropiada en ciertos casos, sin embargo normalmente se utilizan fuentes de luz artificiales por ser mucho más fácilmente controlables. A continuación se describen las principales fuentes de luz artificial utilizadas en visión artificial:

- *Lámparas incandescentes.* Este es el tipo de iluminación más común, generalmente con filamentos de tungsteno o halógeno-tungsteno. Es una forma simple de suministrar luz continua en un amplio rango de potencias. Sin embargo, emiten una gran cantidad de energía en forma de calor y pierden intensidad luminosa por envejecimiento. Esta última característica, puede causar problemas al sistema de visión, salvo que se utilicen cámaras con control automático de iris o algoritmos de procesamiento poco sensibles a cambios en el nivel de iluminación.

- *Tubos fluorescentes.* Son más eficientes que las lámparas incandescentes y suministran luz más difusa, característica muy útil cuando no son deseables reflexiones especulares, aunque presentan algunos problemas de parpadeo.

- *Fibra óptica.* Se suele utilizar cuando la zona a iluminar es de difícil acceso. La lámpara empleada es generalmente de tipo cuarzo-halógeno.

- *Láser.* Es la fuente de luz habitualmente utilizada para generar luz estructurada. Los láseres más utilizados son los de helio-neón y los de diodo semiconductor. Las potencias empleadas son bajas, variando desde 1 a 5 mW.

- *Flashes.* Un flash es un dispositivo que genera una fuente de luz muy intensa, con una duración muy breve. Si los objetos se mueven a velocidades altas, se suele utilizar flashes para "congelar" la imagen y evitar obtener imágenes borrosas. La luz suministrada por el flash, debe ser varias veces mayor que la luz ambiental para conseguir el efecto deseado de captación instantánea de la imagen.

- *Ultravioleta.* La lámpara de xenón tiene una respuesta espectral dentro de la banda ultravioleta. Existen aplicaciones en las que esta luz, al incidir sobre sustancias fluorescentes, genera un alto contraste entre dichas sustancias y el resto de la escena.

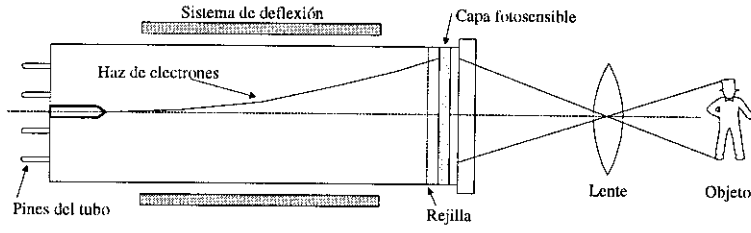


Figura 2.2: Cámara de tubo de tipo Vidicon.

Cámaras de tubo. La cámara de tubo más utilizada es el Vidicon. Existen otros tipos con características mejoradas en cuanto a la composición química del sensor, pero el principio de funcionamiento es el mismo. Las cámaras de tubo fueron populares en las primeras aplicaciones de visión artificial, pero actualmente se han visto desplazadas por las cámaras de estado sólido. Constan de un tubo de rayos catódicos sobre cuya cara frontal se proyecta la imagen de la escena (fig. 2.2). La parte interna de la cara frontal del tubo, presenta un sustancia fotosensible cuya resistencia es inversamente proporcional a la intensidad de la luz incidente. Mediante un haz de electrones se va realizando un barrido por la superficie fotosensible, en los puntos donde no incide luz la resistencia es grande, se produce una acumulación de cargas negativas y la corriente producida es baja. Por el contrario, en las zonas iluminadas la resistencia es baja, con lo cual hay una descarga y el haz de electrones produce un flujo de corriente de recarga superior al caso de no iluminación. Esta variación de corriente producida durante el barrido electrónico genera la señal de vídeo que es proporcional a la intensidad de la imagen captada. El haz de electrones recorre la superficie completamente a una frecuencia de 30 veces por segundo en cuadros de 525 líneas de las cuales 480 contienen información de la imagen. Existen otros modos de barridos, para evitar parpadeos, usando dos campos entrelazados de 265,5 líneas cada uno, o utilizando 559 líneas con 512 líneas de datos, siendo este último estándar popular en visión artificial por ser el número de líneas una potencia entera de 2.

Las cámaras vidicon tienen limitaciones en las aplicaciones industriales ya que son poco ligeras y manejables, la respuesta espectral es limitada (0,4 a 0,8 μm), y pueden distorsionar la imagen o quemarse la superficie fotosensible por exceso de iluminación. Su vida útil es limitada.

Cámaras de estado sólido. Las cámaras de estado sólido, utilizando dispositivos CCD (Charge-Couppled Devices) o CID (Charge-Injected Devices), constituyen los sensores más extendidos para los sistemas de visión artificial. Todas ellas están compuestas por series lineales o matriciales de dispositivos elementales de detección; así, considerando el tipo de imagen capturada, encontramos dos categorías principales: cámaras de línea o cámaras de área. Estas categorías aparecen según la distribución espacial de los elementos sensibles a la luz. La utilización de uno u otro tipo, depende de la clase de aplicación que se plantee. Las cámaras lineales son indicadas en situaciones en la que el objeto se mueve (p. ej. sobre una cinta de un sistema de inspección), obteniéndose finalmente imágenes sobre toda la superficie del objeto.

Los dispositivos sensibles a la luz de las cámaras de estado sólido, se fabrican

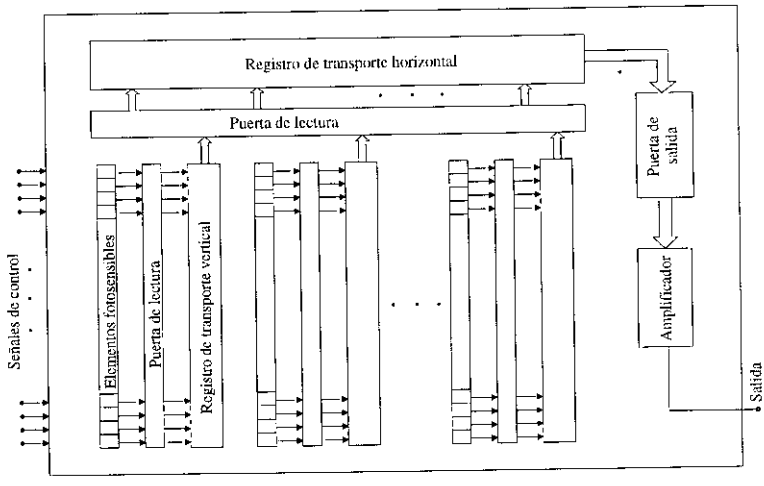


Figura 2.3: Elemento matricial de un sensor CCD.

con materiales semiconductores y contienen configuraciones lineales o matriciales de elementos fotosensibles espaciados con gran precisión (fig. 2.3). Cada una de estas células fotosensibles se puede direccionar independientemente mediante un circuito lógico de direccionamiento; de esta forma ya no se utiliza el sistema de lectura por haz de electrones usado por las cámaras Vidicon, lo cual permite obtener una mayor miniaturización. La densidad de integración de los sensores matriciales supera actualmente los 200 elementos por milímetro, permitiendo obtener 1000 puntos de muestreo en vectores CCD de 5 mm de longitud.

Los componentes CCD están basados en *elementos de transferencia de cargas en serie*, principio utilizado también en las memorias serie RAM. Cada línea consta de una serie de celdas que actúan como un condensador de tipo MOS, acumulando la carga producida por los fotones incidentes sobre cada celda. Estas cargas pasan mediante unas puertas a los registros de transporte, desde donde finalmente se llevan a un amplificador que genera un voltaje proporcional a la carga de la celda direccionada en ese momento. A diferencia, los componentes CID se basan en una matriz de condensadores MOS direccionables individualmente de forma similar a las memorias RAM. Cuando se lee una celda, un amplificador repone la carga del condensador por inyección de corriente, generándose la propia señal de vídeo.

Las cámaras de estado sólido ofrecen importantes ventajas sobre las Vidicon: Son más compactas y ligeras, los elementos fotosensibles no se degradan con el uso, poseen baja distorsión, son insensibles a los campos magnéticos, tienen baja persistencia (efecto que consiste en una permanencia de la imagen durante un tiempo pequeño cuando varía la imagen repentinamente), mejor respuesta espectral (0,4 a 1,1 μm) y hay un bajo efecto de florecimiento (fenómeno ligado a la nitidez que ocurre cuando se ilumina excesivamente un punto, el cual propaga su efecto fotoconductor a las zonas próximas, desvirtuando la imagen).

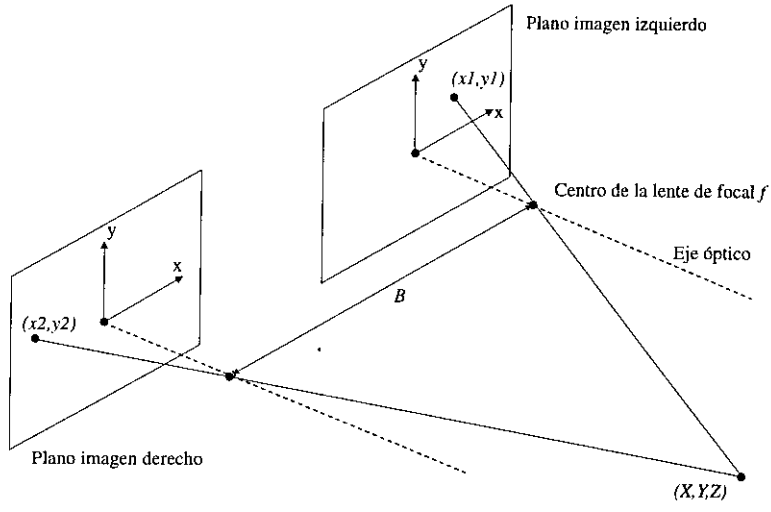


Figura 2.4: Modelo de la estéreo visión.

2.2.1.1 Estéreo visión

Mediante estéreo visión se obtienen dos vistas diferentes de una escena, normalmente utilizando dos cámaras separadas una cierta distancia a través de lo que se conoce como *línea base*. Si conocemos la posición de un mismo punto de la escena en las dos imágenes, es decir conocemos (x_1, y_1) y (x_2, y_2), el objetivo consiste en averiguar las coordenadas tridimensionales (X, Y, Z) de dicho punto en la escena (fig. 2.4). Considerando que las dos cámaras no convergen entre sí, que sus focales son idénticas e iguales a f, y que están separadas una distancia B a lo largo de la línea base, podemos utilizar las ecuaciones 2.1 para obtener la localización espacial del punto.

$$\begin{cases} Z = f - (fB/(x_2 - x_1)) \\ X = x_1(f - Z)/f \\ Y = y_1(f - Z)/f \end{cases} \tag{2.1}$$

Una vez que se conocen las coordenadas (x_1, y_1) y (x_2, y_2) correspondientes a los puntos homólogos, estas ecuaciones son muy sencillas de calcular. Sin embargo, la obtención de estas parejas de puntos es una tarea muy problemática, existiendo una gran cantidad de trabajos de investigación orientados a aportar nuevas soluciones. A estas técnicas se las conoce como *métodos de correspondencia.*

El método de correspondencia más evidente puede consistir en realizar un emparejamiento de regiones próximas en ambas imágenes y realizar la comparación mediante técnicas de correlación. También, si una imagen contiene características fácilmente distinguibles, como por ejemplo esquinas, es posible realizar la comparación a un nivel descriptivo de alto nivel, manipulando menos cantidad de información y por tanto más eficientemente. En este sentido un trabajo reciente [1] ha explorado esta posibilidad realizando una correspondencia jerárquica con dos niveles de resolución. En

un primer nivel se realiza la correspondencia de regiones segmentadas obteniendo mapas en profundidad poco precisos, y en el segundo nivel se realiza la correspondencia de las esquinas de las regiones previas, las cuales son determinadas con mayor precisión, y por tanto la información tridimensional obtenida es más precisa. Otros métodos utilizados para realizar la correspondencia se basan en técnicas de etiquetado por relajación, mediante programación dinámica, por correspondencia de subgrafos donde los nodos representan las primitivas de la imagen, o también restringiendo el área de búsqueda de correspondencias utilizando un mapa de profundidades poco preciso captado por un telémetro láser [197].

El principal problema de la visión esteroscópica es que no se suelen obtener mapas de distancias muy densos y esta densidad solo depende de las propiedades de la imagen capturada. Ello es debido a que las regiones de la imagen que son uniformes y no presentan puntos característicos, no tienen elementos que puedan ser analizados para realizar la correspondencia. Además el procesamiento necesario para realizar estos algoritmos es costoso y se hace necesario la utilización de tarjetas con procesadores especiales para acelerar el proceso. La oclusión también es un problema puesto que si en una imagen una región es visible y en la adyacente no lo es, entonces tampoco es posible realizar la correspondencia. Como ventajas podemos decir que es un método pasivo, que el coste de dos cámaras es bajo y que en el caso de imágenes bien estructuradas y con objetos claramente diferenciados se pueden obtener mapas en profundidad suficientemente precisos para algunas aplicaciones tales como el guiado de robots móviles ($\pm$45 mm a distancias superiores a 2 metros, con una línea base de 300 mm [1]).

2.2.1.2 Visión dinámica

Una forma muy interesante de obtener información tridimensional de una escena consiste en utilizar una única cámara para capturar una secuencia de imágenes, las cuales deben ser analizadas para obtener la información tridimensional. Esta técnica se puede ver como una generalización de la visión esteroscópica, donde el desplazamiento de la cámara no está solamente limitado a un desplazamiento lateral, sino que la cámara se puede mover a distintas posiciones y además podemos obtener más de una pareja de imágenes.

Existen dos enfoques fundamentales para este problema: Métodos basados en movimientos grandes y métodos con desplazamientos pequeños y sucesivos. En el primer caso se suelen capturar pocas imágenes separadas por intervalos de tiempo elevados puesto que hay que realizar movimientos de la cámara grandes. El método de análisis utilizado consiste en localizar características comunes en la pareja de imágenes y realizar el emparejamiento para obtener las correspondencias correctas. Es una estrategia similar a la visión esteroscópica y se requiere el conocimiento del desplazamiento realizado por la cámara para obtener la información tridimensional, la cual suele ser poco densa.

En la segunda aproximación, se captura una secuencia de imágenes a una alta frecuencia mientras la cámara se mueve a una determinada velocidad. En este caso, el análisis se centra en la obtención del flujo óptico entre las sucesivas imágenes o en el uso de derivadas espacio-temporales. Estas observaciones deben estar combinadas con la medida de la velocidad de la cámara (en lugar del desplazamiento) para determinar la estructura tridimensional de la escena. Sin embargo, no existe un método robusto

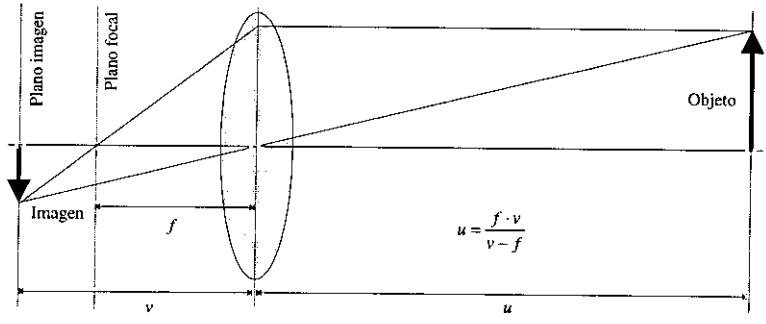


Figura 2.5: Principio de medida de distancias mediante focalización.

y preciso para realizar la correspondencia de características o para el cálculo de los campos de flujo óptico, y las aproximaciones existentes requieren unos algoritmos costosos computacionalmente. La reconstrucción tridimensional no suele ser muy fiable debido a la presencia de mucho ruido y a los fenómenos de oclusión presentes. Por este motivo es usual ver estrategias híbridas en las que se combina la velocidad de las cámaras con su desplazamiento y la extracción de características.

Recientemente están apareciendo soluciones mediante métodos activos, tomando como referencia trabajos como el de Bajcsy [5]. Estas técnicas tratan de realizar una adquisición de datos de una forma inteligente o adaptativa. El propósito consiste en obtener las posiciones o la secuencia de imágenes óptima necesaria para obtener la reconstrucción tridimensional de un objeto cuyo modelo geométrico conocemos. De esta forma, análisis que son complejos de realizar mediante una estrategia pasiva, pueden ser calculados más fácilmente mediante una solución activa, fundamentalmente porque la información captada es la más adecuada para el propósito que nos estábamos planteando.

Un trabajo reciente en esta dirección es aplicable a la caracterización y a la localización de objetos tales como cilindros, círculos y esferas en entornos estructurados [31]. En este trabajo se muestran resultados que mejoran a los precedentes que utilizan visión dinámica pasiva, obteniendo errores máximos en profundidad entre 2 y 20 mm, mejorando las precisiones obtenidas respecto a la visión dinámica pasiva que estaba en torno los 4 cm.

2.2.1.3 Enfoque

Esta técnica utiliza el enfoque o grado de contraste en la imagen para determinar la distancia. Para ello solo es necesario utilizar una cámara y mover la lente de enfoque según el eje de visión para obtener una secuencia de imágenes. Por cada una de las imágenes se debe realizar una medida del grado de contraste o enfoque de pequeñas regiones en la imagen, pudiéndose calcular la distancia a las regiones que presenten un grado de enfoque adecuado. Mediante el conocimiento de la distancia focal de la lente f y la distancia v entre el plano imagen y el plano de la lente, es posible calcular la distancia u al plano objeto (fig. 2.5).

Para obtener una mejor resolución en profundidad, es necesario utilizar lentes con

aperturas grandes para disminuir la profundidad del foco. Sin embargo es inevitable la pérdida de capacidad de discriminación según aumenta la distancia al objeto. Por tanto estos métodos son aconsejables para rangos que no superen 1 metro de distancia, si se quieren obtener precisiones menores a 1 cm.

Otro problema de estas técnicas estriba en que no es posible extraer información tridimensional si las superficies son homogéneas y no presentan algún tipo de textura. En efecto, si por ejemplo observamos una pared blanca, la imagen enfocada y las desenfocadas son iguales, y por tanto, es imposible conocer donde se produce el enfoque. Este hecho ha estimulado a utilizar patrones de luz con una gran resolución espacial para iluminar la escena y crear una textura sobre los objetos. De esta manera es posible medir la distancia incluso a superficies homogéneas. El problema aparece de nuevo cuando se intente utilizar esta técnica en ambientes exteriores donde el patrón de luz difícilmente va a generar una imagen con suficiente contraste.

En un reciente trabajo [154] se utiliza la técnica de desenfoque, en vez del análisis por enfoque, utilizando un patrón luminoso para poder ver tanto superficies con textura como sin ella. La técnica basada en el desenfoque solo requiere la utilización de dos imágenes de la misma escena. Mediante un modelado adecuado de los elementos que intervienen en la captación de la imagen, se puede calcular el punto de enfoque, y a partir de aquí proceder de una forma similar a la anterior. En este trabajo se muestran resultados de generación de imágenes 512×480 a 30 Hz y con una resolución de 1 mm, aunque el volumen de trabajo es muy reducido ($30 \times 30 \times 30$ cm).

2.2.1.4 Gradiente de textura

Las técnicas basadas en textura se inspiran en uno de los métodos utilizados por el hombre para extraer información volumétrica del entorno. Cuando observamos una superficie plana posicionada frontalmente con un cierto patrón o textura, vemos que no se produce ninguna deformación en esta. Sin embargo, al observar la superficie con un cierto ángulo, según aumenta la profundidad el patrón que define la textura se va haciendo mas pequeño. A esta variación progresiva de la textura se la conoce como *gradiente de textura*, y es utilizado para obtener información referente a distancias relativas entre superficies de una escena.

Caracterizando las texturas de la imagen, por ejemplo mediante descriptores obtenidos a partir de transformaciones de Fourier, es posible obtener información de distancia relativa. El tamaño de la ventana utilizada para caracterizar las superficies, es bastante crítico y dependerá del tipo de textura presente. Si el patrón de la textura de la imagen es conocido, cosa poco habitual, es posible también obtener coordenadas absolutas.

Una restricción importante a esta técnica, es que la textura debe ser uniforme a lo largo de las superficies, o en otro caso, las medidas se falsearían. Además, todos los elementos de la imagen deben presentar texturas para poder ser analizados, de tal forma que para una superficie uniformemente iluminada (p.ej. una pared blanca), no es posible obtener información tridimensional. Finalmente los algoritmos utilizados son costosos computacionalmente y requieren incluso la segmentación previa de la imagen.

2.2.1.5 Gradientes de iluminación

Esta estrategia trata de obtener la orientación de las superficies en una imagen a partir del análisis de la distribución luminosa a lo largo de la escena. Si consideramos que un tipo concreto de superficie presenta diferentes luminosidades cuando varía la fuente de iluminación y la orientación de la superficie, podemos representar este hecho mediante una función $\Re(p, q, l_x, l_y)$ que depende de p,q (pendientes de la superficie en direcciones x e y) y de l_x,l_y (componentes de la dirección de la luz). A esta función se la conoce como *reflectancia* de la imagen, y contiene el modelo de como debería variar la intensidad de la imagen al cambiar la orientación de la superficie y de la fuente de iluminación, para un tipo de superficie. Si obtenemos este modelo, por ejemplo experimentalmente, podemos plantear el problema como la resolución de un conjunto de ecuaciones del tipo: $I(x, y) = \Re(p, q, l_x, l_y)$, para cada uno de los casos posibles de iluminación. De esta manera, por cada punto en la imagen (x, y), obtenemos la pendiente (p, q) de la superficie en torno a su vecindario.

Este método suministra distancias relativas, no pudiendo dar información absoluta por los saltos bruscos que hay entre las superficies suaves y por la falta de una referencia absoluta. Las superficies que además reciben iluminación indirecta de otros objetos desvirtúan los resultados afectando a la precisión de los datos. Además el método, tal como se planteó, se limita solamente a objetos con un determinado tipo de superficie.

2.2.1.6 Luz estructurada

Las técnicas basadas en luz estructurada utilizan un patrón de luz conocido que se proyecta sobre la escena y mediante una cámara colocada a una cierta distancia y con un cierto ángulo de convergencia, se capta la deformación producida en el patrón por los objetos presentes en la escena. Existen varias estrategias de generación de patrones, siendo las más usuales la proyección de planos de luz verticales que se van moviendo por un barrido horizontal o la generación instantánea de múltiples planos de luz paralelos. En cualquiera de los casos la obtención de distancias absolutas requiere un análisis de la deformación de las líneas de luz, basado en el hecho de que a mayor profundidad en la escena se producirán mayores desplazamientos en las líneas de luz.

El método más simple de analizar se da cuando se genera una sola línea por cada imagen, sin embargo es una alternativa muy poco adecuada computacionalmente hablando, pues hay que analizar una imagen completa por cada línea vertical que proyectemos sobre la escena. Los problemas de oclusión provocan que las imágenes obtenidas por esta técnica no sean completas y haya regiones donde no se puede medir la distancia. La utilización de dos cámaras, en vez de una, colocadas a ambos lados del emisor del plano de luz permite eliminar estas zonas de oclusión y obtener imágenes con 0.25 mm de resolución a 0.5 m de distancia a una frecuencia puntual de 200 Hz [179].

Una alternativa más eficiente consiste en proyectar varios planos paralelos de luz, aunque en este caso el análisis se hace mas complicado pues se debe garantizar que las líneas de luz no se mezclen unas con otras, lo cual generaría un falseado en las medidas de distancia. Además, debido a las oclusiones, las líneas aparecerían interrumpidas siendo difícil saber a cual pertenece cada segmento luminoso. Para solventar este problema, se pueden generar patrones de luz donde las líneas paralelas adyacentes sean distinguibles

codificándolas por color o espacialmente [183, 186]. Además, para facilitar el análisis, la imagen debería contener un plano de fondo como referencia.

Existen otros tipos de estructuración de la luz [214], que permiten obtener la localización y orientación de áreas planas de objetos poliédricos. Consiste en iluminar la imagen con una rejilla rectangular de luz con muy alto contraste. Analizando la imagen en el dominio de Fourier las deformaciones producidas en los rectángulos se corresponden a diferentes frecuencias espaciales identificables como picos en el dominio frecuencial. Para identificar los diferentes planos basta aplicar filtros paso banda centrados en la frecuencia de cada pico.

La principal ventaja de los métodos basados en luz estructurada es la simplicidad de los dispositivos (cámara y fuente de luz) y el bajo coste. La información generada son distancias absolutas y la densidad espacial puede ser ajustable en función del patrón de luz utilizado. Como desventajas podemos citar la necesidad de analizar la imagen para obtener la información tridimensional, la posibilidad de que se generen oclusiones como ocurre en cualquier método de triangulación (fig. 2.11), y la imposibilidad de generar un patrón de luz cuando las superficies son especulares.

Franjas de Moiré. La técnica conocida como franjas de Moiré consiste en iluminar una escena a través de una serie de rejillas que generan un patrón de líneas paralelas igualmente espaciadas y ver el patrón sobre la escena a través de una configuración de rejillas idéntica. La imagen captada mediante una cámara, desplazada lateralmente con respecto a la fuente de luz, representa los contornos a igual distancia. Para conocer si existe cambio entre dos líneas adyacentes y de que signo es, basta realizar un desplazamiento en las rejillas o cambiar la frecuencia de espaciamiento entre ellas. Para realizar esto de una forma flexible, las rejillas de recepción se suelen sustituir por otras virtuales creadas al explorar la imagen en líneas verticales igualmente espaciadas.

Esta técnica puede ser utilizada para medir *distancias relativas* en superficies continuas, pero la medida de *distancias absolutas* se pierde cuando existen discontinuidades en la imagen mayores de 2 mm o la pendiente de la superficie es superior a 60^o. Las oclusiones en la iluminación que se pueden crear debido a la existencia de discontinuidades también son un problema. Por el contrario como ventajas están la captura directa de la información sin necesidad de barrido, la resolución en profundidad obtenible (10 μm en un rango de 5 cm) y el bajo coste.

2.2.2 Técnicas ultrasónicas

Utilizando sensores ultrasónicos es posible medir la distancia a la que se encuentra un objeto. Cuando una onda ultrasónica se transmite por un medio y se encuentra con un cambio de impedancia acústica (p. ej. cuando incide sobre un objeto rígido), parte de la energía se propaga por el nuevo medio y el resto se refleja creando un eco. La técnica más utilizada, *pulso-eco*, consiste en emitir un tren de pulsos y medir el tiempo transcurrido hasta que se recibe el primer eco correspondiente al objeto más cercano. Los transductores son de distinto tipo, siendo los más usados los electrostáticos y los piezo-eléctricos. El mismo transmisor puede ser utilizado tanto para emitir como para recibir la señal de eco. Conociendo la velocidad de propagación del sonido y el tiempo transcurrido en el viaje de ida y vuelta del pulso, se deduce la distancia.

A parte de la medida de distancias, los sensores ultrasónicos presentan un amplio campo de aplicaciones. Entre las aplicaciones más usuales, cabe destacar la medida de parámetros para vigilancia y control de procesos industriales, pudiéndose realizar medidas de flujo, temperatura, distancia, niveles, presión, densidad, porosidad, vibración, viscosidad, grosor, composición, tensión mecánica, presencia de grietas y burbujas en materiales o fugas de gas [133]. A diferencia de las aplicaciones anteriormente citadas, en las que se obtiene el valor de un parámetro físico, existen otro tipo de aplicaciones donde se requieren la obtención de *imágenes acústicas*. Entre estos problemas se encuentran las exploraciones biomédicas para caracterizar los tejidos del cuerpo humano (p. ej. la ecografía), y otras industriales como los ensayos no destructivos para el control de calidad, el sonar en aplicaciones marinas, el registro de entornos para la navegación de robot móviles autónomos, etc.

Basándonos en el principio de medida *pulso-eco* existen dos tipos de configuraciones para la captura de imágenes, la basada en un solo transductor que se desplaza para explorar el espacio (configuración monoelemento) y la basada en una matriz de transductores que permanecen fijos e utilizan estrategias de control del tiempo de emisión entre pulsos, para focalizar sobre diferentes zonas del volumen de exploración (configuración multielemento).

En los dispositivos *monoelemento*, el transductor ultrasónico suele tener acoplado un elemento con perfil esférico o cónico que actúa como una lente acústica, obteniéndose una mejor resolución lateral (fig. 2.6). Si el sensor es desplazado linealmente o girado angularmente se obtienen imágenes en formato rectangular o unas secciones circulares, respectivamente. Una característica común de los sistemas monoelemento es que, debido al tiempo invertido en mover el transductor, la captura de la imagen es lenta. Además, son poco flexibles, haciendo prácticamente necesario un diseño específico para cada aplicación. Por otro lado, operando con lentes de foco fijo, solamente se obtiene buena resolución lateral en un rango de distancias pequeño.

La utilización de sistemas *multielemento*, permite superar las limitaciones presentes en los sistemas monoelemento. Una matriz consta de varios transductores ultrasónicos dispuestos geométricamente, de forma que puedan cubrir el volumen de interés. Se pueden simular lentes acústicas de distintos perfiles sin más que intercalar líneas de retardo en sus elementos. Para simular una lente, basta controlar el instante de emisión de los elementos, con lo cual se logra que los pulsos emitidos por cada elemento de la matriz, lleguen simultáneamente al foco. En recepción, existen líneas de retardo programables capaces de controlar la focalización en tiempo real. Una de las características más notables de los sistemas basados en matrices es su flexibilidad, ya que al variar los retardos se pueden obtener condiciones óptimas de resolución según la aplicación. Otra ventaja de estos sistemas es que, hace innecesario el movimiento mecánico para realizar el barrido, con lo cual se incrementa la frecuencia de captura de imágenes. A pesar de estas ventajas, los sistemas de imagen existentes hoy en día suelen usar un único elemento como transductor, por su simplicidad y la no necesidad de altas velocidades. Sin embargo, en el campo de la medicina se utilizan cada vez con más frecuencia los sistemas multielemento, pudiéndose obtener tomografías de órganos como el corazón en tres dimensiones.

En general, las imágenes ultrasónicas se ven degradadas fundamentalmente por un ruido conocido como *moteado (speckle)*, que es comparable al encontrado usando iluminación

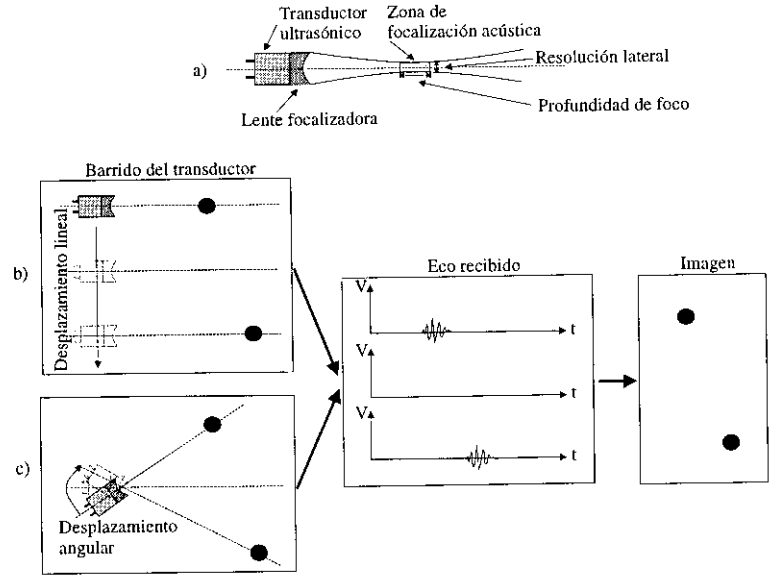


Figura 2.6: Captación de imágenes mediante un sistema ultrasónico monoelemento : a) detalle de la focalización ultrasónica mediante una lente esférica, b) barrido lineal, c) barrido angular.

láser. El *moteado* proviene de las señales ultrasónicas generadas por elementos adyacentes dispersores de las ondas, provocando interferencias destructivas y constructivas de una forma aleatoria en función de la fase relativa con que incidan. Una solución es la combinación de imágenes con diferentes patrones de *moteado* para reducir el ruido. En los casos en los que se obtienen imágenes internas de objetos, la falta de homogeneidad del material distorsiona la imagen resultante. Este efecto se debe a que los ultrasonidos se propagan a diferentes velocidades en función del tipo de medio que atraviesan.

Otros problemas se presentan cuando las superficies sobre las que incide el haz son especulares y se encuentran inclinadas, en este caso, la reflexión forma un ángulo con el eje de emisión y puede ocurrir en función de la anchura del lóbulo, que muy poca energía sea reflejada hacia el detector. Además, por el mismo motivo, la presencia de otros objetos en la proximidad pueden reflejar de nuevo las ondas hacia el detector y provocar medidas falsas. Este efecto de reflexión casi especular en los ultrasonidos, es parecido al comportamiento de una pelota al impactar sobre una pared, y se debe a la alta longitud de onda ($0,68 < \lambda < 8,5$ mm, para frecuencias entre 500 kHz y 40 kHz) de las señales ultrasónicas comparándose con la rugosidad de la superficie de los objetos. En el caso de la luz, este fenómeno no es tan crítico pues su longitud de onda ($0,4 < \lambda < 0,7$ μm) es pequeña respecto a la rugosidad de las superficies y se suele producir reflexión difusa, exceptuando el caso de superficies muy pulidas (fig. 2.7).

Otro aspecto poco favorable de los ultrasonidos consiste en la atenuación que se produce en la amplitud de la señal cuando se transmite por el aire. A medida que aumentamos la frecuencia de vibración, la atenuación aumenta limitando considerablemente el rango

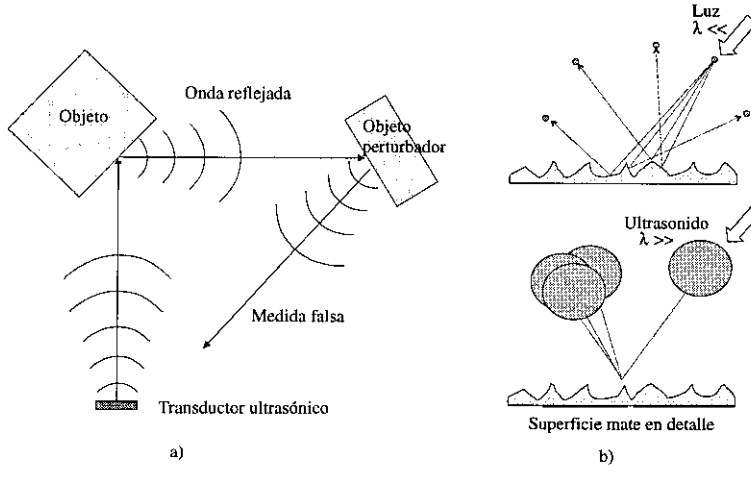


Figura 2.7: a) Reflexión ultrasónica sin retorno directo de señal al sensor y medida falsa por reflexiones múltiples. b) Modelo de la pelota elástica de tamaño igual a su longitud de onda para comprender los fenómenos de reflexión especular en los ultrasonidos y de reflexión difusa en la luz.

de utilización. Resultados experimentales muestran que las máximas distancias de medida en aire están en torno a 10, 3 y 2 metros para frecuencias de 20, 50 y 100 kHz, respectivamente [25].

2.2.3 Palpación táctil

La mayoría de los seres vivos hacen uso de la información táctil. La idea predefinida que tenemos de un sensor táctil, es la de una piel artificial que intenta emular el comportamiento del sentido del tacto de los seres vivos, es decir, el sentido que nos permite caracterizar un objeto mediante un contacto directo de la piel con el objeto. Sin embargo, existen otras modalidades de sensores muy relacionados con la sensación táctil. Estos sensores son los *propioceptivos* o *cinestésicos*, mediante los cuales se conoce la posición y orientación de las articulaciones de por ejemplo un brazo manipulador y el par al que se ven sometidas dichas articulaciones. Los sensores propioceptivos utilizados para captar estas posiciones son de diversos tipos: potenciómetros, codificadores ópticos, transformadores diferenciales variables (LVDT) y sincros. En cuanto a la medida de la fuerza y el par, se puede utilizar la propia corriente que consumen los motores, células de carga, galgas extensométricas, cristales piezoeléctricos y otros [177, 85].

Las *pieles táctiles* o sensores *exteroceptivos*, están formados fundamentalmente de un conjunto de elementos sensoriales básicos dispuestos de tal forma que crean una matriz sensitiva. Mediante estas pieles es posible captar propiedades de los objetos como la temperatura, la forma y textura de su superficie, así como la posición, orientación y el tamaño de estos. Los requisitos deseables para un sensor táctil, exigen resoluciones espaciales de 1 a 2 mm, tamaño de la matriz de entre 10×10 a 20×20 elementos, umbral

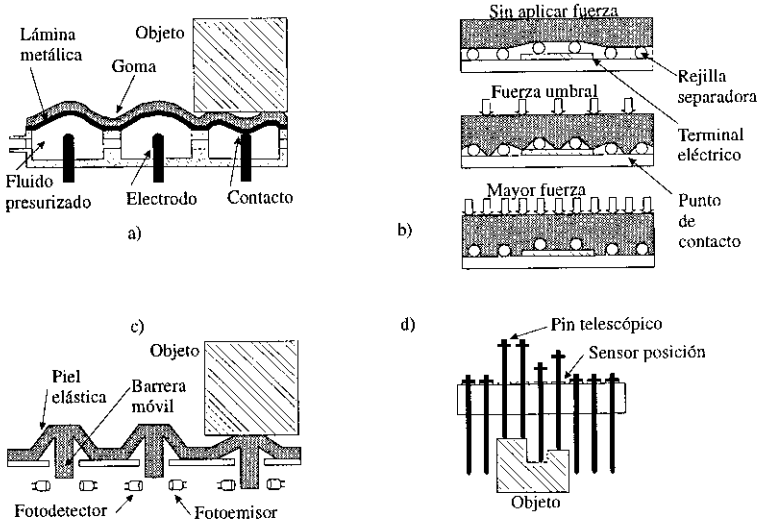


Figura 2.8: Técnicas para la captación de imágenes mediante sensores táctiles: a) Neumática, b) Piezoresistividad, c) Optomecánica, d) Pines retráctiles.

de sensibilidad de 0.005 a 0.1 Newtons, ausencia de histéresis, frecuencia de muestreo por celda de 100 Hz a 1kHz y robustez. Se han utilizado muchas tecnologías para realizar estas matrices táctiles. A continuación indicamos las más representativas [177, 43]:

- *Neumáticos.* El principio de funcionamiento se basa en la deformación de una lámina metálica moldeada al estilo de los conocidos teclados de membrana (fig. 2.8a). Bajo esta lámina metálica existe una cavidad presurizada con un fluido y un electrodo. Cuando se toca un objeto la lámina se deforma y entra en contacto con el electrodo, dando cuenta de su presencia. Este tipo de técnica presenta una serie de dificultades ya que no es posible obtener una buena resolución espacial, hay una fuerte histéresis y la información de salida de cada elemento es pobre, ya que es solo binaria (hay contacto o no).

- *Piezoresistivos.* Los materiales piezoresistivos cambian la resistencia eléctrica cuando varía la presión. Los materiales más utilizados son los elastómeros conductivos, que se fabrican con goma y materiales conductores o semiconductores (p.ej. plata, carbono) [80, 174]. Si se configuran los elementos sensibles del sensor de forma que se produzca un cambio en la superficie de contacto al variar la presión externa (fig. 2.8b), la resistencia de contacto también variará, incluso en mayor escala que debido al efecto piezoresistivo. También se suelen utilizar cordones de elastómero con sección en forma de D y circuitos impresos con pistas cruzadas e individualmente direccionables. Las principales desventajas encontradas son la memoria, histéresis y la dependencia con la temperatura.

- *Piezoeléctricos.* Los cristales de cuarzo producen una carga eléctrica cuando se aplica presión sobre el cristal. Este fenómeno, conocido como piezoelectricidad,

se da en los cristales que no tienen centro de simetría y presentan un dipolo variable con la presión. Existen otros materiales diferentes al cuarzo, polímeros de polivinilo-fluor (PVF), que presentan un efecto piezoeléctrico mayor y que han sido utilizados recientemente para la fabricación de sensores táctiles. Las láminas de PVF son flexibles adaptándose bien a formas sencillas, sin embargo, son sensibles al calor, presentando cierta piroelectricidad que puede interferir la medida de presión [44, 49].

- *Ópticos.* Se caracterizan por utilizar una fuente de luz. Algunos sensores utilizan el principio de reflexión total, que se produce cuando la luz se propaga por un medio más denso que el medio externo que le rodea, y a la vez se cumple, que el ángulo de incidencia es menor que un cierto ángulo crítico. Si un objeto aparece en el medio externo, cesará en esa zona la reflexión total y aparecerá una luz emergente procedente de ese punto [194]. Para capturar esta luz emergente, se puede utilizar una cámara o una matriz de fotodiodos, pero los sensores basados en esta técnica son difíciles de construir. Existen otras versiones optomecánicas que utilizan, por cada elemento de la matriz, un fotodetector y un fotoemisor alineados, cuya transferencia de luz se puede ver interrumpida, por una barrera móvil acoplada a la piel elástica (fig. 2.8c) [173]. Debido a que el elemento flexible suele ser goma, presenta problemas de deriva, memoria, histéresis y dependencia con la temperatura. Otros sensores emplean fibra óptica, y se basan en los diferentes comportamientos en la transmisión de la luz cuando se manipula la fibra óptica. La luz se propaga a través de la fibra óptica con muy pocas pérdidas, pero cuando la superficie es rugosa la luz puede entrar y salir en esa zona. Si dos fibras ópticas pasan cerca una de otra y ambas tienen la superficie rugosa en el punto de cruce, entonces, la luz pasa entre ambas fibras produciéndose un acoplamiento que depende de la separación. También es utilizado el principio según el cual se producen pérdidas en la transmisión de luz cuando una fibra óptica se dobla en algún punto. Otra de las muchas variantes que existen, utilizan dos fibras ópticas (una de entrada de luz y otra de salida) por cada elemento sensible, y una superficie elástica reflectiva. La medida se basa en la detección del grosor de esta piel óptica.

- *Magnéticos.* Algunos sensores táctiles están basados en el efecto Hall y en la magnetorresistencia [123]. Cuando por un material conductor circula una corriente, y en la dirección perpendicular a esta corriente existe un campo magnético, se genera una fuerza sobre las cargas eléctricas que al desplazarse provocan la aparición de un potencial que es conocido como efecto Hall y que es proporcional a la corriente y al campo magnético aplicado. Las cargas eléctricas que viajan a mayor velocidad son desviadas de su ruta, aumentando el camino recorrido, y por tanto, provocando un aumento en la resistencia eléctrica del material. Este efecto es conocido como efecto magnetoresistivo que provoca una variación de la resistencia de un material cuando cambia la intensidad de un campo magnético externo. Este efecto magnetoresistivo tiene aplicaciones directas en la fabricación de sensores táctiles, bien empleando dipolos magnéticos incorporados en pieles elásticas, o bien, generando campos magnéticos mediante hilos conductores.

- *Ultrasónicos.* El principio de medida de distancias *pulso-eco* por ultrasonidos, se puede aplicar en la medida del grosor de superficies. Si utilizamos una piel elástica

que cambia su grosor al entrar en contacto con un objeto, podemos medir el tiempo requerido por el pulso ultrasónico en viajar por la piel, reflejarse y volver de nuevo. Estas variaciones en la distancia nos dan idea del contacto que se está produciendo, pudiéndose medir deformaciones del orden de décimas de milímetro [68].

- *Capacitivos.* Acoplando una serie de electrodos en direcciones transversales y separándolos mediante un medio dieléctrico, se pueden medir pequeños cambios de distancia entre los electrodos, teniendo en cuenta la variación en la capacidad que se produce al separar los electrodos de un condensador [103].

Los sensores que acabamos de revisar se caracterizan por manifestar un acoplamiento pobre entre la superficie de la piel y el objeto con el que entran en contacto, por tanto si existen curvaturas en el objeto, el área de contacto de la piel con dicho objeto es pequeña, siendo adecuadas estas pieles táctiles para objetos con superficies relativamente planas. Como consecuencia, se ha investigado también la realización de sensores táctiles capaces de adaptarse a la forma tridimensional del objeto, los cuales son ventajosos a la hora de realizar reconocimiento de objetos. A continuación se indican brevemente los principios de algunos de estos sensores:

- *Galgas extensométricas.* Se utiliza una ampolla de piel elástica, que tiene su cámara interna rellena de espuma para dotarla de una gran flexibilidad. Las galgas extensométricas de goma se acoplan a la cara interna de la piel elástica, cuya deformación es registrada como una tensión o compresión en las galgas [21].

- *Agujas retráctiles.* Una matriz de agujas retráctiles se utiliza como interfase entre el objeto y un sensor de posición, que mide el desplazamiento de estos tentáculos al entrar en contacto con el objeto (fig. 2.8d).

2.2.4 Técnicas basadas en telemetría láser

El láser es un componente frecuentemente utilizado en sistemas metrológicos, debido a que es una fuente de energía sencilla de generar y de adaptar a diferentes técnicas de medida. Por la importancia de este componente, antes de comenzar relatando las diferentes técnicas de medida con láser, vamos a presentar brevemente la historia, fundamentos y aplicaciones de esta forma de energía.

El primer láser se construyó en 1960 por el físico norteamericano Theodore H. Maiman. Sin embargo, para llegar a este momento histórico, se necesitó la investigación y desarrollo de teorías y experiencias previas por parte de otros muchos investigadores. Todos los estudios parten de la teoría cuántica de Planck y de la teoría de *emisión estimulada* de Einstein que postula que un átomo o una molécula pueden ser estimulados con el fin de que liberen la energía que contienen, bajo la forma de radiación electromagnética. En 1950 el físico Alfred Kastler descubrió el *bombeo óptico* que constituye la base común de los sistemas láser. Sin embargo fueron los científicos norteamericanos Weber y Townes (1953), los que llevaron las teorías de Kastler a la práctica realizando el primer máser (sistema similar al láser pero que genera energía electromagnética en la frecuencia de las microondas). De forma paralela los investigadores moscovitas Basov y Prokhorov

desarrollaron investigaciones en el mismo campo, obteniendo casi al mismo tiempo los primeros máseres de amoniaco.

La palabra *láser* proviene del acrónimo inglés LASER que deriva de "Light Amplification by Stimulated Emission of Radiation", e intenta explicar su funcionamiento. Estos términos indican que el láser es una luz amplificada que se ha generado mediante la emisión estimulada de radiación electromagnética. Dada una sustancia, si la excitamos con energía (flashes, radiofrecuencia, etc.), estamos generando un aumento en la energía de los electrones de los átomos o en los estados de vibración y rotación de las moléculas que la componen. Si la transición energética es adecuada como para alcanzar unos niveles de energía suficientemente altos, se dice que se ha producido un *bombeo óptico*. Desde este nivel energético de bombeo (nivel 3) se producen transiciones naturales a un nivel energético más estable (nivel 2) en el que se produce una acumulación de electrones o moléculas con esa energía. Se dice que se ha producido una *inversión de la población*. En el nivel 2 es poco probable que se produzcan transiciones de forma espontánea a un nivel de mínima energía o nivel 1. Sin embargo cuando se produce alguna, el fotón generado provoca emisiones estimuladas del nivel 2 al 1, las cuales a su vez generan más emisiones estimuladas, generándose una cascada de transiciones estimuladas entre los niveles 2 y 1. Este salto energético genera energía electromagnética de una frecuencia dada por la relación $\nu = (E_2 - E_1)/h$, siendo h la constante de Planck. Por este motivo la luz láser es monocromática, obteniéndose diferentes tipos de frecuencia en función de los saltos energéticos característicos de la materia utilizada o *medio activo* para crear el láser. Así mismo, debido a que la emisión es estimulada y no aleatoria, y a que se encierra el medio activo entre dos espejos paralelos, separados a una distancia múltiplo de la longitud de onda deseada, la luz es coherente, esto es, está en fase, presentando una alta direccionalidad y una muy baja dispersión por lo que la *intensidad radiante* (w/srad) del haz es muy elevada.

En función del medio activo utilizado se encuentran diferentes tipos de láser, agrupándose en tres categorías: sólidos, líquidos y gaseosos. Entre los láseres sólidos se encuentran los de neodimio, rubí (primero que se creó en 1960), cristal y diodos semiconductores. Los gaseosos más representativos son los de CO_2, Helio-Neón y Argón. Otros láseres menos populares son los líquidos como los colorantes y alcohólicos. De los láseres anteriores, los láseres de diodo están teniendo una gran difusión en aplicaciones donde no se requieren potencias elevadas (1mW a 5mW), debido a su bajo coste, su pequeño tamaño y a la calidad del haz que cada vez es mayor.

Asi pues, la radiación láser, que ha revolucionado muchos aspectos de la ciencia y de la tecnología, es una energía electromagnética que se encuentra en las bandas visible, infrarroja o ultravioleta (fig. 2.9), y que se distingue del resto de fuentes de energía en estas mismas bandas, en que la luz láser es *monocromática*, altamente *directiva* y *coherente*, es decir, la luz no interfiere consigo misma mientras no encuentre un medio que provoque una dispersión [138]. La gran direccionalidad del haz láser (divergencia < 1 mrad), posibilita la generación de un haz altamente colimado manteniéndose una alta densidad de energía incluso a grandes distancias. Las propiedades luminosas del láser permiten utilizar lentes, espejos y fibras ópticas para distribuir esta energía, pudiéndose focalizar su luz en un punto generando incluso una fusión nuclear. La coherencia del láser lo hacen ideal para ser utilizado en casos en los que las interferencias jueguen un papel importante. Estas características únicas permiten aplicar el láser a innumerables

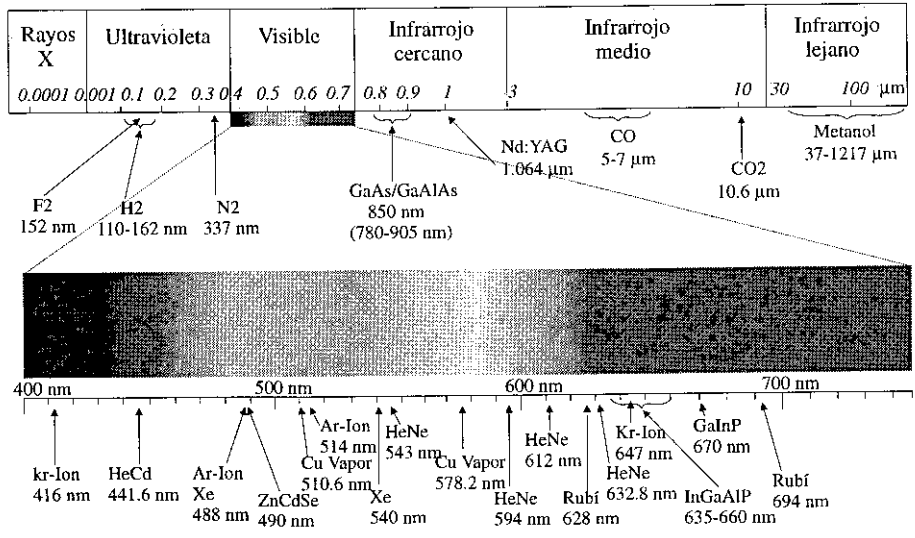


Figura 2.9: Líneas espectrales de emisión de diferentes tipos de láser.

campos:

- *Industria:* Corte, soldadura, endurecimiento y marcado de materiales muy diversos; control de calidad e inspección, guiado de vehículos, seguimiento de blancos, calibración, creación de prototipos a partir de modelos CAD y en la generación de modelos CAD digitalizando moldes.
- *Ofimática:* Impresoras, fotocopiadoras, "scanners", punteros láser, etc.
- *Militares:* Guiado de misiles, técnicas antimisiles, armas con visores láser, visión nocturna, láser radar, etc.
- *Metrología:* Telemetría, alineación, vibraciones, sismógrafos, topografía, cronómetros, medidores de velocidad de blancos, etc.
- *Medicina:* Cirugía, oftalmología, fotocoagulación, operación de cataratas, técnicas de exploración como la endoscopia, tratamiento de tumores, laserterapia, etc.
- *Investigación óptica:* Holografía, interferometría, computadores ópticos, etc.
- *Comunicaciones:* Fibra óptica, redes, CD-ROM, CD-Disc, Video-Disc, Photo-Disc y el próximo soporte digital DVD.
- *Ocio y espectáculos:* Realidad virtual, proyecciones, video juegos, etc.

Una vez vistos los fundamentos del láser, estamos en condiciones de pasar a describir las técnicas utilizadas para medir distancias. A continuación veremos las cuatro técnicas más utilizadas: Triangulación, diferencia de fases, tiempo de vuelo y modulación en frecuencia.

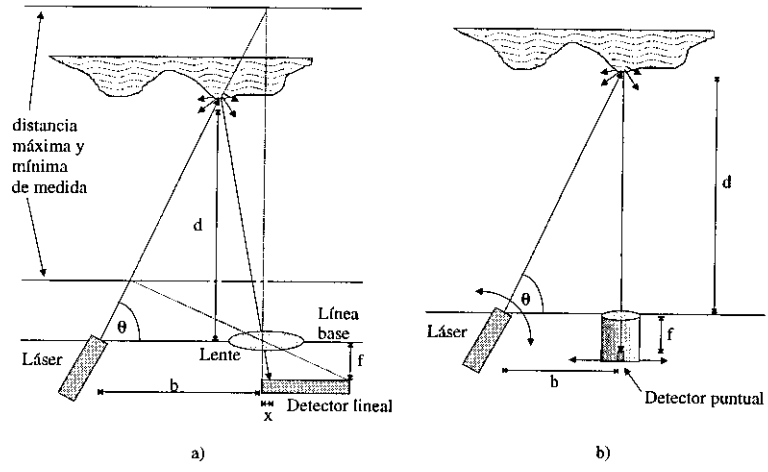


Figura 2.10: Principio de medida de distancias por triangulación: a) utilizando un sensor lineal, b) utilizando un sensor puntual.

2.2.4.1 Telemetría láser mediante triangulación

La técnica consiste en utilizar una fuente de energía láser y un detector de radiación puntual o lineal. Si se conoce el ángulo del eje de emisión del haz láser y la distancia a la que se encuentra el emisor del detector (*línea base*), mediante triangulación se puede deducir la distancia a la que se encuentra un determinado objeto dispersor de la luz. Esta técnica se podría ver como una versión unidimensional de las técnicas de medida de distancias mediante luz estructurada vistas anteriormente.

Como hemos dicho el elemento sensor puede ser lineal o puntual (figs. 2.10a y 2.10b). Dependiendo cual de ellos se utilice se encuentran dos métodos diferentes de triangulación. Cuando se utiliza un sensor lineal, el emisor láser se coloca de forma fija con un ángulo θ respecto a la línea base y el sensor receptor se coloca a una distancia fija b. Como detector lineal suele usarse un sensor analógico de tipo PSD (Position Sensing Device), que genera dos corrientes eléctricas en sus extremos cuyos valores son función del punto de focalización del haz de radiación láser recibido. Midiendo dicho lugar de focalización del haz, x, obtenemos el último parámetro necesario para deducir la distancia d:

$$d = \frac{fb}{x + f/\tan\theta} \tag{2.2}$$

donde f es la focal de la lente (fig. 2.10a).

Utilizando un sensor puntual, es necesario efectuar un barrido angular continuo del haz láser a lo largo del plano definido por la línea base y el eje de emisión, variando el ángulo θ. Además de una forma sincronizada, el detector puntual se debe desplazar linealmente (o angularmente en otras modalidades), hasta que se produce una detección. En este momento, el detector se encuentra enfrentado al punto de impacto (fig. 2.10b) y la

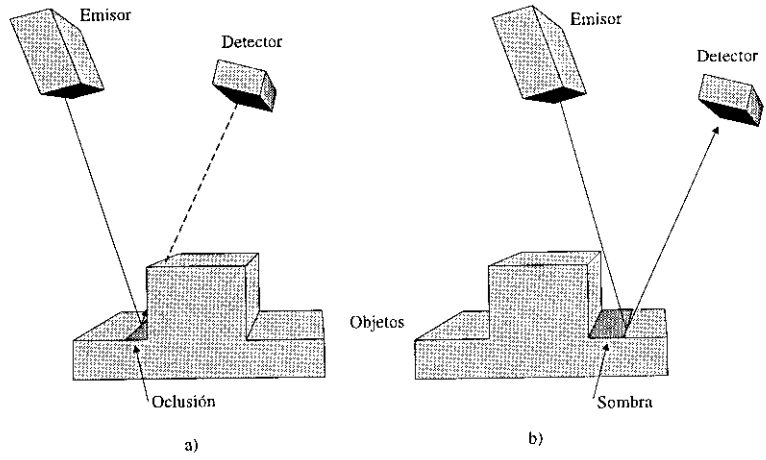


Figura 2.11: Regiones donde no es posible medir la distancia utilizando el principio de triangulación: a) zona sin retorno de la señal (oclusión), b) zona no iluminada (sombra).

distancia se puede calcular de esta forma tan sencilla:

$$d = b \tan \theta \tag{2.3}$$

Cualquiera de las técnicas de triangulación vistas, permiten obtener la distancia absoluta a un blanco, y mediante barridos se pueden obtener imágenes de distancia. Estas imágenes de distancia se obtienen de una forma directa sin necesitar realizar un análisis en la imagen, siendo esta una gran ventaja. Sin embargo, una de las desventajas que presentan es la aparición de sombras y oclusiones debido a que el eje de emisión y el de recepción no son coaxiales (fig. 2.11). Este fenómeno se incrementa cuanto más separados se encuentre el emisor del receptor, aunque es en este caso cuando mejor resolución en profundidad se obtiene. Utilizando separaciones en la línea base pequeñas para reducir el efecto de oclusión y sombras, el rango máximo de medida, con una aceptable resolución entorno a 1 mm, queda limitado entre 0,5 m y 2 m [132]. Además, la precisión en la medida de profundidad no se mantiene constante en todo el rango, siendo más precisas las medidas en las proximidades al sensor y menos precisas cuando medimos a mayores distancias, debido a que los incrementos en x se hacen menos apreciables al aumentar la distancia. Por estos motivos, las técnicas de triangulación se aplican fundamentalmente para medidas de superficies continuas a distancias cortas (100 mm a 400 mm) y con altas resoluciones (0,5-10 micras).

2.2.4.2 Telemetría láser por diferencia de fases

Esta técnica consiste en medir la diferencia de fase entre el haz emitido y el recibido correspondiente a una fuente de luz láser continua modulada en amplitud. En la figura 2.12 se muestra un diagrama de bloques del sistema de medida láser desarrollado en el "Stanford Research Institute" por Nitzan y Duda [155]. Se utiliza un láser modulado a

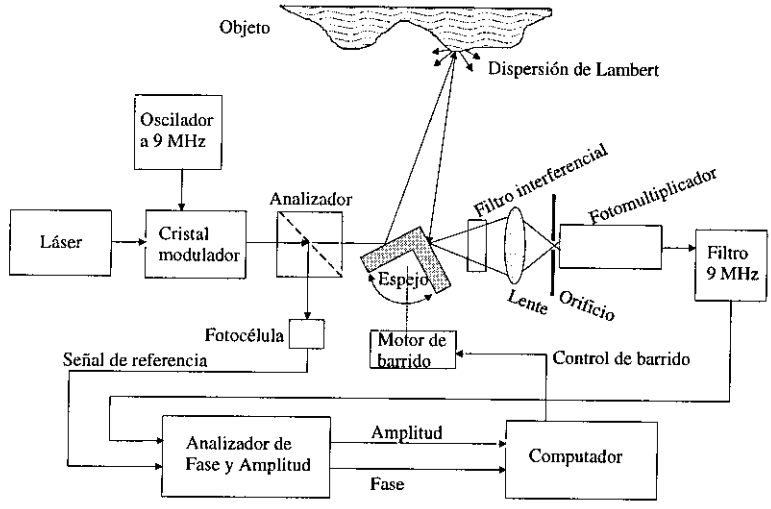


Figura 2.12: Diagrama de bloques de un telémetro por diferencia de fase.

9 MHz, que pasa a través de un divisor de haz o superficie semiespejada que desvía el haz por dos caminos diferentes. El haz que es transmitido a través del divisor de haz, mediante otro espejo es transmitido al medio exterior donde incide sobre la superficie del objeto a medir. Suponiendo una dispersión Lambertiana de la luz, la energía contenida en el ángulo sólido definido entre el punto de dispersión y el área de recepción del sensor, es capturada por el dispositivo de medida para posibilitar el cálculo de la distancia. Mediante un filtro interferencial, una lente y un pequeño orificio, se capta mediante un fotomultiplicador la energía que proviene según el eje de emisión. Este hecho implica que los ejes de emisión y recepción del haz láser van a ser idénticos. La señal eléctrica generada en el fotomultiplicador es filtrada mediante un filtro paso-banda centrado a la frecuencia de modulación de 9 MHz, con el objeto de eliminar la energía provenientes de otras fuentes diferentes a la del telémetro láser. Posteriormente la señal filtrada es comparada con la señal de referencia mediante un analizador de amplitud y fase. La fase detectada es utilizada para calcular de una forma directa la distancia. La amplitud de la señal informa de la atenuación producida en el haz debida a diversos factores como la dispersión de la luz producida al incidir el haz sobre una superficie o la fracción de energía que el sensor es capaz de captar a través de la óptica de recepción.

A diferencia de la técnica de triangulación vista antes, el resto de métodos telemétricos por láser utilizan el mismo eje para transmitir la energía incidente y la reflejada. Esto implica que no se van a producir los problemas de oclusión y sombra explicados anteriormente, lo cual es una gran ventaja. En cuanto a la precisión en la medida, obviamente, depende del rango para el cual esté diseñado el sensor, sin embargo, no ocurre como en el método de triangulación donde la precisión varía dentro del rango de medida.

Una característica común a todos los sistemas telemétricos, es el gran rango dinámico que se puede encontrar en la señal retornada, que puede llegar a ser de 100 dB. La amplitud

de la señal disminuye con el cuadrado de la distancia, depende de la absorción de energía en la superficie del objeto, y además disminuye al aumentar el ángulo entre el eje de emisión y la normal a la superficie. Para evitar los problemas derivados de una señal de baja intensidad se puede aumentar la energía del láser, lo cual puede resultar peligroso en el caso de presencia humana, o se puede realizar un promediado sobre las medidas individuales, lo que implicará una mayor lentitud en la generación de la información.

Para tener un dato cuantitativo, vamos a ver la capacidad de discriminación temporal necesaria para medir en un rango de 2 metros con una resolución de 1 mm, utilizando un láser modulado en amplitud a 10 MHz. Como el periodo de la señal es de $0,1\mu s$ y queremos dividir el rango de 2 metros en 2000 niveles, deducimos que tenemos que discernir diferencias de fase de $\pi/1000$, correspondientes a 50 ps, lo cual unido a la debilidad de la señal retornada, da idea de la complejidad del problema, haciendo imprescindible un procesamiento de integración o promediado de los datos. Esto conduce a tiempos de muestreo relativamente altos (de 0.05 ms a 10 ms), de forma que cuando se quiere obtener una imagen de 200×200 puntos, se requieren tiempos de captura totales entre 2 segundos y 5 minutos, dependiendo del diseño y de la tecnología utilizada [110].

La medición por diferencia de fases puede provocar la obtención de medidas ambiguas. Este hecho se da cuando las señales se desfasan en más de un periodo, no habiendo manera de conocer en cual de los posibles rangos virtuales estamos trabajando, a no ser que se varíe la frecuencia de modulación. Por este motivo, la utilización de estos sensores es recomendable hacerla en entornos donde la máxima distancia esté dentro del rango en el cual no se produzca un desfase mayor de 360^o.

2.2.4.3 Telemetría láser por tiempo de vuelo

En este caso el haz emitido es pulsado y lo que se mide es el tiempo que invierte el pulso de energía en retornar al sensor por una reflexión desde el objeto a medir. En la figura 2.13, se presenta un esquema de un sistema de medida que utiliza este principio [102]. El pulso emitido por el láser incide en la superficie del objeto y a través del mismo eje de emisión retorna al dispositivo de medida, donde la señal luminosa se convierte en eléctrica mediante un fotodetector. Un elemento discriminador genera un pulso digital en el momento que se detecta el eco, siendo esta detección independiente de la magnitud de este eco. Posteriormente un conversor *tiempo/amplitud de pulso*, compara los instantes de emisión y recepción generando un tren de pulsos digitales cuya frecuencia es fija pero cuya amplitud varía con el intervalo temporal detectado. Finalmente se realiza un promediado de la altura de estos pulsos digitales para obtener una medida más precisa y estable.

Las características de este método son bastante semejantes a las de la técnica de diferencia de fase, aunque en el caso de medida por tiempo de vuelo, debido a que la velocidad de transmisión de la luz es muy alta, los tiempos implicados son muy pequeños. Esto complica la medida de rangos pequeños; por ello las aplicaciones más abundantes se centran en mediciones de distancias superiores a varias decenas o cientos de metros. Así, si queremos hacer una medida con una capacidad de discriminación de 1 mm, teniendo en cuenta la velocidad de la luz, debemos ser capaces de discriminar 3 ps. Este hecho limita en gran medida las resoluciones en distancia alcanzables mediante esta técnica de telemetría que típicamente están en torno a 20-30 mm, pudiéndose alcanzar los 5 mm

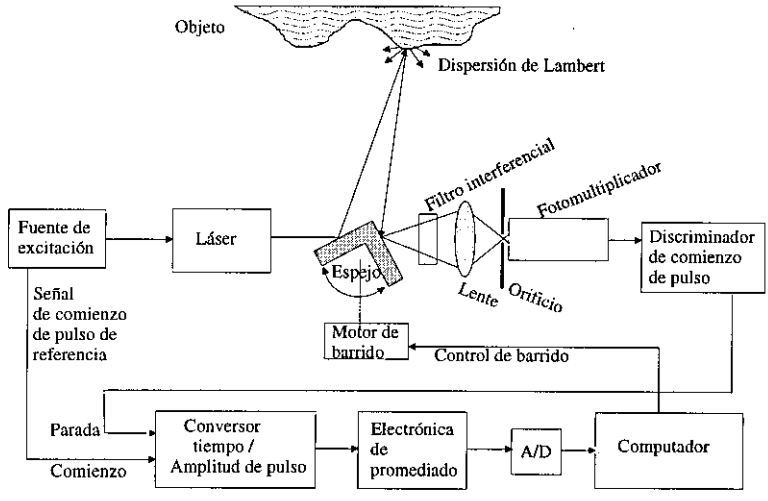


Figura 2.13: Diagrama de bloques de un telémetro por tiempo de vuelo.

con tiempos de integración próximos a 1 segundo [101].

2.2.4.4 Telemetría láser por frecuencia modulada

Esta técnica utiliza una onda electromagnética modulada en frecuencia mediante una señal triangular periódica. Por tanto, la onda es una señal cuya frecuencia varia de forma lineal con el paso del tiempo. Al comparar la señal de referencia con la señal recibida tras incidir sobre la superficie a medir, debido a que existe una diferencia de frecuencia entre ellas, se genera una señal modulada en amplitud a una frecuencia característica que al ser identificada permite calcular la distancia al objeto. La distancia calculada es directamente proporcional a la frecuencia de modulación detectada y es tan precisa como la linealidad en la generación del barrido de frecuencias [50, 67].

La mayoría de los dispositivos que utilizan esta técnica no usan energía láser, sino microondas u ondas milimétricas. Por ello, se necesitan pequeñas antenas parabólicas para su emisión y recepción, los haces son menos direccionales, aunque la capacidad de penetrar en la atmósfera y de lograr largos alcances es superior que en el caso de utilizar energía infrarroja o visible láser. Las aplicaciones típicas de estos sensores están en la medida a largas distancias (>100 m) con precisiones poco exigentes (1 m). Sin embargo, los diodos láser se pueden sintonizar en frecuencia variando su temperatura, lo cual unido al incremento en la linealidad de su control, permite emplean láseres utilizando esta técnica con resultados cada vez más prometedores.

2.3 Análisis, selección y configuración del sistema de medida

2.3.1 Principio de medida: Telemetría láser por diferencia de fases.

Una vez revisadas las principales técnicas para la generación de información de distancia, que en algunos casos también viene acompañada de información de reflectancia, vamos a realizar un análisis global de todas ellas haciendo una estimación del método más adecuado a los objetivos que se plantearon en la sección 2.1.3. En este sentido hemos creado una tabla comparativa (tabla 2.1) donde se presentan todos los métodos vistos anteriormente, analizando las principales condiciones que deben satisfacer. Cada una de las columnas de la tabla representa una *condición*, cuyo cumplimiento indica un aspecto positivo para nuestros objetivos. De esta forma, si representamos por un $\oplus$ la satisfacción de una condición, un método ideal será aquel que cumpla todas las condiciones y se representará mediante una fila completa de signos positivos. Las condiciones no satisfechas, o aspectos negativos, vendrán indicadas por un signo negativo $\ominus$, y las condiciones que sean muy variables y difícilmente evaluables, se representan con un $\odot$.

La tabla 2.1 nos permite clasificar las diferentes técnicas sensoriales en función del grado de adecuación a nuestros objetivos particulares. Los resultados de evaluación obtenidos utilizando tablas con este formato dependen de los objetivos concretos perseguidos y del desarrollo futuro de las tecnologías implicadas. En nuestro caso, podemos observar que no hay ningún método absolutamente perfecto o que cumpla todos los objetivos que nos hemos planteado. Sin embargo, podemos apreciar que la telemetría láser, en particular aquella que opera según el principio de diferencia de fases (AM), es el método más destacado al cumplir la mayoría de las condiciones críticas.

En efecto, este método de medida presenta un buen comportamiento en casi todos los aspectos, siendo sus puntos débiles la relativa lentitud en la exploración de una imagen densa, los riesgos producidos por incidencias directas del haz láser sobre el ojo humano (excepto en la clase I que es inocuo), y el relativo alto coste del equipo. Por lo demás, la técnica es capaz de suministrarnos medidas de distancia absolutas, con resoluciones y precisiones buenas, no viéndose afectada apenas por parámetros externos, siendo por tanto una técnica bastante robusta. La medida se realiza a través del eje de iluminación y la información de distancia es generada de forma directa sin ningún tipo de interpretación posterior. Adicionalmente, muchos de estos dispositivos generan información correspondiente a la atenuación que sufre el haz láser durante el proceso de medida, lo cual permite obtener imágenes de reflectancia más estables que las captadas mediante cámaras CCD, al ser independientes de la iluminación existente sobre la escena. Por todos estos motivos hemos decidido utilizar esta técnica de medida en nuestro trabajo.

Técnicas de captación de distancia y reflectancia	Distancia absoluta	Reflectancia disponible	Rango de medida[1] 1-4 m	Precisión profundidad[1] $\leq \pm 10$ mm	Resolución profundidad[1] ≤ 1 mm	Resolución espacial $\leq 2-3$ mm	Captura imagen (167×167) < 5 seg	Generación distancia sin análisis	Robustez[2]	Ejes captación coaxiales	Inocuo al ojo humano	Coste
Estéreo	⊕	⊕	⊕	⊖	⊖	⊖	⊖	⊖	⊖	⊖	⊕	⊕
Visión dinámica	⊕	⊕	⊕/⊖	⊖/⊕	⊖	⊖	⊖	⊖	⊖	⊖	⊕	⊕
Enfoque	⊕	⊕	⊖	⊕	⊕	⊖	⊕	⊖	⊖	⊕	⊕	⊕
Textura	⊖	⊕	⊕	⊖	⊖	⊖	⊖	⊖	⊖	⊕	⊕	⊕
Gradiente ilum.	⊖	⊕	⊕	⊖	⊙	⊕	⊕	⊖	⊖	⊕	⊕	⊕
Luz estructurada	⊕	⊕	⊖/⊕	⊕/⊖	⊕/⊖	⊕	⊕	⊖	⊙	⊖	⊕	⊕
Moiré	⊖	⊕	⊖	⊖	⊕	⊕	⊕	⊖	⊙	⊖	⊕	⊕
Ultrasonidos	⊕	⊖	⊖/⊕	⊕/⊖	⊕/⊖	⊖	⊖	⊕	⊙	⊕	⊕	⊕
Tel. láser triang.	⊕	⊕	⊖/⊕	⊕/⊖	⊕/⊖	⊕	⊙	⊕	⊕	⊖	⊖	⊖
Tel. láser AM	⊕	⊕	⊕	⊕	⊕	⊕	⊙	⊕	⊕	⊕	⊖	⊖
Tel. láser t. vuelo	⊕	⊕	⊕	⊖	⊖	⊕	⊙	⊕	⊕	⊕	⊖	⊖
Tel. láser FM	⊕	⊕	⊕	⊖	⊖	⊖	⊙	⊕	⊕	⊕	⊖	⊖

[1]Rango, precisión y resolución son parámetros dependientes, por tanto, el cumplimiento de una condición puede forzar a que otras condiciones sean falsas. La aparición de la barra "/" significa que cada línea se debe leer de 2 formas: 1) leyendo los signos a la izquierda de la barra y 2) mirando los signos de la derecha. Por ejemplo, Distancia ⊕ Rango ⊕/⊖ Precisión ⊖/⊕ debe leerse: 1) Distancia ⊕ Rango ⊕ Precisión ⊖, y 2) Distancia ⊕ Rango ⊖ Precisión ⊕.

[2]Robustez ante: Iluminación ambiental, presecia/ausencia texturas, discontinuidades superficies o cambios en temperatura.

Tabla 2.1: Tabla comparativa de diferentes técnicas para la captación de distancia y/o reflectancia, mostrando su grado de adecuación a las especificaciones fijadas como objetivo.

Modelo	Rango medida (m)	Precisión (mm)	Resolución (mm)	Frecuencia max. medida puntual (Hz)	Potencia láser (mW)	Tamaño imagen	Diámetro haz a 2 m (mm)
Acuity 4000-LIR	0.5-15	±5	0.8	50000	8	puntual	3
Acuity 4000-LV	0.5-12	±15	4	50000	5	puntual	3
Sick DME-2000	0.1-2	±5	1	30	3.4	puntual	3
ESP-Tech ORS-1	0.6-4.5	±75	25	2000	2	100 × 1	25
ERIM-ASV	0.6-10	±61	20	32768	–	128 × 128	150
ERIM-ALV [198]	0.6-20	±61	20	32768	–	256 × 64	150
Perceptron LASAR	2-40	±2	–	163840	–	1024× 1024	–
Perceptron 2 [122]	0.6-40	±100	10	131072	–	256 × 256	10
Perceptron 3 [110]	2-4	±4	0.5	147000	–	1000 × 1000	–
Odetics	-10	–	40	13653	50	128 × 128	–
Boulder Elect-Optics	-43	–	170	91750	–	256 × 256	–
Sandia SRI	-30	–	300	524288	20000	256 × 256	–

Tabla 2.2: Telémetros láser: Diferencia de fases (AM).

2.3.2 Telémetro láser: Selección y análisis tecnológico

2.3.2.1 Selección: Acuity Range 4000-LIR

Una vez decidida la técnica de medida a utilizar, debemos proceder a la selección del dispositivo existente más ajustado a las especificaciones dadas en la sección 2.1.3. Hemos realizado una amplia revisión de diferentes medidores de distancia láser, presentando a modo de resumen unas tablas donde se reflejan los diferentes parámetros característicos de estos sistemas de medida. Se han incluido indistintamente tanto sensores puntuales, lineales o regionales, es decir, aquellos que realizan medidas de un punto fijo, con barrido en un eje o que captan un matriz de datos de profundidad al barrer en dos ejes, respectivamente. A pesar de partir con la idea de elegir un dispositivo que opere mediante el principio de diferencia de fases, vamos también a presentar, a modo ilustrativo, modelos que operan con los otros tres principios de medida, de esta forma cuantificaremos el análisis cualitativo realizado en la sección 2.2.

En las tablas 2.2, 2.3, 2.4, 2.5 se presentan los telémetros láser que operan según los principios de diferencia de fases, triangulación, tiempo de vuelo y modulación en frecuencia, respectivamente. Aquellas casillas con un fondo oscuro significan que el dispositivo presenta unas especificaciones que no satisfacen los requisitos establecidos como necesarios para nuestro propósito, y por tanto el modelo deja de ser adecuado. Si alguno de los parámetros no es conocido se indica mediante un guión.

Como era de esperar, los modelos basados en técnicas de triangulación presentan buenas

Modelo	Rango medida (m)	Precisión (mm)	Resolución (mm)	Frecuencia max. medida puntual (Hz)	Potencia láser (mW)	Tamaño imagen	Diámetro haz a 2 m (mm)
MIT-CSD	0.5-3	±150	–	25000	500	puntual	–
Hycan 3D	0.1-0.18	±0.05	0.003	10000	–	800 × Adj.	–
Keyence 3060	0.001-0.06	±0.003	0.0001	50	–	puntual	–
Servo Robot Jupiler	0.5-2.5	–	1.4	20000	100	512 × 1	–
LAP LMS6024	0.3-8	–	1	500	5	puntual	1.5
Monocrom MKLT30	0.5-3	±10	2	1000	5	puntual	3
Hamamatsu H3065	0.35-0.65	–	0.5	4000	–	puntual	15
Zimmer 600/200	0.45-0.65	±0.2	0.05	1000	20	puntual	14
Selcom 2005	1.2-2.2	±1	0.5	16000	10	puntual	4×8
DynaVision SPR02	0.4-4	–	–	667	–	puntual	–
HYMARC [132]	0.5-1.5	±4	–	10000	2000	512 × 512	–

Tabla 2.3: Telémetros láser: Triangulación.

Modelo	Rango medida (m)	Precisión (mm)	Resolución (mm)	Frecuencia max. medida puntual (Hz)	Potencia láser (mW)	Tamaño imagen	Diámetro haz a 2 m (mm)
Riegl LD90-3100HA	1-30	±10	2	2	1	puntual	–
Riegl LD90-210	1-15	±5	1	0.8	1	puntual	–
Leica DISCO	0.2-30	±5	1	0.3	1	puntual	–
SEO LRF-200	1-100	±30	–	3000	–	puntual	–
SEO HDPSS	1.5-75	±15	–	2000	–	puntual	–
SEO Tresense	1-30	±100	–	18000	–	Adj. × 1	–
Odetics 3D	-100	±150	12.5	400000	240	256 × 128	–
ANU Laser [101]	1-4	±3	–	100	–	64 × 64	–

Tabla 2.4: Telémetros láser: Tiempo de vuelo.

Modelo	Rango medida (m)	Precisión (mm)	Resolución (mm)	Frecuencia max. medida puntual (Hz)	Potencia láser (mW)	Tamaño imagen	Diámetro haz a 2 m (mm)
Dig. Optronics Vision	0-1	±12	4	65536	–	256 × 256	–
Dig. Opt. Metrology	0-2.5	–	0.05	10	–	256 × 256	–
Millitech-S	0.5-100	–	500	131072	10	256 × 512	–
Millitech-I	0.2-30	–	50	–	5	puntual	50
Clemens VRSS	0.3-90	±1000	–	200	10	puntual	200
VORAD	0.3-105	±3000	–	30	0.5	puntual	140

Tabla 2.5: Telémetros láser: Frecuencia modulada (FM).

precisiones a costa de unos rangos de medida muy limitados y unas frecuencias de lectura bajas. Las técnicas basadas en tiempo de vuelo consiguen rangos y precisiones elevadas pero a costa de utilizar largos periodos de integración con lo cual las frecuencias de medida son muy bajas. En el caso de la frecuencia modulada, cuando se utiliza radiación milimétrica se consiguen altos alcances con bajas precisiones en profundidad y una pobre resolución espacial al ser el frente de ondas emitido poco puntual. Las técnicas que miden el desfase de la señal modulada en amplitud dan las mejores soluciones, con parámetros más próximos a los requeridos. Se obtienen buenos rangos de medida con altas frecuencias de muestreo utilizando poca energía, y en algunos casos las precisiones son las deseadas. En concreto, de todos los modelos presentados y considerando todas las técnicas, solamente hay un caso en el cual todos los parámetros requeridos se satisfacen. Es el caso del modelo AC4000-LIR de la casa Acuity. Por tanto este es el sensor seleccionado y con él dotaremos al robot Agribot del modo de detección automático, requerido para poder aplicar la *estrategia de detección mixta.*

2.3.2.2 Análisis tecnológico: Índice de eficiencia tecnológica

En el apartado anterior hemos visto que de todos los modelos analizados solamente uno satisfacía nuestros requisitos y no de una forma muy holgada, es decir, algunos parámetros estaban próximos a la frontera que marca el límite entre un método válido y otro que no lo es. En concreto la potencia estaba muy próxima a la máxima admisible, la frecuencia de muestreo máxima (50000 puntos/segundo) permite captar imágenes de 167 × 167 en poco menos de un segundo, y la precisión y resolución de ±5 y 0.8 son muy próximas a las requeridas. Esto hace pensar que existe actualmente una limitación tecnológica que no permite mejorar homogéneamente todos estos parámetros generando sensores más avanzados. Lo más que se puede hacer es mejorar en algún parámetro a consta de perjudicar en otro. Existe un compromiso que hace que si se quiere mejor precisión se necesite mayor tiempo de integración y viceversa. Si queremos mayor rango de medida

perdemos precisión, y si queremos mantener ambas características hay que aumentar la potencia de emisión o de nuevo reducir la frecuencia de medición. Existe por lo expuesto, un compromiso estrecho entre todos los parámetros.

Según el desarrollo que se mostrará en la sección 2.4.2, podemos ver como la relación señal-ruido (SNR) en el fotodetector de un dispositivo de medida depende de diversos factores (ec. 2.22 o 2.4). Entre estos factores están la longitud de onda del láser λ, el periodo de integración T, la potencia o flujo radiante de emisión F_i, la reflectividad de la superficie medida ρ, el ángulo de incidencia sobre la superficie θ, la distancia de medida D, el área de recepción óptica de la señal A y otros parámetros relacionados con la eficiencia en la conversión de la señal electromagnética en eléctrica, que ya se explicarán en la sección 2.4.2 y que ahora no tienen mayor interés.

$$SNR_{(f)} = \sqrt{\frac{\lambda T}{hc}\eta\frac{\beta A F_i}{\pi}\frac{\rho\cos\theta}{D^2}} \tag{2.4}$$

Si aceptamos que la repetitividad en la distancia medida σ_D es inversamente proporcional a la relación señal ruido [155], entonces el compromiso existente entre los parámetros periodo de integración T, distancia de medida D, repetitividad de la medida σ_D y flujo energético incidente emitido por el sensor F_i, se puede expresar de la siguiente forma:

$$\sigma_D \propto \frac{D}{\sqrt{T \cdot F_i}} \tag{2.5}$$

Mediante esta relación de compromiso, podemos hacer múltiples interpretaciones; p.ej. podemos ver que si queremos obtener un sensor con muy buena repetitividad, entonces tenemos que disminuir nuestras pretensiones en rango, aumentar los tiempos de integración o incrementar la potencia de emisión. Hay que destacar que esta relación contiene un signo de proporcionalidad, con lo cual el valor absoluto de los parámetros relacionados va a ser función del valor concreto de la constante de proporcionalidad. Esto indica que dicha constante está relacionada con la calidad de la técnica de medida. Por tanto podemos definir un *índice de eficiencia tecnológica*, M (ec. 2.6), como el factor de proporcionalidad que liga la relación de la ecuación 2.5. Este índice M mejora el índice presentado por Besl [13], que no considera la influencia del término F_i, con lo cual sensores que obtienen buenas especificaciones a base de utilizar altas potencias tienen ventaja sobre los que trabajan con menores potencias, ya que la calidad de aquellos sensores queda sobrevalorada.

$$M = \frac{D}{\sigma_D\sqrt{T \cdot F_i}} \tag{2.6}$$

Utilizando M podemos deducir que una técnica de medida será más eficiente cuando se consigan mejores rangos de medida con buenas repetitividades, pequeños tiempos de integración y se requiera una energía mínima. En este caso el valor de M será grande, y en el caso de tecnologías menos eficientes M tendrá un valor absoluto más reducido.

El compromiso existente entre los parámetros T, D y σ_D, lo podemos ver de una forma gráfica en la figura 2.14. En esta gráfica el origen (0,0,0) del sistema de referencia

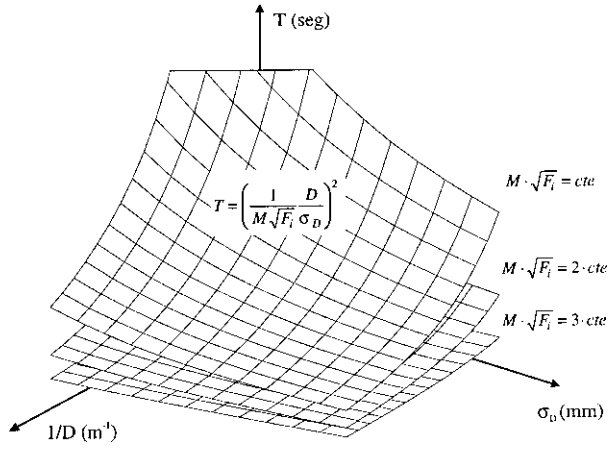


Figura 2.14: Compromiso existente entre algunos de los parámetros que caracterizan a un telémetro láser.

corresponde con un sistema de medida ideal, que tendría un rango de medida infinito, una repetitividad absoluta y frecuencias de muestreo infinitas. Cada una de las superficies se caracterizan por tener un $M \cdot \sqrt{F_i}$ constante. Si este producto aumenta la superficie se aproxima paulatinamente al origen de referencia, lo cual implica que a igualdad de tecnología podemos mejorar las especificaciones simplemente aumentando la potencia de emisión F_i. Sin embargo si la potencia de emisión láser es un parámetro que debe estar acotado, p.ej. por razones de seguridad, la solución para obtener sensores mejores pasa por incrementar la eficiencia M de la técnica.

En la tabla 2.6 se presentan algunos de los modelos comerciales de los que se dispone de suficiente información para calcular su factor M. Para obtener la repetitividad en la distancia σ_D hemos asumido que los medidores están bien calibrados y que la precisión está dada en términos de $\pm 2\sigma$, con lo cual podemos obtener la repetitividad en la medida sin más que dividir el valor absoluto de la precisión entre dos. Los modelos están enumerados siguiendo un orden decreciente en su índice de eficiencia M. Podemos ver que el sensor elegido (AC4000-LIR) no solamente se ajusta a las especificaciones requeridas sino que también utiliza la técnica más eficiente (M=474). Por tanto podemos decir que existe una barrera tecnológica que ninguno de los sensores analizados ha podido superar, que está en torno a $M = 500$. En la figura 2.15 se muestra esa barrera y los vectores de especificaciones, indicados mediante cruces, de algunos de los sensores analizados. Podemos ver igualmente como el volumen de las especificaciones deseadas ($\sigma_D \simeq 1mm$, $T = 1 \times 10^{-4} - 1 \times 10^{-5}$, $\sqrt{F_i}/D = 0.23 - 0.75$ con $F_i = 5$ mW) es seccionado por la superficie $M = 500$, lo cual indica que estas especificaciones son muy exigentes estando parte de ellas fuera del alcance de la tecnología actual.

En un futuro inmediato, si se requieren especificaciones aún más exigentes que las planteadas en nuestro trabajo, habría que mejorar las técnicas de medida o en su defecto, no quedaría más remedio que aumentar la potencia de emisión F_i. Para seguir respetando las normas de seguridad frente a daños en la retina de una persona, este aumento de

#	Modelo	$\sqrt{F_i}/D$	T (seg.)	σ_D (mm)	M
1	Acuity 4000-LIR	0.19	2×10^{-5}	2.5	474
2	Acuity 4000-LV	0.18	2×10^{-5}	7.5	160
3	Selcom 2005	1.6	6×10^{-5}	0.5	160
4	Odetics 3-D	0.15	2.5×10^{-6}	75	54
5	Monocrom MKLT-30	0.75	1×10^{-3}	5	8.5
6	Riegl LD90-3100 HA	0.03	0.5	5	8.5
7	Leica DISCO	0.03	3.3	2.5	8.5
8	Riegl LD90-210	0.06	1.25	2.5	5.3
9	ESP-Technologies ORS-1	0.35	5×10^{-4}	37.5	3.4
10	Sick DME-2000	0.92	0.033	2.5	2.4

Tabla 2.6: Eficiencia tecnológica M de algunos telémetros láser comerciales.

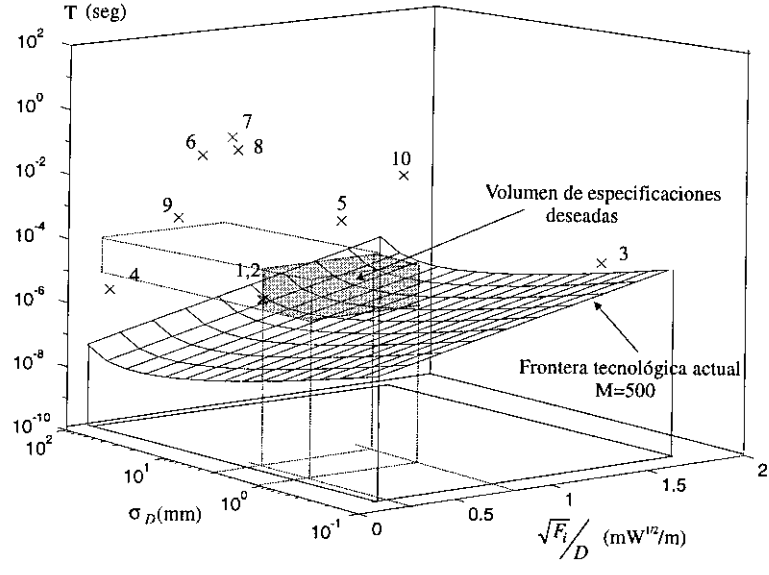


Figura 2.15: Espacio de especificaciones mostrando algunos de los telémetros analizados (representados por aspas y numerados según tabla 2.6), la frontera tecnológica M=500, y el volumen de especificaciones deseado.

energía debería ser acompañado de la sustitución de fuentes láser visibles o en el infrarrojo cercano por láseres trabajando en el infrarrojo medio o lejano. De esta forma, utilizando láseres en estas bandas del infrarrojo, se podrían utilizar potencias bastante más elevadas que las actuales mejorando las especificaciones de los medidores. El motivo se debe a que a partir de longitudes de onda superiores a 1540 nm [132], el agua presente en los tejidos de ojo absorbe la energía que llega a la retina, y además la radiación trasmitida a través del cristalino pierde capacidad de focalización al aumentar la longitud de onda, con lo cual la densidad de energía que incide en la retina es mínima.

2.3.3 Configuración física del sistema de exploración

El telémetro láser AC-4000-LIR es un sistema de medida puntual, por tanto para obtener una imagen necesitamos realizar un barrido en acimut y elevación. La torreta de exploración, utilizada en el robot Agribot para mover el telémetro láser en el modo de señalización asistida, permite realizar este barrido, sin embargo debido a las altas velocidades de barrido requeridas y a las altas inercias generadas al mover todo el sensor, esta solución no es la más adecuada ya que para obtener imágenes de 167×167 puntos se emplea más de un minuto.

Las especificaciones fijadas en la sección 2.1.3, indican que una imagen que capte 50×50 cm, adquiriendo 167×167 puntos, debe tardar menos de 5 segundos. Esto significa que el eje de movimiento más rápido debe realizar 33 barridos en un segundo, o un barrido en 30 ms. Si la distancia media de operación es de 2500 mm, tenemos que la velocidad angular requerida es de 6.7 radianes por segundo. Por el contrario el barrido lento solo requiere una velocidad de 0.04 rad/s. Considerando una resolución espacial de 3 mm/punto, tenemos que a la distancia media de trabajo de 2500 mm, el incremento angular mínimo de nuestro sistema de barrido $d\alpha$ es de 1.2 miliradianes. Con el objeto de satisfacer estas especificaciones hemos optado por una solución que realiza una deflexión del haz láser utilizando un espejo motorizado (fig. 2.16).

2.3.3.1 Componentes del sistema de deflexión

El sistema por deflexión de haz que hemos diseñado tanto para la adquisición de imágenes como para la señalización asistida de frutos, consta básicamente de un espejo motorizado en acimut y elevación, el telémetro láser infrarrojo Acuity-4000-LIR, un diodo láser visible con luz estructurada, una ventana de integración para hacer coaxiales los dos haces láser y unas monturas de ajuste fino para alinear los haces. El motivo para incorporar un sistema de señalización mediante un láser visible es doble; por un lado es necesario para permitir trabajar de una forma asistida con un operario señalando la fruta, y por otro es útil para determinar la posición en torno a la cual se desea captar una imagen. Por estos motivos, y debido a que el telémetro láser trabaja con un haz invisible, se tuvo que incorporar el sistema de señalización mencionado. A continuación presentamos una breve descripción de estos componentes, aunque si se desean más detalles, en el apéndice A presentamos las hojas técnicas de los distintos componentes.

- *Motores.* El motor utilizado para el eje de barrido rápido incorpora una etapa de reducción (1:29.6) precargada, con objeto de eliminar histéresis por holguras,

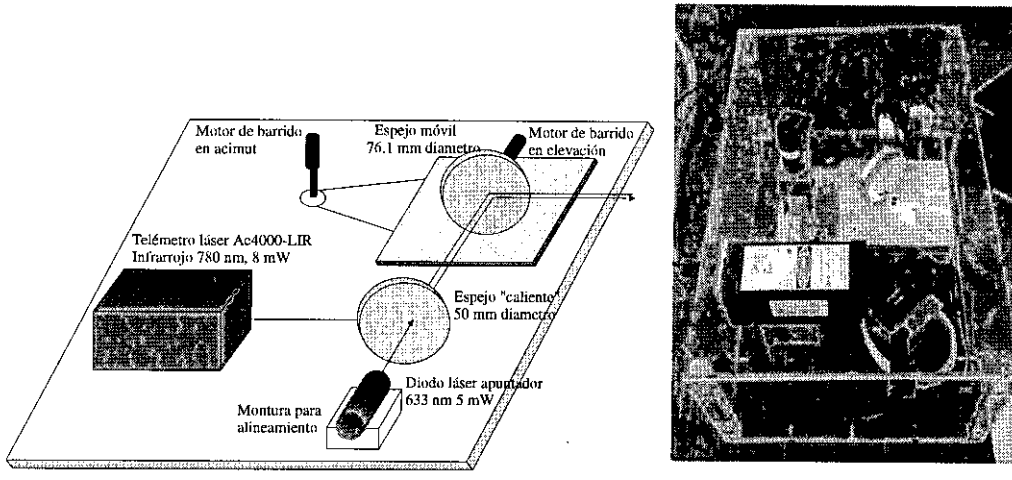


Figura 2.16: Sistema de exploración por deflexión de haz.

un codificador óptico de 1000 pasos por revolución, que permite giros máximos tras la etapa de reducción de 135 rpm, o lo que es lo mismo 14.1 rad/s, generando 5.6 pasos en cada intervalo de barrido angular $d\alpha = 1.2$ mrad. En el otro eje de barrido, el motor que se utilizaba en la antigua torreta de localización del Agribot es adecuado, ya que utilizando un codificador de 500 pasos por revolución y una reducción de 1:751.68 permite una velocidad máxima de 0.7 rad/s.

- *Tarjetas de control de motores y etapas de potencia.* Los dos motores de corriente continua son controlados desde el ordenador mediante una tarjeta basada en el microcontrolador LM629. Las señales de *signo* y *pulso* de baja potencia que generan las tarjetas ante determinadas acciones de control, antes de actuar sobre los motores, pasan por una etapa de potencia que amplifica en voltaje y corriente la señal modulada por ancho de pulso. Ambos componentes electrónicos se desarrollaron en el Instituto de Automática Industrial (IAI-CSIC) constituyendo un trabajo de final de maestría [180].

- *Diodo láser apuntador.* El puntero láser visible que sirve al operador para saber donde está señalando, es un diodo láser rojo de 5 mW a 633 nm. Con esta potencia y longitud de onda, se consigue que el riesgo de daños al ojo humano sea limitado (véase la sección siguiente para más detalles relativos a la seguridad). Debido a que la longitud de onda está próxima a la zona de alta sensibilidad (555 nm), la visibilidad del haz es muy clara en entornos sin luz ambiental intensa, requiriéndose el uso de filtros centrados en 633 nm para poder vez el haz al incidir sobre una superficie cuando se trabaja en el exterior con luz solar. Para facilitar la detección inicial de la posición del haz y el seguimiento de este, hemos estructurado el haz acoplando un generador de cruz, con lo que se mejora su visibilidad, aunque también supone un descenso significativo de su densidad de energía.

- *Espejos.* El espejo deflector que va motorizado en acimut y elevación es un

espejo redondo de 76.2 mm de diámetro con recubrimiento multicapa dieléctrico que proporciona reflectividades superiores al 99% en el infrarrojo cercano y reflectividades mayores del 80% en la banda visible de 633 nm. El espejo utilizado para integrar el haz visible e invisible es redondo y de 50 mm de diámetro, se suele denominar "espejo caliente"(hot mirror) ya que refleja la radiación infrarroja y transmite la luz visible, con pérdidas del 1-2% en la reflexión infrarroja y del 20% en la transmisión del visible.

- *Telémetro y tarjeta de adquisición de alta velocidad.* El telémetro láser es un sistema de medida puntual que puede trabajar a frecuencias de muestreo de hasta 50 kHz, suministrando información de forma periódica de la distancia y de la reflectancia. Esta información se puede leer por un puerto serie RS-232-C o de una forma especial ya que también se transmite por modulación de ancho de pulso. Debido a que necesitamos altas velocidades de lectura, el canal de comunicación serie no es adecuado y utilizamos una tarjeta de alta velocidad que demodula la información de distancia y reflectancia que le llega. Para evitar que se produzcan pérdidas de los datos si el programa de lectura de la información no se activa con la suficiente cadencia, la tarjeta de adquisición de alta velocidad dispone de un buffer de 512 kb con lo cual la integridad de los datos se garantiza.

- *Monturas de alineamiento.* Para asegurar que la medida de distancia del telémetro se corresponda con el punto donde el haz visible este incidiendo, es necesario que ambos haces estén perfectamente alineados en el mismo eje. Para conseguirlo, aparte de los ajustes groseros existentes para la posición del telémetro y del diodo láser, hemos añadido dos ajustes finos en acimut y elevación en el espejo caliente y en el puntero láser utilizando dos monturas de posicionamiento preciso a tornillo.

2.3.3.2 Aspectos de seguridad ante la radiación láser

Según los estándares de seguridad (ANSI, American National Standards Institute y IEC, International Electrotechnical Commision [89]), todo producto que emplee radiación láser debe atender unas mínimas normas de seguridad, relativas fundamentalmente a ciertas especificaciones de diseño y construcción de los equipos, y a la correcta información de peligrosidad en etiquetas o pilotos de advertencia. Según estas normas, y atendiendo a la potencia de emisión de la radiación y su relación con los riesgos al ojo o a la piel humana, los láseres se pueden clasificar en cinco clases o categorías:

- *Clase I.* Totalmente seguros.
- *Clase II.* Potencia menor de 1 mW. Es permisible la exposición directa del haz al ojo durante cortos periodos de tiempo.
- *Clase IIIa.* Potencia menor de 5 mW. La exposición directa produce daños en la retina.
- *Clase IIIb.* Potencias superiores a 5 mW e inferiores a 0.5 W. Tanto la exposición directa como difusa es dañina a la retina y cornea.
- *Clase IV.* Potencias superiores a 0.5 W. Daños en retina, piel y peligro de combustión.

Wavelength λ(nm) \ Exposure time t(s)	$<10^{-9}$	10^{-9} to 10^{-7}	10^{-7} to 10	10 to 10^3	10^3 to 3×10^4
200 to 302.5	3×10^{10} W.m^{-2}	30 J.m^{-2}			
302.5 to 315		C_1 J.m^{-2} ($t<T_1$); C_2 J.m^{-2} ($t>T_1$); $C_1=5600\cdot t^{0.25}$		$C_2\times10^{-3}$ W.m^{-2}; $C_2=1.0^{0.1(\lambda-295)}$	
315 to 400		C_1 J.m^{-2}		10^4 J.m^{-2}	10 W.m^{-2}
400 to 1 400	2×10^{11} W.m^{-2}	200 J.m^{-2}	$1.1\times10^4\ t^{0.25}$ J.m^{-2}	2 000 W.m^{-2}	
1 400 to 10^6	10^{11} W.m^{-2}	100 J.m^{-2}	5 600 $t^{0.25}$ J.m^{-2}	1 000 W.m^{-2}	

Tabla 2.7: MPE para exposición directa de radiación láser sobre la piel [89].

El sistema de exploración que hemos presentado utiliza dos haces láser, uno visible correspondiente al diodo visible de 633 nm con 5 mW de potencia, y otro invisible utilizado por el telémetro Ac-4000-LIR de 780 nm y 8 mW de potencia. Debido a su potencia, estos láseres quedan clasificados en las clases 3a, en el caso del visible, y en la clase 3b para el láser con haz invisible. Simplemente por el hecho de utilizar dos láseres correspondientes a las clases 3a y 3b (aunque este último esta muy próximo a la clase 3a) podemos deducir con total certeza que existen riesgos de daño en la retina ante exposiciones directas del haz sobre el ojo, o bien, a través de reflexiones especulares.

Vamos a tratar de cuantificar estos factores de riesgo de tal forma que podamos ser más explícitos en las precauciones que se deben ser tomadas para poder trabajar con seguridad en el entorno de este sistema de exploración. Estudiaremos tanto el daño a los ojos como el daño sobre la piel ante exposiciones directas y ante reflexiones difusas, viendo cuales son los umbrales de energía de riesgo y deduciendo las protecciones oculares apropiadas para trabajar de forma segura o las distancias mínimas a las que se debe trabajar del equipo láser para no sufrir daños.

Según las normas, entendemos por MPE (Maximum Permissible Exposure) o exposición permisible máxima, el nivel de radiación para el cual, ante circunstancias normales, las personas pueden ser expuestas sin sufrir efectos adversos sobre la piel o la retina. Este concepto suele tener unidades de irradiancia $[w/m^2]$ y está relacionado con la longitud de onda de la radiación, el tiempo de exposición y el tipo de tejido expuesto. En las figuras 2.17 y 2.18 se presentan unas gráficas donde se dan los valores de MPE para diferentes tiempos de exposición y diferentes longitudes de onda, para daños en la retina con exposición directa o difusa respectivamente. La tabla 2.7 muestra MPE para el caso de incidencia directa sobre la piel.

Cuando un láser no es seguro existen normalmente dos soluciones: trabajar siempre a una distancia de seguridad del equipo o utilizar gafas de protección. La *distancia nominal de daño* ó NOHD (Nominal Ocular Hazard Distance), es aquella distancia a la cual la exposición producida sobre el ojo tiene una irradiancia que se ha atenuado suficientemente con la distancia como para que no exista riesgo. Por tanto la determinación de NOHD nos

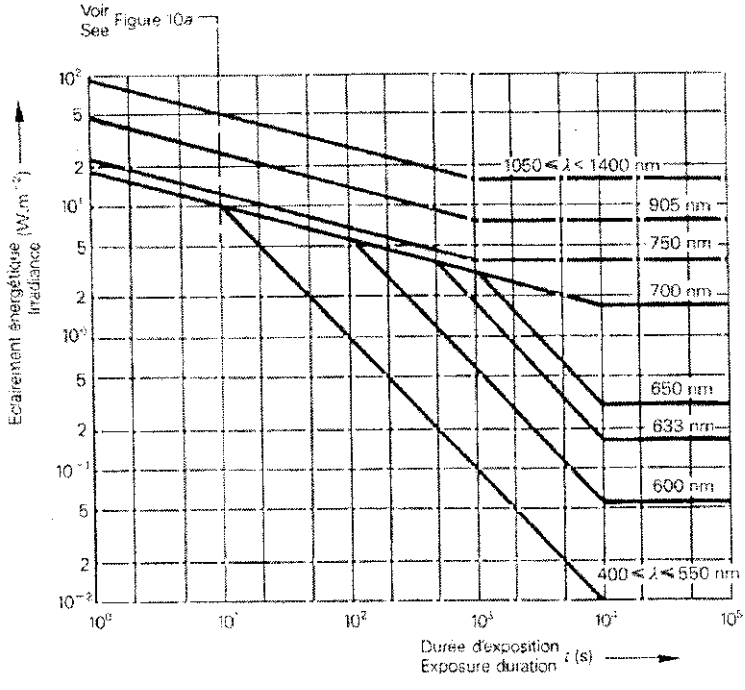


Figura 2.17: MPE para exposición ocular directa con duraciones superiores a 1 segundo para longitudes de onda entre 400 y 1400 nm [89].

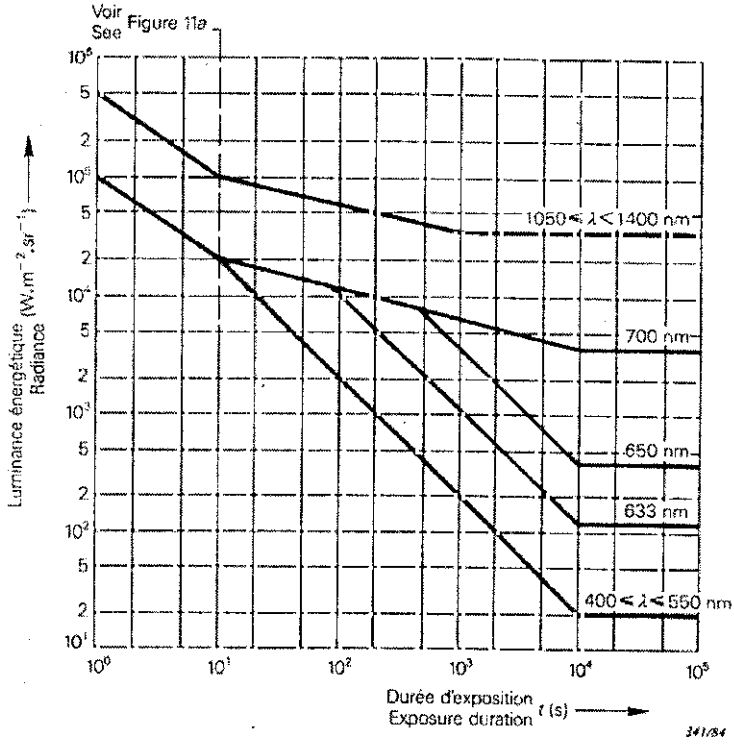


Figura 2.18: MPE para exposición ocular procedente de reflexiones difusas con duraciones superiores a 1 segundo para longitudes de onda entre 400 y 1400 nm [89].

va a marcar la distancia de trabajo a partir de la cual existe seguridad. En el caso de que esta distancia sea muy grande, resultando imposible desde un punto de vista práctico trabajar tan alejado del equipo, la solución pasa por utilizar gafas de protección. La densidad óptica D de un filtro protector se define como el logaritmo decimal negativo del cociente entre el flujo radiante transmitido y el flujo incidente (ec. 2.7). Por tanto, para determinar la densidad óptica de las gafas a utilizar, basta calcular esta relación entre la potencia del láser y la MPE de seguridad que los filtros de las gafas pueden dejar transmitir.

$$D = -\log_{10} \frac{F_t}{F_i} \tag{2.7}$$

Para el caso del láser de 8 mW con una longitud de onda de 780 nm, tenemos que su divergencia α es de 0.5 mrad y el diámetro del haz a es de 3 mm, por tanto la irradiancia a la salida del sensor es de 1132 W/m^2. En el caso de exposición sobre el ojo y piel de forma directa o difusa, tenemos los siguientes resultados:

1. *Exposición directa al ojo de 100 segundos.*
   - *MPE.* Según la figura 2.17 tenemos que MPE=6 W/m^2.
   - *NOHD.* La distancia nominal de daño se calcula aplicando $NODH = \frac{\sqrt{(4F_i)/(\pi MPE)}-a}{\alpha}$. En nuestro caso esta es de 76.4 metros, con lo cual se requiere el uso de gafas para evitar daños por posibles reflexiones especulares o exposiciones directas al haz por accidente.
   - *Densidad óptica de las gafas.* La gafas recomendadas deberán tener una densidad óptica de 2 o superior, ya que: $D = -\log_{10} \frac{6}{(48\cdot 10^{-3})/(\pi a^2)} = 2.27$
2. *Exposición difusa al ojo de 100 segundos.*
   - *MPE.* Según la figura 2.18 tenemos que MPE=10000 $W/m^2 sr^2$.
   - *NOHD.* Es de 0.57 m, con lo cual se requiere el uso de gafas para reflexiones difusas si se trabaja a menos de 57 centímetros del punto de impacto del haz.
   - *Densidad óptica de las gafas.* La gafas recomendadas anteriormente serán adecuadas también para este caso.
3. *Exposición directa sobre la piel de 100 segundos.*
   - *MPE.* Según la tabla 2.7 tenemos que MPE=2000 W/m^2.
   - *NOHD.* Es de -1.48 m, lo cual indica que no existe riesgo de daños en la piel, y por tanto no hay necesidad de utilizar ninguna prenda o cubierta protectora.

En el caso del láser visible la potencia emitida es ligeramente inferior al caso anterior, 5 mW, y la divergencia y el diámetro del haz a la salida del diodo son iguales (707 W/m^2 a la salida del diodo). Como consecuencia los valores de MPE y NOHD obtenidos para el diodo visible son muy similares a los presentados anteriormente, con lo cual se debería recomendar el uso de gafas de densidad óptica 2 o superior para longitudes de onda de

633 nm. Sin embargo debido a que existe un generador de cruz que reduce la densidad de energía por un factor variable con la distancia, que para 2 metros está en torno a 100, no es necesario el uso de estas gafas para reflexiones difusas y solamente podría existir un riesgo ante incidencias directas. Esta última circunstancia también queda minimizada ya que la respuesta temporal, ante un destello de haz visible, en cerrar la pupila y girar la cabeza (*respuesta de aversión*) es de 0.25 segundos, con lo cual la exposición no es lo suficientemente prolongada como para producir daños ya que el MPE de respuesta de aversión es de 25 W/m^2 (fig. 2.17) y la densidad de potencia del haz con la cruz generada a 2 metros es de 707/100=7 W/m^2.

Por tanto, para trabajar de forma segura con este sistema de exploración no se necesita ninguna protección ante el haz visible, y se puede trabajar igualmente sin protección para el haz infrarrojo si se producen siempre reflexiones difusas estando a más de 60 centímetros de distancia. Como esta circunstancia no tiene porque cumplirse siempre, ya que bien debido a la presencia de superficies especulares o bien a que por accidente podemos mirar directamente el haz infrarrojo, se recomienda el uso de gafas de protección (D=2, $\lambda = 780$ nm) para trabajar de forma segura y no confiar en que las circunstancias poco probables anteriormente citadas no se vayan a dar nunca.

2.3.3.3 Adquisición de las imágenes

El sistema de exploración desarrollado, gracias a que el telémetro suministra tanto información de distancia, D, como de reflectancia, $\Re$, permite obtener dos imágenes por cada escena captada (fig. 2.19). Las imágenes en distancia también se conocen como mapas de distancia, imágenes o mapas de profundidades, imágenes 3-D ó 2.5-D o mapas de alturas superficiales. Si las medidas se referencian respecto a un sistema de coordenadas ortogonal se dice que la imagen de distancias es de la forma XYZ. A diferencia, si la distancia medida indica la profundidad a lo largo de un vector cuya dirección queda definida por dos índices, se dice que la imagen de distancias tiene la forma $D_{i,j}$. Nuestro caso es este último y los índices son las coordenadas angulares en acimut y elevación. Este formato provoca una cierta deformación de la imagen que consiste en que las superficies planas aparecen ligeramente curvadas, especialmente si el ángulo de barrido es grande. Sin embargo el formato $D_{i,j}$, en general, permite un procesamiento más sencillo que sobre imágenes en formato XYZ ya que en éste caso no existe un claro ordenamiento de los puntos a la hora de procesarlos [13].

Para asegurar que tanto en las imágenes $D_{i,j}$ como $\Re_{i,j}$ exista una correspondencia entre los índices (i,j) que indican los ángulos de acimut y elevación, y la distancia D o reflectancia $\Re$ medidas, se han ensayado dos modos de sincronización entre las lecturas que el telémetro suministra de forma periódica y los ángulos que hay que leer de los codificadores de los motores. Estos modos los denominados síncrono y asíncrono.

En el caso síncrono mientras se realiza cada uno de los barridos verticales de los que se compone la imagen, se va supervisando la posición angular actual. En el momento que los motores están en la posición adecuada almacenamos en las coordenadas (i,j) de los mapas $D_{i,j}$ y $\Re_{i,j}$, los valores de distancia y reflectancia que en ese momento suministra el telémetro. El perfil trapezoidal de velocidad que sigue cada barrido, con sus consiguiente aceleración inicial y deceleración final, no influye en el proceso de adquisición pues solamente se almacenan datos en el momento que se comprueba que estamos midiendo

a) b) c)

Figura 2.19: Escena de un árbol con cuatro frutos: fotografía natural (a), y las correspondientes imágenes de distancia (b) y reflectancia (c) captadas mediante el sistema de exploración.

en el lugar deseado (fig. 2.20a). Mediante este método de sincronización los ángulos de inicio y fin de barrido en elevación se corresponden con los extremos superior e inferior de la imagen a captar, ya que el proceso de medida puede realizarse incluso en las zonas de aceleración y deceleración.

En el método asíncrono, el almacenamiento de los datos de distancia y reflectancia se hace sin consultar la posición angular actual. El barrido vertical se hace realizando un rastreo mayor que la altura de la imagen, de tal forma que las regiones exploradas mientras se está en zona de aceleración/deceleración no se contemplan y solamente almacenamos los datos captados mientras el movimiento se realiza con velocidad uniforme (fig. 2.20b). Como el telémetro genera la información a intervalos regulares y el movimiento es uniforme, solamente con realizar una sincronización al inicio de cada barrido es suficiente. Este modo de barrido tiene sentido debido a que el modo síncrono requiere una carga extra de trabajo al necesitar consultar constantemente la posición angular del motor de barrido vertical. Además, las tarjetas de control que estamos utilizando (basadas en LM629) no permiten consultas a más de 500 μs con lo cual trabajar a más de 2 kHz en el modo síncrono se hace imposible.

Cualquiera de los dos modos de barridos son válidos a la hora de generar imágenes, sin embargo si se pretende trabajar a menos de 2 kHz el método que recomendamos es el síncrono, pues requiere menor exploración, y para más de 2 kHz hay que utilizar el modo asíncrono que permite realizar barridos más rápidos aunque también necesita mayores exploraciones.

2.3.3.4 Limitaciones encontradas

En principio con el sistema presentado se podría trabajar en el modo asíncrono a 5.6 kHz obteniendo imágenes de 167 × 167 puntos en 5 segundos, que era nuestro objetivo. Sin embargo, existe un factor que limita actualmente nuestro sistema, nos referimos a las inercias existentes en los cambios de dirección del barrido. Estas inercias hacen que las aceleraciones/deceleraciones máximas alcanzables sean de 200 rad/s^2, lo que implica

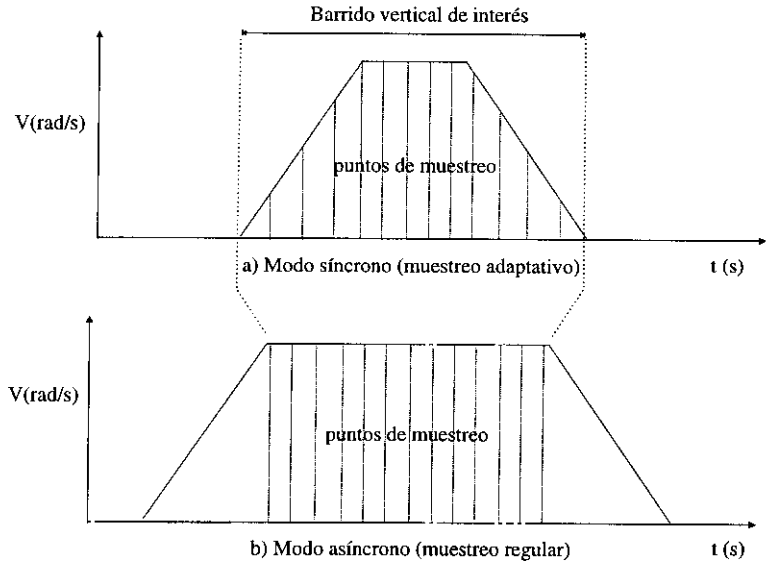


Figura 2.20: Modos de sincronización en el barrido vertical.

que para alcanzar una velocidad de 6.7 rad/s en el eje de elevación se requieren 33 ms de aceleración y otro tanto de deceleración. Realizando 167 aceleraciones y deceleraciones en cada imagen tenemos un tiempo de 11.2 segundos solamente dedicados a esta tarea. Si además sumamos a los 11.2 segundos el tiempo que se invierte en el barrido con movimiento uniforme que es de 5 segundos en el modo asíncrono, tenemos tiempos totales de 16.2 s para captar una imagen de 500 × 500 mm con una resolución de 3 mm/punto que genera 167 × 167 puntos.

En la práctica los tiempos experimentales rondan los 20 segundos ya que también hay que considerar el periodo de estabilización de la posición al comienzo de cada barrido. Somos conscientes de que este sistema de exploración no es adecuado para una operación en tiempo real ya que se queda lejos de lo deseado, sin embargo nos permite continuar estudiando la viabilidad de las técnicas láser en entornos agrícolas y además constituye una muy buena base experimental para plantearse mejoras futuras, las cuales deberán ir encaminadas hacia la reducción de las inercias o al empleo de sistemas rotatorios que no supongan tener que variar la velocidad.

En el apartado referente al nivel de señal con el que el telémetro realiza el cálculo de la distancia, podemos decir que es un aspecto bastante crítico, ya que a las pérdidas de señal que se producen en el proceso de medida hay que añadir las pérdidas introducidas al guiar el haz láser por medio del espejo móvil y el "espejo caliente". Una selección de componentes con reflectividades no próximas al 100% para la longitud de onda de 780 nm, puede hacer que las pérdidas de señal acumuladas en las cuatro reflexiones sobre estos componentes, atenúe la señal de tal forma que si medimos sobre objetos de baja reflectividad, la reflectancia final captada sea baja y las medidas de distancia sean muy poco fiables.

En nuestro caso los componentes ópticos tienen reflectividades del 99.5% en el espejo móvil y del 98% en el caliente. Esto hace que considerando las reflexiones en el camino de ida y vuelta tengamos una pérdidas totales del 5% de la energía. A esto hay que añadir la sección del espejo caliente que es de 50 mm de diámetro, debido a que la óptica de recepción del telémetro es de 60 mm de diámetro y el espejo está dispuesto a 45^o, existe una pérdida adicional de un 10%, con lo cual la energía total perdida con el sistema de deflexión de haz presentado es del 15%, frente al sistema operando directamente sobre el blanco. Este factor, con estos niveles de pérdidas no es muy significativo, ya que solo supone una reducción uniforme en los valores del mapa de reflectancia y un aumento del nivel de ruido en la imagen de distancias como consecuencia de la menor señal captada.

2.4 Caracterización del telémetro láser

En esta sección nos planteamos la caracterización del telémetro láser Ac-4000-LIR, con el fin de conocer otras características adicionales a las suministradas por el fabricante en las especificaciones del aparato. En especial estamos interesados en identificar cuales son los factores que influyen en el valor de la reflectancia que proporciona el sensor como salida, y ver de que forma este valor influye en la calidad de las medidas de distancia. Los dos apartados siguientes presentan y desarrollan ambos aspectos, indicando también la utilidad de sendos estudios.

2.4.1 Modelado de los factores que afectan a la reflectancia registrada

La imagen de distancia, obtenida mediante el sistema de exploración presentado en la sección anterior, tiene una directa interpretación puesto que sus gradientes dan idea de la forma de los objetos. Sin embargo la imagen de reflectancia no tiene una inmediata significación puesto que los valores puntuales de reflectancia dependen de diversos factores. La correcta interpretación del parámetro de reflectancia va a permitir predecir como será la distribución de una región del mapa de reflectancia cuando asumamos la presencia de un objeto determinado, o lo que es lo mismo, una determinada distribución en el mapa de distancia. Este hecho hace posible que podamos utilizar tanto la información de distancia como la de reflectancia en el análisis de formas, así como la obtención de propiedades ópticas, aspectos ambos que mostraremos en el capítulo 4. Al integrar ambas fuentes de información en las etapas de procesamiento de las imágenes, estaremos facilitando la obtención de resultados más fiables y útiles a la hora de realizar un proceso de reconocimiento. Es por ello que para poder llevar a cabo esta integración, será indispensable un perfecto conocimiento del significado y de como están relacionados los parámetros distancia y reflectancia.

En esta subsección vamos a presentar la caracterización del telémetro láser Ac-4000-LIR, modelando matemáticamente la reflectancia registrada por el sensor como una función que depende de diversos factores que intervienen en el proceso de medida. Este modelo se basa en principios genéricos, por tanto no se limita a este dispositivo concreto y puede ser aplicado a diversos sistemas telemétricos operando con los mismos principios de medida.

2.4.1.1 Planteamiento inicial del modelo

El telémetro láser genera un haz de energía electromagnética que al incidir sobre una superficie se ve afectado por fenómenos de atenuación, reflexión y dispersión. Dependiendo del tipo de superficie, de su orientación y de su posición espacial, la amplitud de la radiación recibida por el sensor varía. Los principios generales en los que se basan estos fenómenos son conocidos [9, 155, 36, 84, 193, 66], sin embargo, también será necesario modelar las variables específicas de nuestro telémetro para poder cuantificar todos los fenómenos a los que se ve sometida la señal final de reflectancia, desde que se emite el haz láser hasta que ya en forma de corriente eléctrica se amplifica y se almacena.

Inicialmente vamos a suponer que la superficie sobre la que incide la radiación es un difusor perfecto, o superficie Lambertiana. Este tipo de superficies se caracteriza porque la radiancia, o luminancia en el caso del ojo humano, es constante independientemente del ángulo desde donde se observe. Una superficie que tiene características muy próximas a las de un difusor perfecto es el sulfato de Bario ($BaSO_4$), aunque cualquier otra superficie de las denominadas "mate" (papel, madera, escayola,...) presenta características próximas a las de un difusor perfecto. Debido a que la luminancia es constante, si la iluminación es fija, al observar estas superficies desde ángulos distintos siempre tendremos la misma sensación de claridad (luminancia constante $[w/m^2sr]$). Sin embargo si nos olvidamos del parámetro que capta el ojo humano o sensores como cámaras (luminancia o radiancia, respectivamente), y nos referimos a la intensidad radiante o intensidad luminosa $[w/sr]$, la distribución de este parámetro en un difusor perfecto no es uniforme y obedece a una función que depende del coseno del ángulo de observación. Este hecho es conocido como ley de Lambert, y se expresa así:

$$I = I_{\theta=0} \cos\theta \tag{2.8}$$

donde $I_{\theta=0}$ representa la intensidad máxima que se emite en la dirección perpendicular a la superficie. En la figura 2.21 podemos ver los perfiles en intensidad y radiancia correspondientes a una superficie difusora perfecta.

Si expresamos $I_{\theta=0}$ en términos del flujo radiante incidente $F_i[w]$, y del coeficiente de reflexión o reflectividad difusa ρ, podemos reformular la ley de Lambert como:

$$I = \frac{F_i}{\pi}\rho \cos\theta \tag{2.9}$$

donde el término constante π tiene dimensiones de estereoradian para que las unidades a ambos lados de la expresión sean las de una intensidad radiante $[w/sr]$.

La reflectividad difusa, ρ, es la relación entre el total del flujo radiante difuso reflejado y el total del flujo radiante incidente F_i. Este coeficiente depende de la longitud de onda utilizada y del tipo de superficie, sin embargo en nuestro caso al utilizar una onda monocromática, ρ solo dependerá del tipo de superficie. Si se utiliza un láser visible ρ variará desde 1, para una superficie idealmente blanca o que refleja todos los colores, hasta 0, para una superficie de color negro o que absorbe todos los colores. En cuanto al flujo radiante, F_i, utilizando el telémetro Ac-4000-LIR a la máxima potencia será de 8 mW.

Teniendo en cuenta que la superficie se encuentra a una distancia D del telémetro,

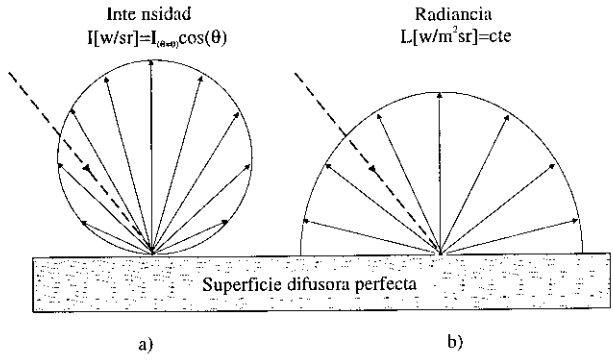


Figura 2.21: Distribución de intensidad (a) y radiancia (b) de una superficie difusora perfecta.

que el área de la óptica de recepción del telémetro, A, es un círculo de radio 30 mm ($A = \pi \cdot 30^2 = 2827mm^2$), que la óptica interna atenúa la señal por un factor constante β, y que el ángulo sólido subtendido por el receptor es $\Omega = A/D^2$, tenemos que el flujo radiante capturado por los sensores fotoeléctricos del propio telémetro, F_c $[w]$, viene dado por la siguiente expresión:

$$F_c = \beta I \Omega = (\frac{\beta A F_i}{\pi}) \frac{\rho \cos\theta}{D^2} \tag{2.10}$$

Debido a que la energía electromagnética es transformada en señal eléctrica mediante un sensor fotoeléctrico y posteriormente se realiza una amplificación logarítmica de esta señal, finalmente podemos modelar la amplitud de la señal recibida, o reflectancia, por

$$\Re = cte_1 \log(cte \cdot F_c) = cte_1 \log(cte_2 \frac{AF_i}{\pi} \frac{\rho \cos\theta}{D^2}) \tag{2.11}$$

donde cte_1 representa el factor de amplificación y cte_2 representa varios factores: el factor de atenuación óptica β, el factor de conversión y la sensibilidad del fotodiodo η, y la preamplificación en la señal eléctrica.

Tanto cte_1 como cte_2 son desconocidos, por tanto para determinarlos se realizó la recogida de los datos experimentales que se muestran en el apéndice B, para una superficie difusora casi perfecta (superficie mate), y posteriormente se realizó un ajuste por mínimos cuadrados. Realizando tantos ajustes como diferentes distancias ensayadas, se pudo comprobar que cte_1 tenía un valor medio en torno a 80 con unas variaciones máximas del 6%, mientras que cte_2 tenía una valor promedio de 5.8×10^7 con variaciones del 200%.

Analizando la figura 2.22a, se puede apreciar cómo el modelo planteado, ante cambios en θ, se ajusta bastante bien a los datos experimentales, con lo que se confirma que para esta superficie casi difusora perfecta la ley de Lambert es un buen modelo. Analizando la figura 2.22b se puede observar que los datos experimentales, ante cambios en D, no se ajustan al modelo planteado y se produce una clara discrepancia. Esto significa que

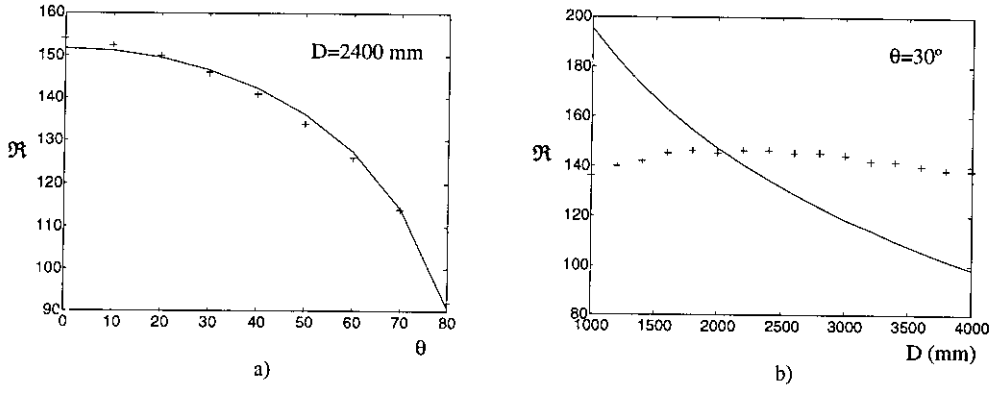


Figura 2.22: Adaptación del modelo inicial expresado en la ec. 2.11 (trazo continuo) a los datos experimentales ('+') ante variaciones en θ (a) y ante variaciones en D (b).

la señal no se atenúa según el cuadrado de la distancia. Como esta ley evidentemente se cumple, se deduce que debe existir otro factor que varíe con la distancia y que provoque el comportamiento global observado. Es por tanto necesario considerar el término cte_2, que por lo visto en el ajuste inicial tenía una gran dispersión (200%), como un factor no constante y que depende de la distancia D. Por tanto cte_2 será reemplazado por una función $f(D)$ que depende de la distancia, que al incorporarlo a la ecuación 2.11 genera el siguiente modelo corregido:

$$\Re = c_1 \log((aD^3 + bD^2 + cD + d)\frac{AF_i}{\pi}\frac{\rho\cos\theta}{D^2}) \tag{2.12}$$

donde $f(D)$ lo hemos desarrollado como un polinomio de orden tres de tal forma que obtengamos sus coeficientes de forma empírica, sin necesidad de realizar un desarrollo analítico de su comportamiento.

El telémetro no dispone de ningún tipo de compensación o amplificación variable en función de la distancia D, por tanto la causa de esta dependencia añadida de la señal con la distancia lo atribuimos fundamentalmente a un fenómeno de reducción de la señal luminosa transmitida a través de la óptica de recepción cuando la distancia al sensor disminuye. Este hecho puede ser debido a la disminución de la capacidad de focalización de la lente cuando la distancia disminuye y a la presencia del orificio previo al fotodiodo que limita la energía que no fue focalizada correctamente.

Realizando un ajuste por mínimos cuadrados a los datos experimentales, obtenemos los coeficientes a, b, c y d (-4.10×10^{-3}, 34.25, -2.71×10^4 y 7.98×10^6 respectivamente). El término cte_1 lo hemos considerado temporalmente igual a 80, por ser una estimación bastante estable de la amplificación logarítmica utilizada en el sensor y por no influir significativamente su elección arbitraria en la determinación de $f(D)$.

En la figura 2.23a, podemos ver la representación gráfica de $f(D)$, apreciándose una actuación casi lineal con un punto de inflexión en torno a 2500 mm que corresponde a la distancia de máxima sensibilidad del sensor. La gráfica de la derecha (fig. 2.23b)

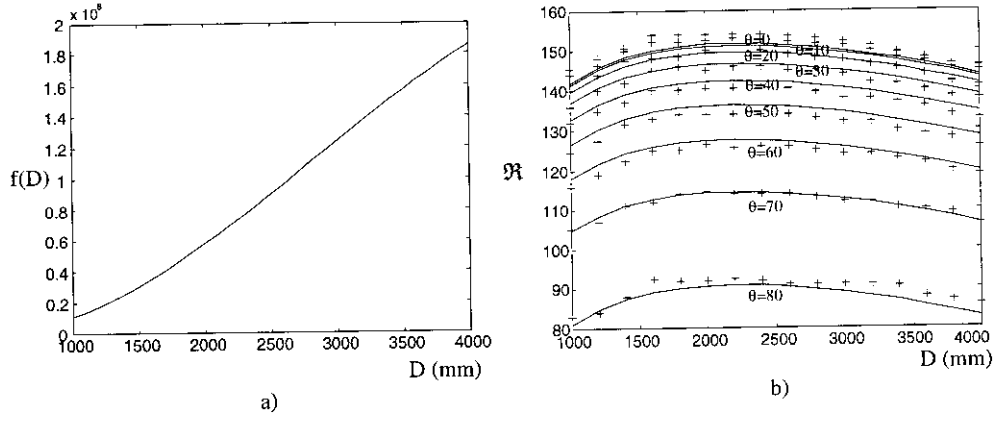


Figura 2.23: a) Representación de la función empírica $f(r)$. b) Adaptación de los datos experimentales del apéndice B al modelo corregido (ec.2.12).

muestra la adaptación del modelo corregido planteado a todos los datos experimentales incluyendo variaciones en θ, pudiéndose apreciar una clara mejoría en el ajuste.

2.4.1.2 Modelo general: Superficies no ideales

Observando de nuevo la figura 2.23b podemos ver que el modelo define reflectancias similares para el caso de incidencias con un ángulo $\theta = 0^o$ que igual a 10^o. Sin embargo los datos experimentales no manifiestan este comportamiento, existiendo una diferencia media de una unidad en la reflectancia entre ambas situaciones. Atribuimos este efecto a que la superficie elegida no es una superficie difusora perfecta y existe un pequeño porcentaje de reflexión difusa que se propaga siguiendo el eje de reflexión especular.

Existen trabajos previos [84, 36, 7] en los cuales se ha tratado de modelar el comportamiento difuso de la luz al incidir sobre una superficie. Cuando la luz incide sobre una superficie, parte se absorbe, parte se transmite y el resto se refleja en un grado que depende de cual sea el factor de reflexión. La energía reflejada a su vez se puede dividir en dos componentes: *especular* y *difusa*. La luz especular, siguiendo la *ley de Snell* de la óptica geométrica, proviene de las caras con superficies suficientemente pulidas para que las irregularidades superficiales sean menores que la longitud de onda de la luz. La luz difusa proviene de las múltiples reflexiones que se producen en las microcaras orientadas aleatoriamente de una superficie mate, y también se genera por dispersión interna cuando la luz incidente penetra dentro de la superficie del material.

Según lo expuesto hasta ahora, la reflexión o es especular o es difusa, siguiendo en este último caso la ley de Lambert. En la realidad las superficies comunes no son ni de un tipo ni del otro, sino que presentan propiedades intermedias. En este sentido se ha intentado modelar superficies mates como una combinación lineal de estos comportamientos extremos, pero los resultados obtenidos no son satisfactorios. La solución más próxima a la realidad se obtiene descomponiendo la reflexión en tres lóbulos difusos [193]. En

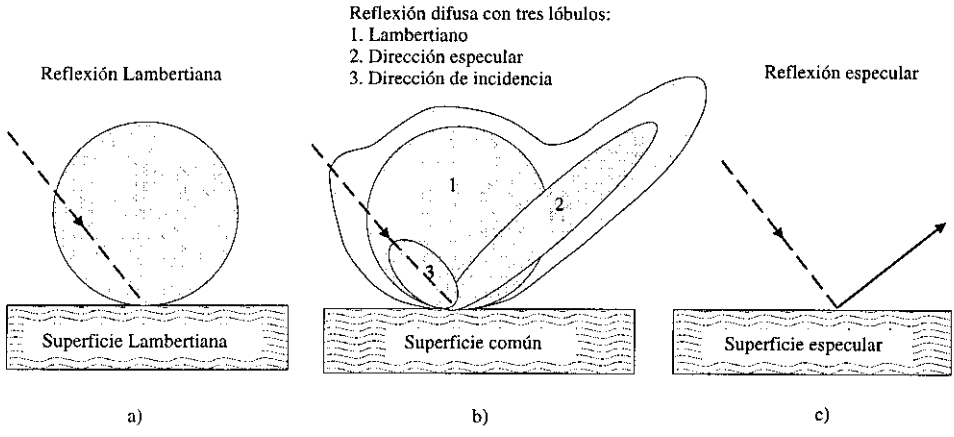


Figura 2.24: Distribuciones de intensidad radiante reflejada para: (a) difusor ideal, (b) superficie común, (c) superficie idealmente especular.

el esquema central de la figura 2.24 podemos ver la distribución de *intensidad radiante* $[W/sr]$ de los tres lóbulos, y en los esquemas laterales los casos ideales correspondientes a difusores perfectos y superficies especulares.

En el caso de superficies comunes (fig. 2.24b), el lóbulo difuso más importante es el lambertiano, que sigue siempre la dirección normal a la superficie, y es debido a las reflexiones múltiples en el interior de la superficie del material. El segundo lóbulo difuso en orden de importancia es el que sigue la dirección especular, comportándose de forma semejante a una reflexión especular aunque con la salvedad de que en este último caso no existe dispersión, y en nuestro caso sí. Esta componente difusa tiene su origen en las reflexiones que se producen en las microcaras orientadas aleatoriamente de una superficie rugosa. El último lóbulo, que se transmite en la dirección de incidencia de la luz, tiene una importancia menor y no siempre se manifiesta, y cuando lo hace puede aproximarse como un valor constante.

Por tanto, siguiendo estas pautas, nuestro modelo para superficies no ideales va a contemplar no solo el lóbulo lambertiano, sino también el lóbulo en la dirección especular. De esta forma si representamos por g a la fracción de la señal reflejada de una forma que tiene una dispersión lambertiana; $1-g$ sería la fracción reflejada de forma difusa en la dirección especular. Representando por K_{esp} a la distribución relativa de la energía al variar θ correspondiente al lóbulo difuso en la dirección especular, y por K_{dif} a la correspondiente distribución lambertiana, podemos plantear que la distribución total, K, se puede expresar así:

$$K = gK_{dif} + (1-g)K_{esp} \tag{2.13}$$

Donde K es adimensional y muestra la fracción de energía recibida por el sensor para un determinado ángulo de incidencia θ y una determinada superficie con fracción difusa g.

Es un hecho conocido que $R_{dif} = \cos\theta$, sin embargo en cuanto a R_{esp} no hay tal unanimidad de criterios. Se suele plantear un modelo basado en la distribución de

las pendientes correspondiente a las pequeñas caras microscópicas que forman una superficie. Representando por m la pendiente de estas caras existen dos distribuciones utilizadas: La distribución gaussiana $K_{esp} = ce^{-(\theta/m)^2}$ [201], y la distribución de Beckman $K_{esp} = \frac{c}{m^2\cos^4\theta}e^{-(\tan\theta/m)^2}$ [9]. Sin embargo, para simplificar al máximo el modelo, hemos considerado la relación que indudablemente existe entre el factor g y la pendiente de las micro-caras m. De esta forma hemos planteado una distribución en la cual el término m desaparece con lo que queda simplificada dependiendo solo de θ y de g.

$$K_{esp} = \frac{e^{-(\theta/c_1 g^{c_2})^2}}{g^{c_2}} \tag{2.14}$$

Por tanto el modelo completo, válido para todo tipo de superficies desde las muy rugosas hasta las más especulares, es el siguiente:

$$\Re = c_1 \log((aD^3 + bD^2 + cD + d)\frac{AF_i}{\pi}\frac{\rho}{D^2}(g\cos\theta + (1-g)\frac{e^{-(\theta/c_3 g^{c_4})^2}}{g^{c_4}})) \tag{2.15}$$

Realizando un ajuste por mínimos cuadrados obtenemos los resultados que se muestran en la figura 2.25a, donde se puede ver como los datos experimentales se ajustan mucho mejor que en el modelo restringido a superficies mates perfectas (fig. 2.23b). Asumiendo que la superficie utilizada en los experimentos refleja de forma difusa el 75% ($\rho = 0.75$) de la luz con longitud de onda de 0,78 μm, hemos deducido que $g = 0.95$ lo cual indica que el 95% de la energía reflejada lo hace según una distribución difusa lambertiana y el 5% restante sigue una distribución difusa en la dirección especular. Utilizando otras superficies con propiedades ópticas diferentes, se pudo comprobar la adaptación del modelo general a los datos experimentales. Como muestra de esta adaptación véase el gráfico de la figura 2.25b, donde se aprecia la distribución total correspondiente a la interacción de las dos componentes difusas. Evidentemente la superficie con $g = 0.78$ presenta un pico más pronunciado que el resto en torno a incidencias frontales a la superficie. A igualdad de reflectividad, debido a que la superficie más pulida refleja más energía de forma difusa en la dirección especular y a que la energía total reflejada se debe conservar, la energía reflejada lambertianamente disminuye proporcionalmente al aumentar la especularidad de la superficie, apreciándose una menor reflectancia para ángulos de incidencia entre 10^o y 90^o que corresponden a reflexiones difusas puras. Véase un ejemplo de este hecho en la figura 2.25b, donde se muestra el caso de dos superficies con la misma reflectividad, $\rho = 0.75$, pero con distintos grados de difusión $g = 0.95$ y $g = 0.78$

Las constantes obtenidas por los ajustes realizados proporcionaron los siguientes valores: $c_1 = 80.9$, $c_3 \simeq 20$ y $c_4 \simeq 5$. Por tanto estamos ya en condiciones de presentar el modelo general de reflectancia correspondiente a la señal generada por el fotosensor a partir de una onda reflejada sobre una superficie real:

$$\Re = 80.9\log((aD^3 + bD^2 + cD + d)\frac{AF_i}{\pi}\frac{\rho}{D^2}(g\cos\theta + (1-g)\frac{e^{-(\theta/20g^5)^2}}{g^5})) \tag{2.16}$$

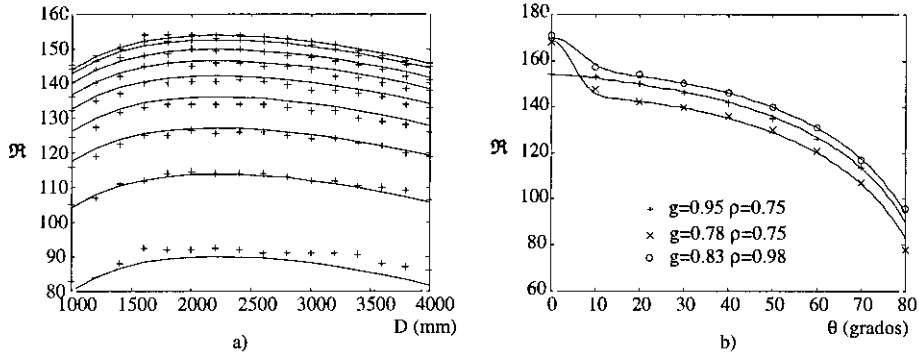


Figura 2.25: Adaptación del modelo general (ec. 2.16) con: a) los datos experimentales del apéndice B, b) superficies con diferentes ρ y g.

La figura 2.26 nos presenta de una forma gráfica el comportamiento de la reflectancia, $\Re$, en función de la distancia, D, y el ángulo de incidencia, θ, para superficies caracterizadas por el grado de reflexión difusa g y la reflectividad ρ. Las figuras 2.26a y 2.26d muestran dos casos con g=1, es decir, superficies difusoras perfectas, donde no hay puntos de inflexión y los cambios en la reflectividad, ρ, no afecta a la forma de la superficie modelada, aunque si generan desplazamientos relativos a lo largo del eje $\Re$. En las figuras 2.26a,b,c se puede observar el paulatino aumento de la inflexión en torno a $\theta = 0^o$, a medida que el coeficiente de reflexión difusa g va disminuyendo, y por tanto, aumenta la reflexión difusa en la dirección especular. Es importante destacar que este tipo de reflexión se podrá dar en cualquier dirección en función de cual sea la orientación de la superficie, sin embargo, solamente aquellas reflexiones con θ próximas a cero (incidencias normales a la superficie) serán captadas por nuestro telémetro láser. Por este motivo, podemos decir que la energía captada en el rango de 10 a 90 grados se deberá únicamente a reflexiones difusas lambertianas, mientras que en el rango de 0 a 10 grados la energía recibida corresponde a la suma de las dos componentes difusas.

2.4.1.3 Modelo aproximado: Superficies difusoras perfectas

Como acabamos de indicar, nuestro sensor, al tener el eje de emisión y recepción iguales, va a captar solamente la reflexión difusa lambertiana, salvo en el caso de incidencias normales a la superficie donde ambos términos difusos contribuyen, pero esta circunstancia tiene una baja probabilidad. Por este motivo tiene sentido plantearse utilizar un modelo aproximado que no considere la reflexión difusa en la dirección especular. Además puede ser conveniente trabajar con un modelo aproximado por motivos de simplicidad computacional. En este caso podemos deducir a partir del modelo general de la ecuación 2.16 la situación correspondiente a un comportamiento totalmente difuso. Basta considerar $g = 1$ y obtendremos el modelo aproximado siguiente:

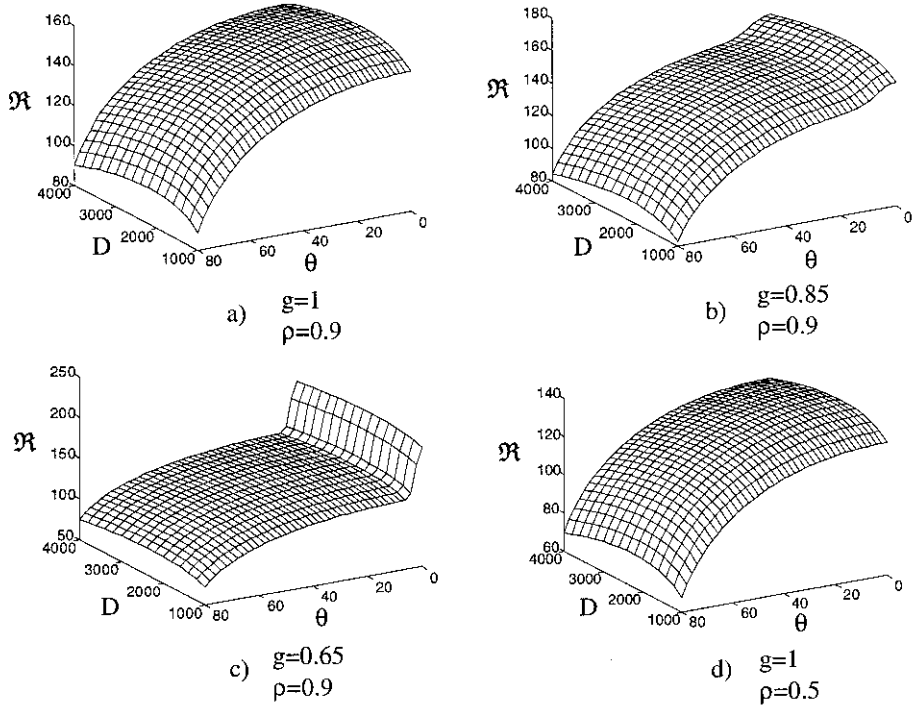


Figura 2.26: Modelo general. En a) b) y c), deformación de la superficie de reflectancias al variar $g = (1, 0.85, 0.65)$, para una $\rho = 0.9$. En d), disminución uniforme de la reflectancia al disminuir la reflectividad, $\rho = 0.5$ y $g = 1$.

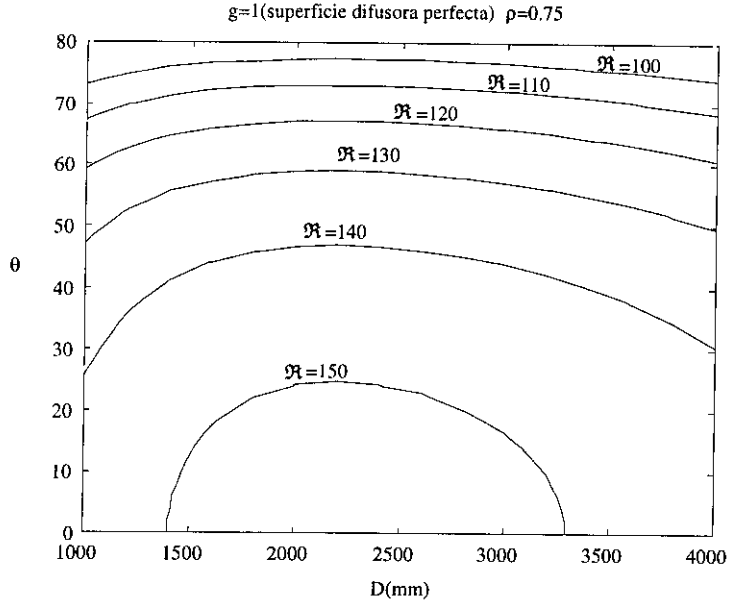


Figura 2.27: Contornos de reflectancia constante para el caso de una superficie difusora perfecta con una reflectividad $\rho = 0.75$.

$$\Re = 80.9 \log((aD^3 + bD^2 + cD + d)\frac{AF_i}{\pi}\frac{\rho \cos\theta}{D^2}) \tag{2.17}$$

Representando los contornos de reflectancia correspondientes a una superficie supuestamente difusora ideal de reflectancia $\rho = 0.75$ obtenemos el gráfico de la figura 2.27, donde se puede apreciar que la reflectancia va a depender fundamentalmente del ángulo de incidencia con la superficie, θ, y va a variar poco dentro del rango de 1 a 4 metros en el que vamos a trabajar.

2.4.1.4 Reflectividad y fracción difusa: Propiedades características de una superficie

Según el modelo expresado en la ecuación 2.16, vemos que existen cinco variables que quedan relacionadas entre sí. Estas variables son la reflectancia, $\Re$, la distancia a la superficie, D, el ángulo de incidencia del haz láser con respecto a la normal a la superficie, θ, la reflectividad de la superficie, ρ, y la fracción difusa g. De estos parámetros, D y $\Re$ son directamente medibles a través del telémetro láser, mientras que θ se puede estimar mediante el análisis de la imagen de distancias generada. Los otros dos parámetros que quedan por deducir, ρ y g, solo dependen de las propiedades ópticas de la superficie,

por lo tanto su estimación permitiría caracterizar la superficie sobre la cual estamos realizando la medida.

La obtención de estos parámetros discriminantes, ρ y g, junto con el análisis previsto basado en la forma, debería proporcionar un método más potente para realizar el análisis de las imágenes, para lo cual los parámetros ρ y g deben ser diferentes para el objeto que para el fondo. No solamente se podría facilitar el reconocimiento, sino que además, se podría hacer una caracterización más precisa del objeto reconocido, no incluyendo solamente su posición y tamaño, sino también propiedades relativas a la superficie del objeto. Estas propiedades superficiales del objeto permitirían realizar una clasificación más rica dentro de los objetos reconocidos. En el caso del reconocimiento de frutas, estas propiedades podrían servir para obtener el grado de madurez del fruto o la presencia de defectos.

2.4.2 Modelado de la repetitividad de la distancia

Cuando realizamos una medida con el telémetro sobre un blanco, obtenemos un valor indicando la distancia a la que se encuentra este. En nuestro caso particular de análisis de escenas mediante imágenes de profundidad interesa, más que la precisión absoluta de la distancia, la sensibilidad y la dispersión en la medida de un punto. Una forma de cuantificar esta sensibilidad consiste en el cálculo de la repetitividad en distancia, o desviación estándar σ_D, obtenida al medir repetidamente sobre un punto fijo. Esta varianza en la distancia depende de muchos factores, como la distancia, la orientación de la superficie, la temperatura, el tipo de superficie y la frecuencia de muestreo. Si la repetitividad de la señal es pobre (σ_D altas), el resultado obtenido cuando generamos una imagen en distancias, es que esta imagen presenta una baja relación señal-ruido y las superficies obtenidas no representan de una forma fiel a las reales. Este hecho hace que la imagen de distancias sin preprocesamiento sea poco adecuada para realizar un reconocimiento basado en la forma y se hace necesario una restauración de la imagen.

Existen muchas técnicas de restauración de imágenes, las cuales se estudiarán con detalle en el capítulo 3. Estas técnicas suelen aplicar filtros gausianos, filtros no lineales, filtros adaptativos o ajustes a curvas predefinidas, siendo unas estrategias más adecuadas que otras en función de las características del problema. En algunos de estos casos es necesario realizar una estimación del ruido presente en la imagen de tal forma que sea posible establecer un criterio para distinguir lo que es ruido de lo que es una discontinuidad propia de la escena la cual debe ser preservada. En la mayoría de los casos la estimación del ruido presente en un punto local de la imagen se hace mediante estadísticas sobre un vecindario muy reducido. Esta estimación suele ser aceptable pero no siempre plenamente satisfactoria debido al reducido número de datos que intervienen en la estimación. El conocimiento local de la repetitividad σ_D correspondiente a cada punto de la imagen, mediante un método determinístico, permitiría obtener una estimación local más fiel del ruido presente en la imagen y por tanto realizar una restauración adaptativa respetando discontinuidades en la imagen y suavizando regiones continuas contaminadas de ruido, mejorándose significativamente la calidad de la imagen. Por este motivo, la obtención de un modelo que nos permita obtener de una forma directa la repetitividad σ_D de cada punto de la imagen, sería deseable para nuestro propósito de restauración de imágenes de distancia.

La restauración de la imagen también se puede hacer tomando varias medidas de distancia en un mismo punto y posteriormente promediando. Esto es posible gracias a que la desviación estándar de la señal se reduce según la raíz cuadrada del número de medidas realizadas. Sin embargo esta técnica requiere una enorme cantidad de medidas reiteradas para conseguir obtener buenas reconstrucciones en las zonas con alto valor de σ_D. Si se utiliza una aproximación activa calculando, en cada punto de la imagen, el número de medidas necesarias para que al promediar se obtenga una σ_D fija, el número total de medidas disminuiría al estar realizando promedios de tamaño variables sobre cada punto de la imagen [155]. Sin embargo esta técnica requiere un barrido de la escena a velocidades variables para poder realizar en cada punto el número de medidas adecuadas, lo cual redunda en una baja velocidad de barrido.

La obtención del modelo que nos de una σ_D local, permitiría realizar una restauración adaptativa en el sentido descrito anteriormente, y a la vez la velocidad de barrido sería la más alta posible, puesto que solamente es necesario realizar una medida en cada punto de la imagen. A continuación presentamos el desarrollo de este modelo. Su aplicación para la reconstrucción de imágenes de distancia se verá en el capitulo 3.

2.4.2.1 Repetitividad en régimen estático

Existen múltiples fuentes de ruido en un sistema óptico de medida como el que estamos utilizando, entre ellos está el ruido fotónico, de cuantificación, del diodo láser, ruido ambiental, térmico y ruidos de amplificación. El ruido térmico originado en el fotodetector o elementos resistivos tiene una distribución gausiana de media cero, siendo directamente proporcional a la temperatura e inversamente proporcional a la impedancia. Este ruido es independiente de la imagen captada y en los casos en que existe suficiente amplificación en el fotodetector queda enmascarado por el ruido fotónico [169]. En particular, los factores más significativos que afectan en mayor medida a nuestro sensor son dos: el ruido fotónico en el fotodiodo y el debido a la cuantificación. El primer foco de ruido es generado en el diodo fotodetector, lo que se pone especialmente de manifiesto cuando el nivel de la señal recibida es muy baja. Representaremos la desviación estándar de este ruido por $\sigma_{D(f)}$. El segundo foco proviene de la cuantificación en la digitalización de la señal, siendo además variable debido a que el sensor es programable y se pueden seleccionar diversos parámetros que afectan a la resolución en la medida. En este caso su desviación estándar la representaremos por $\sigma_{D(c)}$. Suponiendo que estos dos ruidos no están correlacionados, la varianza de la suma de dos señales es la suma de las varianzas de las señales individuales, por tanto podemos plantear la siguiente ecuación

$$\sigma_D = \sqrt{\sigma_{D(f)}^2 + \sigma_{D(c)}^2} \tag{2.18}$$

Cuando el nivel de potencia de la señal recibida es muy bajo, el ruido fotónico en el fotodetector es significativo frente a la señal. Si representamos por E a la energía que corresponde a un fotón, podemos plantear la conocida relación

$$E = \frac{hc}{\lambda} \tag{2.19}$$

donde λ es la longitud de onda de la radiación láser utilizada, h es la constante de Planck

$(6,62 \times 10^{-34}\ J \cdot s)$ y c es la velocidad de la luz.

El número de fotones n recibidos en el fotodiodo del telémetro durante el intervalo de muestreo T, vendrá dado por

$$n = \frac{F_c T}{E} = \frac{\lambda T}{hc} F_c \tag{2.20}$$

donde F_c es el flujo radiante capturado por el telémetro láser, cuya expresión puede verse en la ecuación 2.10.

El número de fotoelectrones n_{fe} generados por el fotodiodo de eficiencia cuántica η en el intervalo de tiempo T, vendrá dado por $n_{fe} = \eta n$. Asumiendo que el proceso de fotoemisión es un proceso de Poisson [169], la desviación estándar en la fotoemisión σ_{fe} es igual a la raíz cuadrada del número n_{fe} de fotoelectrones emitidos. Planteando la relación señal-ruido en el fotodetector $SNR_{(f)}$ como el cociente entre los electrones n_{fe} emitidos en el intervalo T y el ruido de fotoemisión σ_{fe}, obtenemos la siguiente expresión:

$$SNR_{(f)} = \frac{n_{fe}}{\sigma_{fe}} = \frac{n_{fe}}{\sqrt{n_{fe}}} = \sqrt{n_{fe}} = \sqrt{\eta \frac{\lambda T}{hc} F_c} \tag{2.21}$$

Sustituyendo F_c, por la expresión de la ecuación 2.10, tenemos

$$SNR_{(f)} = \sqrt{\frac{\lambda T}{hc} \eta \frac{\beta A F_i}{\pi} \frac{\rho \cos\theta}{D^2}} \tag{2.22}$$

Identificando términos con la ecuación 2.17, obtenemos que la relación señal-ruido se puede expresar de la siguiente forma:

$$SNR_{(f)} = \sqrt{\frac{\lambda}{hc} T 10^{\Re/80.9}} = 20 \times 10^8 \sqrt{T 10^{\Re/80.9}} \tag{2.23}$$

Es conocido [155], que la desviación estándar de la distancia σ_D de un sistema telemétrico basado en diferencias de fase, es inversamente proporcional a la relación señal-ruido $SNR_{(f)}$. De esta forma podemos expresar la desviación estándar correspondiente al ruido fotónico de esta manera:

$$\sigma_{D(f)} \propto \frac{1}{\sqrt{T 10^{A/80.9}}} \tag{2.24}$$

En cuanto al ruido debido a la cuantificación, el fabricante del telémetro asegura que se rige según la siguiente expresión:

$$\sigma_{D(c)} = \frac{D_{max}}{9 \times 10^7 T} \tag{2.25}$$

donde D_{max} es un factor programable que indica la máxima distancia a la cual se pretende medir (expresada en mm), y T es el periodo de muestreo expresado en segundos.

Finalmente, podemos plantear el modelo que representa la desviación estándar de la distancia medida por el telémetro láser mediante la ecuación 2.26, donde se puede ver

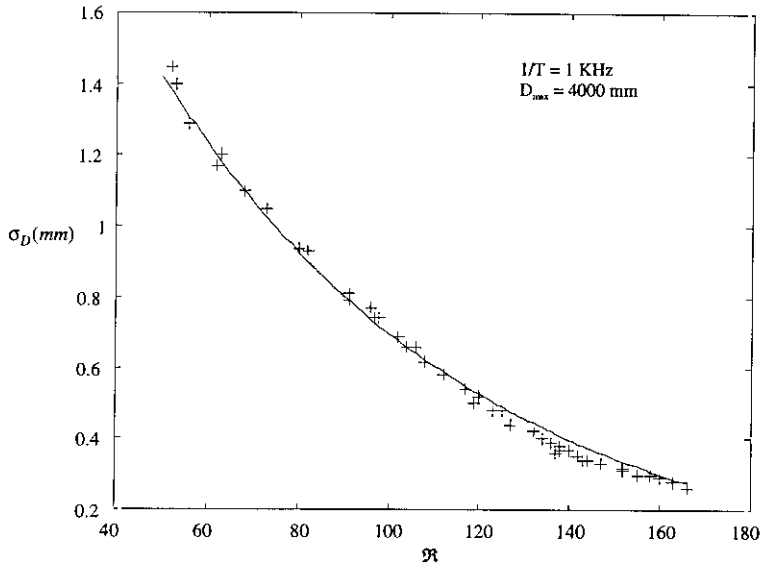


Figura 2.28: Adaptación del modelo de repetitividad en distancia (ec. 2.26) a los datos experimentales presentados en el apéndice B.

que aparecen tres factores que influyen en la fiabilidad de la medida: $\Re$, T y D_{max}. Y existe un factor constante k que se determinará en función de los datos experimentales que se presentan en el apéndice B.

$$\sigma_D = \sqrt{\frac{k}{T10^{\Re/80.9}} + \left(\frac{D_{max}}{9 \times 10^7 T}\right)^2} \tag{2.26}$$

En la figura 2.28, se puede apreciar el resultado de un ajuste por mínimos cuadrados del modelo de la ecuación 2.26 a los datos experimentales presentados en el apéndice B, obteniéndose un $k = 8.37 \times 10^{-3}$. Se puede apreciar, como era de esperar, que a medida que aumenta la amplitud de la señal, o reflectancia $\Re$, la fiabilidad en la medida de distancia es mayor, y por el contrario, si esta disminuye σ_D crece.

Según resultados experimentales, la precisión en la medida (o error absoluto entre una referencia considerada correcta y la medida de distancia generada por el telémetro) está siempre acotada entre ± 2 mm para reflectancias entre 160 y 90, excepto para valores entre 90 y 50 en que la precisión es de ± 4 mm, y para reflectancias entre 50 y 35 donde la precisión es escasa: ± 20 mm. Valores de reflectancia inferiores a 35, provocan la inestabilidad del sistema de medida y por tanto cualquier medida con reflectancias en este rango no deben tenerse en cuenta y debe considerarse como una medida falsa o irrealizable.

En las figuras 2.29 y 2.30, se puede ver la influencia de la reflectancia $\Re$ y de la frecuencia

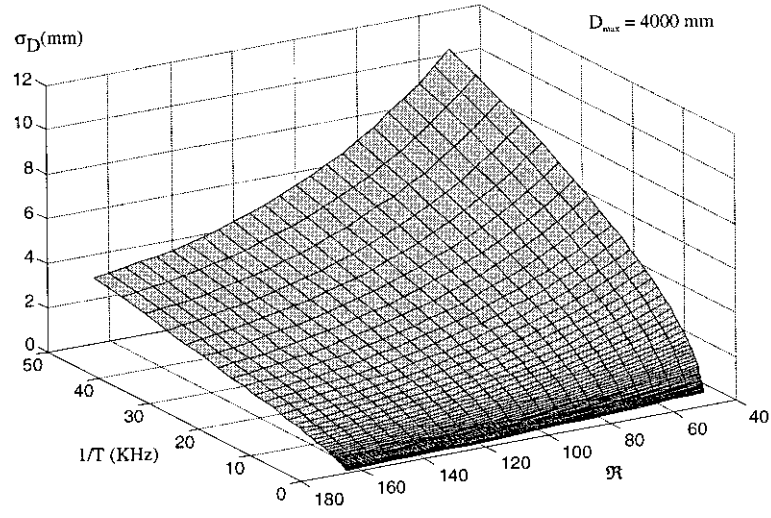


Figura 2.29: Desviación estándar de la distancia medida por el telémetro láser AccuRange4000-LIR en función de la reflectancia $\Re$ y de la frecuencia de muestreo $1/T$.

de muestreo $1/T$, en la desviación estándar. Puede apreciarse como el incremento de la desviación estándar al disminuir la reflectancia, se hace más evidente cuando la frecuencia de trabajo seleccionada se aproxima a la frecuencia de trabajo máxima permitida por el sensor (50 KHz). Teniendo en cuenta que es deseable una frecuencia de trabajo máxima con una desviación estándar mínima, se debe obtener una solución de compromiso seleccionando unos parámetros intermedios que estén lo más próximos a nuestros objetivos. Una frecuencia de muestreo de 5.6 KHz permitiría obtener imágenes de 167×167 puntos en 5 segundos, tomando una medida por cada punto en la imagen, y σ_D quedaría acotada entre 0.75 mm para reflectancias en torno a 160 y 2.1 mm para reflectancias cercanas a 80.

2.4.2.2 Repetitividad en situaciones dinámicas

Hasta ahora todas las experimentaciones han sido realizadas en régimen estático, es decir, sin realizar ningún tipo de barrido o exploración mediante el telémetro láser. Entre los aspectos considerados anteriormente para deducir el modelo de la repetitividad σ_D, no se incluyó ningún parámetro que considerase si el modo de medida era estático o dinámico, puesto que no hay razones objetivas para que este aspecto deba ser incluido. Sin embargo, los resultados experimentales indican que hay una diferencia en la repetitividad de la medida entre un modo estático y uno dinámico o de exploración. La figura 2.31 muestra de una forma gráfica cómo la amplitud del ruido, para un perfil de distancia, es superior en las primeras mil muestras (régimen dinámico) que en las últimas mil muestras (régimen estático).

En la tabla 2.8 se pueden ver algunos de los datos experimentales tomados de forma dinámica para diferentes frecuencias de muestreo $1/T$ y distintas reflectancias $\Re$. Al

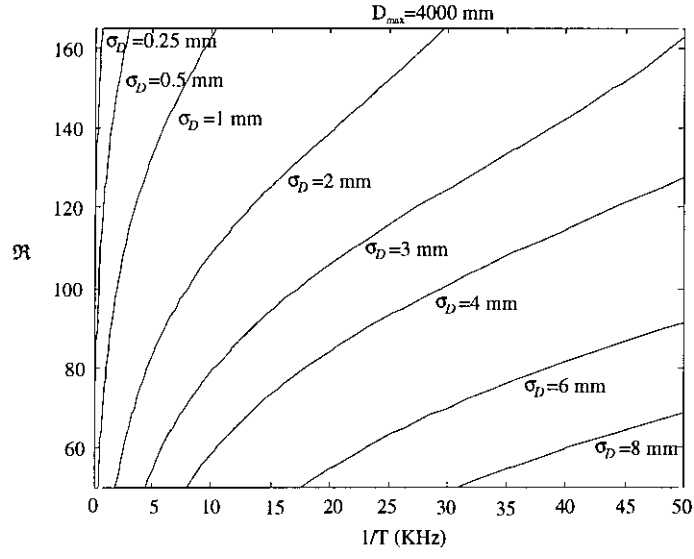


Figura 2.30: Curvas de contorno con σ_D constante correspondientes a la figura 2.29.

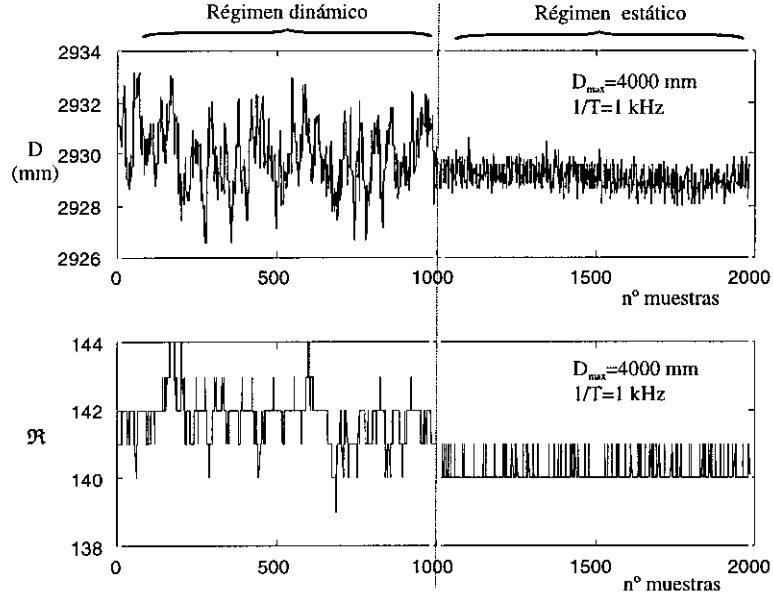


Figura 2.31: Distancia medida correspondiente a un barrido a distancia constante (1000 primeras muestras) y una lectura estática (siguientes 1000 muestras).

1/T (kHz)	$\Re$	σ_{D_est}	σ_{D_din}	$\sigma_{D_adicional}$	$\overline{\sigma_{D_adicional}}$
25	140	2.50	2.70	1.01	
10	140	1.35	1.70	1.03	
5	140	0.70	1.30	1.09	
2	140	0.45	1.15	1.05	
1	140	0.33	1.01	0.95	1.01
25	110	3.60	3.75	1.05	
10	110	1.65	1.92	0.98	
5	110	1.12	1.48	0.97	
2	110	0.86	1.35	1.04	
1	110	0.57	1.10	0.94	

Tabla 2.8: Datos de repetitividad en modo estático y dinámico que muestran la existencia de un ruido adicional de amplitud 1 mm.

comparar la situación dinámica con la estática ante diferentes circunstancias, aparece un ruido constante de magnitud 1 mm que se suma al ruido estático deducido con anterioridad (ec. 2.26). El ruido adicional lo hemos estimado según la ecuación 2.27 y los resultados obtenidos se muestran en la columna de la derecha de la tabla 2.8.

$$\sigma_{D_adicional} = \sqrt{\sigma^2_{D_din} - \sigma^2_{D_est}} \tag{2.27}$$

Las razones para la aparición de este tipo de comportamiento en situaciones dinámicas la atribuimos al principio de medida del sensor, que utiliza una técnica de diferencia de fases con cambio activo de frecuencia de modulación al cambiar la distancia de medida, para lo cual se usa un oscilador electro-óptico que modula el haz láser en función de la propia radiación láser recibida [177]. Es de suponer que la exploración entre dos puntos de un barrido supondrá una pequeña variación en la distancia o tipo de superficie, y de esta forma pueda afectar al lazo de realimentación óptico provocando una cierta inestabilidad en la medida que se corresponda con el ruido adicional detectado. Las vibraciones mecánicas del sensor al realizar la exploración, es otra posible causa de este ruido adicional que aparece en modo dinámico, sin embargo estas vibraciones mecánicas son de menor amplitud que la oscilación necesaria para introducir un ruido con $\sigma = 1$ mm, es decir $\pm 2\sigma = \pm 2$ mm. Las posibles interferencias electromagnéticas generadas por los motores de exploración han sido descartadas al seguir apareciendo el ruido incluso con barridos manuales y al no detectarse ruido en modo estático realizando una fuerte acción de control con la consiguiente emisión de interferencias electromagnéticas.

En cualquier caso, el modelo de repetitividad que nos interesa es el dinámico ya que las imágenes hay que captarlas realizando barridos. Por tanto, el modelo dinámico que utilizaremos vendrá dado por la ecuación 2.28.

$$\sigma_D = \sqrt{\frac{8.37 \cdot 10^{-3}}{T \cdot 10^{\Re/80.9}} + \left(\frac{D_{max}}{9 \cdot 10^7 T}\right)^2 + 1} \tag{2.28}$$

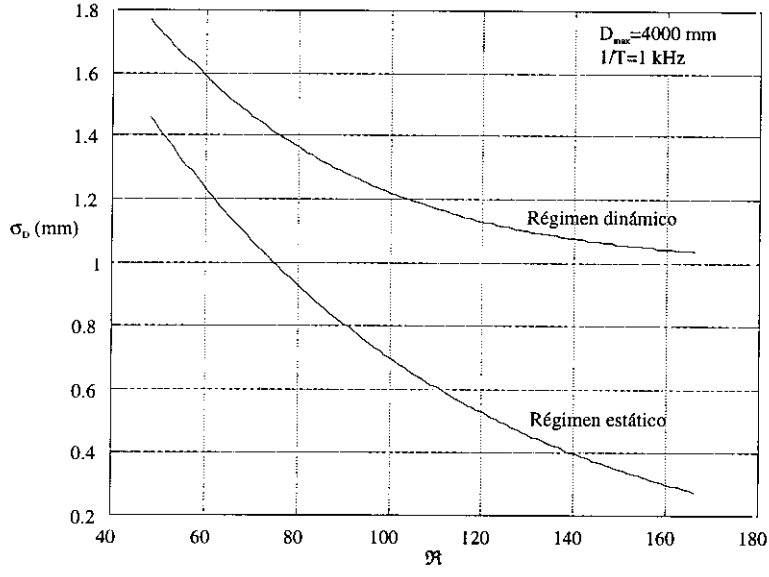


Figura 2.32: Repetitividad en modo dinámico para una frecuencia de muestreo de 1 kHz y D_{max}=4000 mm.

En la figura 2.32 se puede ver la diferencia que existe en repetitividad según el modo sea estático (ec. 2.26) o dinámico (ec. 2.28). Finalmente, en la figura 2.33 podemos ver los contornos de $\sigma_D = cte$ para el caso dinámico. De acuerdo a nuestros objetivos la captura de una imagen de 167×167 puntos en 5 segundos requiere una frecuencia de 5.6 kHz, con lo cual la repetitividad es de 1.2 mm para amplitudes de 160 y de 2.4 mm para amplitudes de 80. Esto significa que nuestros mapas de distancia van a estar contaminados con ruido gausiano con unas amplitudes que variarán entre ± 2.4 y ± 4.8 mm. Este hecho hace que sea necesario aplicar un preprocesamiento a las imágenes de distancia con el propósito de restaurarlas y obtener unas superficies mucho más regulares que permitan realizar un correcto análisis basado en la forma de dichas superficies, incluso utilizando regiones reducidas y muy localizadas. Este proceso de restauración se tratará en el capítulo 3, donde se presentará un nuevo método de restauración que utiliza como estimación del ruido presente en la imagen, el modelo de repetitividad que acabamos de obtener (ec. 2.28).

2.5 Conclusiones

Por lo expuesto anteriormente, los aspectos más destacables que se han tratado a lo largo del capítulo 2 son los siguientes:

- Se ha propuesto una *estrategia de detección y localización mixta* de frutos que opera aplicando un modo de detección automático seguido de una estrategia asistida. Esta estrategia integra los modos de percepción automáticos con los asistidos,

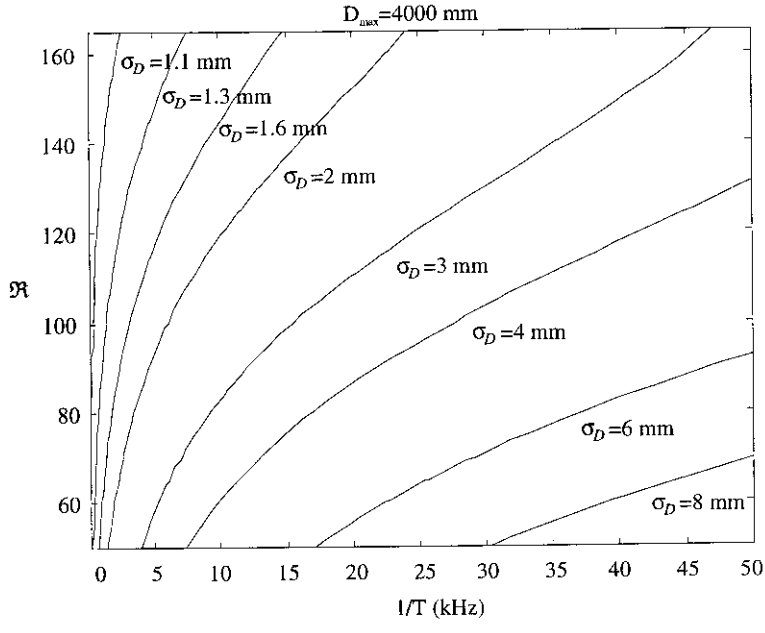


Figura 2.33: Contornos de $\sigma_D = cte$ para el modo dinámico.

permitiendo eliminar las desventajas propias en cada una de las soluciones por separado.

- Hemos propuesto basar el proceso de detección automático de frutas en criterios que utilizan la *forma* de los objetos, las *propiedades ópticas* de las superficies y la *distribución espacial* de los objetos. Para ello se plantea la necesidad de utilizar un sensor que capte información de distancia y reflectancia.

- Se ha realizado un análisis del estado de desarrollo de diferentes *técnicas para captar imágenes de distancia y reflectancia.* Teniendo en cuenta las especificaciones deseadas, la telemetría láser por diferencia de fases es la técnica más adecuada, seleccionando para nuestras investigaciones el telémetro Ac-4000-LIR de Acuity.

- Hemos realizado un *análisis tecnológico* de los sistemas de telemetría láser actuales, definiendo un *índice de eficiencia tecnológico* M, y presentando las relaciones de compromiso existentes entre el rango de medida, la frecuencia de muestreo, la repetitividad de la medida y la potencia de la fuente de emisión. Concluimos que la mejora en las especificaciones, tomando en cuenta los aspectos de seguridad, pasa por mejorar la tecnología o aumentar la potencia láser en el infrarrojo medio.

- Presentamos la configuración física de un *sistema de exploración* telemétrico por deflexión de haz en acimut y elevación para la generación de imágenes de distancia y reflectancia, y para actuar como sistema de señalización en el modo de detección asistido. Se muestran las debilidades de este prototipo que se centran en las inercias

existentes, las cuales limitan la adquisición de imágenes de 167 × 167 puntos a tiempos superiores a 15 s.

- Presentamos un *modelo matemático del telémetro* que relaciona la reflectancia registrada con diferentes factores que intervienen en el proceso de medida (distancia, ángulo incidencia, reflectividad,...). Este modelo permite interpretar la información de reflectancia y es la base para una integración entre las imágenes de distancia y amplitud, y para obtener propiedades ópticas de la superficie de los objetos como la reflectividad ρ.

- Hemos *modelado la repetitividad de la distancia*, tanto en régimen estático como en dinámico, concluyendo que los factores que más influyen son la reflectancia y el periodo de muestreo. Este modelo permite por tanto obtener una estimación del ruido presente en la imagen de distancias a partir de la imagen de reflectancias, y podrá ser utilizado en la restauración de aquella.

Capítulo 3

Restauración del mapa de distancias: Filtrado adaptativo y calidad de restauración

Resumen. *En este capítulo presentamos la elaboración tanto de una nueva técnica de restauración de imágenes como de una métrica que permite evaluar la eficiencia del método de restauración. Inicialmente revisamos diversas técnicas de preprocesamiento de imágenes incluyendo procesos de realce y de restauración. Revisamos también algunos métodos utilizados para evaluar diferentes algoritmos de restauración, poniendo de manifiesto que es una tarea escasamente formalizada. Debido a la necesidad de utilizar un método de evaluación que valore la fidelidad y la suavidad de una restauración, desarrollamos la métrica GRI que es el resultado de combinar dos índices independientes que valoran ambos factores de calidad de restauración. Utilizando la métrica recién definida, clasificamos diferentes métodos de restauración concluyendo que los mejores algoritmos encontrados, o bien no eliminan convenientemente el ruido, y por tanto no existe suficiente suavidad en las superficies, o proporcionan suavidad pero no respetan las discontinuidades. Debido a ello presentamos nuestro propio método de restauración que hemos denominado filtrado* 3σ-MPF *que se basa en un ajuste de planos multiresolución que evoluciona de regiones grandes a pequeñas hasta encontrar un ajuste con* fidelidad-3σ. *En el caso de que no se encuentre dicho ajuste se supone que estamos en una discontinuidad y se aplica un filtro de preservación de bordes MTM. Hemos evaluado esta técnica mediante la métrica GRI, realizando un comparación con otros algoritmos de restauración, obteniéndose resultados siempre superiores del filtro adaptativo* 3σ-MPF *ante degradaciones de ruido gausiano y gausiano-impulsional.*

En el capítulo anterior vimos cómo mediante un sistema de exploración por deflexión del haz de un telémetro láser, éramos capaces de obtener una imagen de distancia y otra de reflectancia. La imagen de distancia, obtenida utilizando el sensor láser configurado para medir a 5.6 kHz, está contaminada de ruido gausiano aditivo con amplitudes entre ± 2.4 mm y ± 4.8 mm. Además, dicha imagen puede contener ruido impulsivo en lugares donde la amplitud de la señal láser recibida es muy baja o donde aparezcan reflexiones altamente especulares. Este tipo de ruido se caracteriza por valores de distancia alejados de los valores reales, sin embargo solo suele afectar a porcentajes muy reducidos de la imagen. La presencia de ambos tipos de ruido supone una desviación de los valores reales frente a los capturados, lo cual redunda en la adquisición de una imagen degradada.

Es indudable que imágenes degradadas, presentan mayores dificultades a la hora de extraer características de sus datos, puesto que la información contaminada puede alejarse significativamente de un determinado modelo asumido. En nuestro caso el modelo de objetos a procesar corresponde a una esfera de radio medio en torno a 35-45 mm (naranjas). Variaciones entre ± 2.4 mm y ± 4.8 mm suponen una degradación importante a la hora de detectar formas y curvaturas en regiones reducidas, como es el caso de escenas con árboles frutales donde debido a las oclusiones solo es visible parte de la superficie de cada fruto. Por tanto, antes de comenzar realizando cualquier tipo de análisis sobre las imágenes, necesitamos aplicar algún algoritmo de restauración eficiente. Debido a que existe un gran rango dinámico en los mapas de profundidad con diversas superficies a muy distintas distancias, lo cual genera discontinuidades muy abruptas en la transición de una superficie a otra, este algoritmo debe ser capaz de eliminar la mayor cantidad de ruido y a la vez preservar las discontinuidades significativas en la imagen. Esto es, el método de restauración debe ofrecer suavidad y fidelidad.

Con el objetivo de aplicar la técnica de restauración más apropiada a nuestro problema, comenzaremos realizando un repaso de las principales técnicas existentes de preprocesamiento de imágenes.

3.1 Técnicas de preprocesamiento de imágenes

Existen dos grandes categorías en las que podemos encuadrar las diversas técnicas de preprocesamiento existentes. La finalidad de cada una de estas categorías se define a continuación:

- *Métodos de realce.* La etapa de preprocesamiento está dirigida a aumentar la calidad subjetiva de una imagen observada por un usuario humano, y a la vez, facilitar el reconocimiento automático. No se pretende incrementar la fidelidad de la imagen tratada respecto a una imagen ideal, cosa que se trata en la restauración, sino realzar detalles inicialmente no apreciables.

- *Técnicas de restauración.* La etapa de preprocesamiento se dirige a mejorar imágenes digitales, degradadas por la existencia de ruido, y facilitar, por tanto, las etapas sucesivas orientadas al reconocimiento automático. El objetivo de la restauración consiste en obtener una imagen lo más próxima a una imagen de referencia ideal, la cual no contiene el ruido presente en la imagen degradada de origen.

En nuestro caso lo que queremos es eliminar el ruido aditivo presente en la imagen de distancias y preservar todas las discontinuidades para no introducir errores en distancia que afecten a la forma de las superficies o a la localización de los objetos. La imagen restaurada queremos que sea lo más fiel posible a un mapa de profundidades ideal. Por tanto el tipo de preprocesamiento que queremos realizar es una *restauración* de la imagen. A continuación presentamos los métodos de preprocesamiento más destacados tanto en la restauración como en el realce de imágenes.

3.1.1 Métodos de realce

Debido a que nuestro objetivo no está dirigido hacia el realce de una imagen, solamente vamos a presentar de una manera ilustrativa las tendencias más destacadas en este campo, lo cual nos permitirá apreciar con mayor claridad la diferencia de enfoque entre realce y restauración.

3.1.1.1 Transformaciones del histograma

Dada una imagen con diversos niveles de gris, podemos obtener su *histograma*, h, mediante un gráfico donde, en abscisas, se representan los n niveles de gris que existen en la imagen y en ordenadas se indica la frecuencia de aparición del nivel de gris correspondiente $h(n)$. Para imágenes digitales, el histograma corresponde a un diagrama de barras donde la altura $h(n_k)$ de cada una de ellas representa el número de puntos (pixels) con un nivel de gris n_k. Frecuentemente los niveles de gris n_k son normalizados entre los valores 0 y 1, donde el cero se corresponde con la mínima intensidad o negro y el uno con el blanco o máxima intensidad. En el caso de imágenes de distancia la normalización se realiza asignando un 1 a la distancia máxima y un 0 a la distancia mínima, y si las imágenes son de color el proceso es análogo con cada uno de los tres histogramas correspondientes a las componentes espectrales RGB.

Las técnicas de realce presentadas en este apartado se basan en modificar el histograma característico de una imagen con la intención de aumentar su visibilidad. Para ello se aplica una función $s_k = T(n_k)$ que transforma los niveles de gris iniciales n_k en otra gama de niveles s_k. Estas son las estrategias habituales de conformación de histogramas:

- *Oscurecimiento.* La transformación $T(n_k)$ es de tipo cuadrática o cúbica, es decir, $T(n_k) = n_k^2$ ó $T(n_k) = n_k^3$. Así se consigue aumentar el contraste en las zonas muy luminosas oscureciendo las zonas menos luminosas.

- *Aclarado.* La transformación $T(n_k)$ es de tipo raíz cuadrada o cúbica, es decir, $T(n_k) = n_k^{1/2}$ ó $T(n_k) = n_k^{1/3}$. Así se consigue aumentar el contraste de las zonas muy oscuras, aclarando las menos oscuras.

- *Aumento de contraste.* Aplicando la función error $fer(x) = \frac{2\int_0^x e^{-y^2}dy}{\sqrt{\pi}}$, se puede conseguir un aumento de contraste tanto en las zonas claras como en las oscuras. El efecto final es la generación de un histograma bimodal, donde cada modo contiene

los tonos oscuros y los claros. La transformación viene dada por [135]:

$$T(n_k) = \frac{fer(\frac{n_k - 0.5}{\sqrt{2}}) + \sqrt{2}/4}{fer(\sqrt{2}/4)} \tag{3.1}$$

- *Ecualización del histograma.* La ecualización del histograma consiste en repartir todos los niveles de gris de manera uniforme entre los puntos de la imagen. De esta forma se obtiene un histograma plano. La transformación $T(n_k)$ que se aplica debe ser esta:

$$T(n_k) = \sum_{n_j=0}^{n_k} h(n_j) \tag{3.2}$$

- *Especificación del histograma.* Este proceso consiste en aplicar una transformación $T(n_k)$ que permita obtener una imagen con un determinado histograma. El procedimiento se divide en dos etapas, primeramente se realiza la ecualización del histograma y a partir de este se aplica una transformación que permite modular el histograma a la forma final deseada.

- *Transformaciones locales.* Los anteriores métodos son globales, es decir, las transformaciones de cada pixel se hace atendiendo a la distribución de tonos de toda la imagen. Los métodos globales son apropiados para mejorar de un modo uniforme la imagen, sin embargo, a veces es necesario resaltar detalles que aparecen solo en regiones pequeñas de la imagen, y los métodos globales no pueden debido a que la influencia en el histograma de una pequeña región no es significativa. La solución está en dividir la imagen en diferentes regiones, y sobre cada una de ellas aplicar una técnica basada en el histograma como las globales que acabamos de ver.

3.1.1.2 Acentuado de bordes

Diversos estudios muestran que una imagen con los bordes acentuados es subjetivamente más agradable de ver que una reproducción fiel de la escena [169]. El realce de los bordes de una imagen puede ser realizado de diversas formas:

- *Enmascarado suavizado.* Combinando linealmente un perfil, muestreado espacialmente a alta resolución, con un perfil interpolado tras muestrear a baja resolución (perfil suavizado), obtenemos un perfil resultante en cuyos bordes presenta unas oscilaciones que visualmente producen un efecto de resaltado de los mismos.

- *Filtro paso-alto.* Convolucionando máscaras de tipo paso-altas se obtiene igualmente un efecto de acentuado de los bordes.

- *Diferenciado estadístico.* Básicamente consiste en dividir cada punto de la imagen $g(i,j)$ por la desviación estándar $\sigma(i,j)$ en torno a su vecindario. Añadiendo un factor de ganancia A, la media deseada m_d, la desviación estándar deseada σ_d y

un coeficiente de proporcionalidad α, la transformación que permite obtener una imagen resaltada $g^*(i,j)$, se puede expresar así:

$$g^*(i,j) = (g(i,j) - \bar{g}(i,j)) \left(\frac{A\sigma_d}{A\sigma(i,j) + \sigma_d} \right) + \alpha m_d + (1-\alpha)\bar{g}(i,j) \qquad (3.3)$$

- *Modificación de los coeficientes de las transformadas.* Transformaciones como la de Fourier o la de Hadamard generan una descomposición espectral de una imagen en coeficientes característicos de la imagen. Una técnica de resaltado consiste en modificar la amplitud de los coeficientes de la transformación sin modificar la fase. Si elevamos la amplitud a un exponente entre 0 y 1, estamos reduciendo en un alto grado los coeficientes con mayor valor, mientras que los de menor amplitud no se reducen tanto. El resultado es una redistribución de la energía que se traduce en una visualización más agradable.

- *Cepstrum generalizado.* Es una operación no lineal que consiste en realizar una transformación como la de Fourier o Hadamard, aplicar el logaritmo a los coeficientes y posteriormente realizar la transformada inversa para obtener de nuevo la imagen en el dominio espacial. Debido a que normalmente las imágenes suelen contener componentes de alta frecuencia de baja magnitud y componentes de baja frecuencia de alta amplitud, al aplicar el logaritmo en el dominio de la transformación se produce una ecualización que atenúa las componentes de baja frecuencia y por tanto provoca un realce de los bordes de la imagen.

3.1.1.3 Coloreado

El color es otra característica que permite resaltar diferentes componentes en una imagen. En este sentido existen dos técnicas básicas usadas para facilitar la visualización de una imagen [169]:

- *Cambio de colores.* La técnica consiste en la reconfiguración de los tonos de color de una imagen que inicialmente tenía colores naturales. Existen varias razones por las cuales puede ser interesante aplicar esta técnica. En primer lugar, cambiando colores de fondo como el color del cielo a rojo o el color del césped a azul, el observador va a prestar más atención a los objetos que si estuviesen coloreados normalmente. El ojo humano es más sensible al color verde (555 nm), con lo cual se podría transformar ciertos colores que son típicos de algunos objetos, en colores dentro de las bandas de alta sensibilidad para que dichos objetos sean más fácilmente distinguibles. Finalmente, también puede ser interesante realizar transformaciones a la banda del azul, pues en ella el ojo es más sensible a cambios de tonalidad y por tanto se aumenta el contraste.

- *Seudocolor.* Esta técnica utiliza imágenes con tonos de gris y realiza una transformación para generar una imagen en color, la cual es más agradable de visualizar. Para ello se utilizan transformaciones lineales o no lineales, las cuales se pueden visualizar como una trayectoria a través de un espacio tridimensional RGB, donde el inicio de la trayectoria corresponde a la transformación aplicada a los tonos negros y el final de la trayectoria indica la transformación que se aplica a los tonos blancos.

3.1.2 Métodos de restauración

La restauración de señales o de imágenes constituye un área del conocimiento que tiene un amplio historial. Es sin duda uno de los campos del procesamiento de la información donde más aplicaciones se pueden encontrar, pues en todo proceso en el que intervenga la captación, transmisión o transformación de información es habitual aplicar una etapa de restauración para recomponer la información original degradada. Para presentar las técnicas más significativas utilizadas en restauración, hemos hecho una clasificación agrupándolas en cuatro clases: lineales, no lineales, adaptativas e iterativas. Algunos de estos grupos pueden solapar con otros, como por ejemplo el caso de los adaptativos o los iterativos que pueden ser tanto lineales como no lineales. Sin embargo, los hemos agrupado en estas clases por presentar cada uno de ellos características bien diferenciadas.

Con el objetivo de presentar los diferentes métodos de restauración con la mayor consistencia posible, vamos a utilizar la siguiente notación:

- (i, j) coordenadas de un punto de la imagen.
- $f(i, j)$ imagen ideal, la cual no tiene ruido.
- $g(i, j)$ imagen captada, la cual está degradada.
- $\hat{f}(i, j)$ estimación de la imagen ideal $f(i, j)$ a partir de la imagen captada $g(i, j)$. Es el resultado de la restauración.
- $N \times N$ tamaño de la imagen.
- V_{ij} conjunto de L puntos pertenecientes a un vecindario en torno al punto (i, j).

3.1.2.1 Filtros lineales

Este tipo de filtros están basados en la teoría de sistemas lineales. La idea básica radica en que normalmente las señales o imágenes a restaurar tienen un contenido espectral de baja frecuencia, mientras que el ruido que las contamina está caracterizado por anchos de banda mayores. La estrategia de filtrado consiste en eliminar las componentes de alta frecuencia y quedarse con el resto. Este filtrado, que normalmente es paso-bajo, se puede hacer tanto en el dominio de la frecuencia como es el dominio espacial aplicando la convolución, siendo el efecto totalmente equivalente aunque no la eficiencia del cálculo. En principio una convolución tiene una complejidad $O(N^4)$, resultando menos eficiente que un filtrado mediante transformada rápida de Fourier FFT $O(N^2 \times log(N))$ [135]. Sin embargo, el filtro diseñado en el dominio de la frecuencia se corresponde en el dominio espacial a un filtro de tamaño $N \times N$, el cual se puede aproximar por una máscara de tamaño muy reducido $n \times n$, con lo cual al aplicar la convolución se obtienen complejidades del orden de $O(N^2 \times n^2)$ que son menores que en el caso de la FFT. Realizando el filtrado mediante convolución de máscaras reducidas se consigue mayor eficiencia computacional que cuando se hace en el dominio de la frecuencia, especialmente cuando $N \gg n$ [169]. En función de la máscara utilizada para la convolución obtenemos las siguientes soluciones:

- *Media aritmética.* Consiste en realizar un promediado entorno al vecindario de cada punto de la imagen. Para ello se puede aplicar, p.ej., una máscara de tamaño 3×3 o 5×5 con todos los valores igual a $1/9$ o $1/25$, respectivamente.

1	1	2	2	2	1	1
1	2	2	4	2	2	1
2	2	4	8	4	2	2
2	4	8	16	8	4	2
2	2	4	8	4	2	2
1	2	2	4	2	2	1
1	1	2	2	2	1	1

Figura 3.1: Máscara gausiana para restauración de tamaño 7×7.

- *Filtro Gausiano.* En este caso la máscara utilizada tiene una distribución gausiana. Para ello se utilizan como mínimo máscaras de tamaño 7×7 (fig. 3.1), pudiéndose utilizar de mayor tamaño pero teniendo en cuenta que se va a producir un suavizado con un alto efecto desfocalizador.

- *FIR (Finit Impulse Response).* Es un filtro utilizado fundamentalmente en el suavizado de señales unidimensionales que, eliminando su causalidad, puede aplicarse a imágenes de una manera similar a los casos anteriores , es decir, convolucionando una máscara que realiza un promedio ponderado sobre el vecindario de un punto.

En otros casos, la restauración puede hacerse sin utilizar máscaras de convolución que se desplazan a lo largo de la imagen. Si disponemos de varias imágenes de la misma escena que están degradadas por la adición de ruido con distribución *normal*, al promediarlas obtenemos una imagen resultante que tiene un ruido cuya desviación estándar es menor que la correspondiente a cada imagen individual. En concreto, la dependencia es inversamente proporcional a la raíz cuadrada del número M de imágenes promediadas:

$$\sigma_{promedio} = \frac{\sigma}{\sqrt{M}} \tag{3.4}$$

Otros filtros lineales, frecuentemente aplicados en el dominio de Fourier son el filtrado inverso y los filtros de Wiener:

- *Filtro inverso.* Dada una imagen degradada $g(i,j)$, que se ha generado al actuar un sistema lineal degradante $H(u,v)$ sobre una imagen ideal $f(i,j)$, podemos reconstruir la imagen ideal sin más que invertir la función degradante $H(u,v)$ y aplicarla sobre $g(i,j)$. Es decir, en el dominio de Fourier tendríamos que la transformada de la imagen reconstruida es: $\hat{F}(u,v) = G(u,v)/H(u,v)$. Este método solo es correcto cuando no existe ruido aditivo y la función degradante $H(u,v)$ no tiene términos igual a cero en alguna de sus componentes espectrales. Si esto se cumple, y por supuesto se conoce con certeza al sistema lineal degradante $H(u,v)$, se obtiene una reconstrucción perfecta.

  Sin embargo, si existe un cierto ruido N(u,v) en la imagen y si hay términos de $H(u,v)$ pequeños, se producen grandes valores en el espectro de la imagen reconstruida $\hat{F}(u,v)$ y por tanto se producen errores en la reconstrucción puesto

que el componente aditivo introducido por el ruido hace que $F(u,v)$ y $\hat{F}(u,v)$ se hagan diferentes, como se puede ver en la siguiente ecuación:

$$\hat{F}(u,v) = \frac{G(u,v)}{H(u,v)} + \frac{N(u,v)}{H(u,v)} = F(u,v) + \frac{N(u,v)}{H(u,v)} \tag{3.5}$$

Otra limitación del filtro inverso estriba en la dificultad en determinar la naturaleza exacta del proceso de degradación, es decir, la obtención de $H(u,v)$. Además este proceso de degradación no debe tener componentes iguales a cero, puesto que de ser así, no es posible aplicar el filtro correctamente y se hace necesario recurrir a algunos arreglos heurísticos.

- *Filtro de Wiener.* En este caso se incorpora un conocimiento a priori del ruido presente en la imagen que se da en términos de su densidad espectral $W_N(u,v)$. La respuesta a un impulso del filtro de Wiener es calculada para minimizar el error cuadrático medio, por ello a este filtro también se le conoce como filtro LMS (Least Mean Square). La función de transferencia del filtro es esta:

$$\frac{H^*(u,v)}{|H(u,v)|^2 + \frac{W_N(u,v)}{W_F(u,v)}} \tag{3.6}$$

En esta expresión se necesita utilizar una estimación de la degradación $H(u,v)$, de la densidad de potencia espectral del ruido $W_N(u,v)$ y de la imagen ideal $W_F(u,v)$. En el caso de que no haya ruido podemos ver que el filtro de Wiener se convierte en un simple filtro inverso, con lo cual se vuelve a ver que el filtro inverso solo se debe aplicar cuando no hay ruido.

3.1.2.2 Filtros no lineales

Una de las familias de filtros no lineales más representativas son los *filtros de orden.* La técnica recibe este nombre debido a que siempre se realiza una ordenación de los valores correspondientes al vecindario de un punto, y posteriormente se aplica una transformación lineal. Si denominamos por $\vec{v}_{ij}$ al vector desordenado que contiene L puntos vecinos al punto (i,j), y a $\vec{v}_{(ij)}$ al mismo vector pero con sus componentes ordenadas, podemos representar la acción de un filtro de orden de la siguiente forma:

$$\hat{f}(i,j) = \vec{a}^T \vec{v}_{(ij)} \tag{3.7}$$

donde $\vec{a}^T$ es un vector que expresa el tipo particular de filtro de orden que se está utilizando. Este vector debe ser simétrico y no debe producir desviaciones en la salida, es decir, $\sum_{i=1}^{L} a_k = 1$ y $a_k = a_{L-k}$. En función del vector $\vec{a}$ utilizado podemos destacar los siguientes filtros de orden [34]:

- *Mediana.* Consiste en ordenar todos los valores y coger el del medio como estimación, es decir, $\vec{a}^T = [0 \ldots 0\ 1\ 0 \ldots 0]$. El filtro de la mediana se caracteriza por eliminar ruido impulsivo y preservar discontinuidades, siendo por tanto de naturaleza robusta al no verse influenciado por la presencia de ciertos valores dispares muy alejados del valor real. Sin embargo no realiza un suavizado satisfactorio cuando el ruido no es impulsional.

- *Mediana ponderada.* El efecto de dar más peso o influencia a unos valores que a otros, se puede conseguir duplicando algunos valores, p. ej. aquellos que estén más cerca del punto central del vecindario. Una vez hecho esto se aplica el filtro de la mediana, con lo cual aumenta la probabilidad de que alguno de los puntos que se han duplicado sean considerados como la mejor estimación.

- *Media truncada (Trimmed mean).* Esta técnica combina las características del filtro de la mediana en cuanto a la supresión de ruido impulsional y preservar bordes, con las propiedades de suavizado de ruido no impulsional de los filtros lineales paso bajos. La idea consiste en eliminar los extremos del vector ordenado $\vec{v}_{(ij)}$ y aceptar como estimación el promedio de los valores centrales no eliminados. Es decir, $\vec{a}^T = [\underbrace{0 \ldots 0}_{M} \underbrace{\frac{1}{L-2M} \cdots \frac{1}{L-2M}}_{L-2M} \underbrace{0 \ldots 0}_{M}]$.

- *Rango medio.* En este caso se promedian solamente los extremos del vector ordenado $\vec{v}_{(ij)}$, con lo cual $\vec{a}^T = [1/2 \ \ 0 \ldots 0 \ \ 1/2]$. Este filtro puede ser una estimación más eficiente que la media cuando la distribución del ruido tiene colas ligeras y suaves.

- *Filtro de orden de mínima varianza (Minimum Variance Unbiased-MVUB).* Suponiendo que hay un ruido estacionario de media cero, podemos plantear la minimización de la varianza del error respecto al promedio del vecindario. Si denominamos R a la matriz de autocorrelación de las medidas de ruido ordenado, podemos definir el vector $\vec{a}^T$ como $\frac{R^{-1}\vec{1}}{\vec{1}^T R^{-1}\vec{1}}$ [34].

Otro grupo de filtros no lineales son los *filtros morfológicos*, los cuales tienen versiones tanto para imágenes binarias como para aquellas con un rango completo de valores. Funcionan aplicando un *elemento estructurante*, o máscara de forma definible y que normalmente es una ventana cuadrada 3×3, sobre toda la imagen realizando operaciones binarias entre el elemento estructurante y los puntos de la imagen correspondientes. En el caso de imágenes no binarias se aplican funciones como el máximo y el mínimo. Estos son algunos de los operadores:

- *Erosión.* Si alguno de los puntos de la imagen bajo el elemento estructurante B vale 0, entonces el punto bajo consideración se hace cero, en caso contrario vale 1. En una imagen no binaria la erosión se obtiene aplicando la función mínimo sobre los puntos de la imagen cubiertos por el elemento estructurante. El efecto visual obtenido es precisamente de erosión de los contornos de los objetos en la imagen, de ahí su nombre.

$$\hat{f}(i,j) = \text{Erosión}(g(i,j)) = \prod_{x,y \in B} g(i+x, j+y) \tag{3.8}$$

$$\hat{f}(i,j) = \text{Erosión}(g(i,j)) = \min_{x,y \in B} g(i+x, j+y) \tag{3.9}$$

- *Dilatación.* El efecto es el de una dilatación o crecimiento de los objetos, para lo cual se aplica un XOR o la función máximo.

$$\hat{f}(i,j) = \text{Dilatación}(g(i,j)) = \bigoplus_{x,y \in B} g(i+x, j+y) \tag{3.10}$$

$$\hat{f}(i,j) = \text{Dilatación}(g(i,j)) = \max_{x,y \in B} g(i+x, j+y) \tag{3.11}$$

- *Cierre.* Es una dilatación seguida de una erosión. Se llama cierre puesto que los pequeños agujeros dentro de los objetos quedan rellenados y los objetos muy próximos quedan unidos.

$$\hat{f}(i,j) = Cierre(g(i,j)) = \text{Erosión}(\text{Dilatación}(g(i,j))) \tag{3.12}$$

- *Apertura.* Es una erosión seguida de una dilatación. En este caso objetos alargados con algunos agujeros o semicortes quedan fraccionados y separados, de ahí su nombre. Igualmente objetos pequeños compuestos de unos pocos puntos desaparecen.

$$\hat{f}(i,j) = Apertura(g(i,j)) = \text{Dilatación}(\text{Erosión}(g(i,j))) \tag{3.13}$$

Además de los filtros de orden y los morfológicos existen muchos más filtros no lineales aplicados en la restauración de imágenes o señales. A continuación citamos algunos:

- *Filtrado homomórfico.* Este tipo de filtro es adecuado cuando existe ruido multiplicativo en la imagen, es decir, $g(i,j) = f(i,j) \cdot h(i,j)$. La técnica consiste en aplicar una transformación no lineal como el logaritmo neperiano para conseguir que el término que corresponde al ruido quede relacionado mediante una adición al término que contiene la imagen sin degradar $f(i,j)$.

$$\log g(i,j) = \log f(i,j) + \log h(i,j) \tag{3.14}$$

  De esta forma se pueden aplicar filtros lineales para separar ambas componentes y posteriormente mediante una exponenciación se puede recuperar la imagen restaurada la cual ya no contiene la degradación multiplicativa.

- *Filtro de histéresis.* Este tipo de filtro elimina fluctuaciones pequeñas, preservando la estructura de las principales transiciones. Funciona como un proceso con histéresis donde hay tantos estados como niveles de gris en la imagen. Una vez que estamos en un estado para salir de él será necesario que se produzca una variación significativa del valor de gris superior a un umbral determinado; si no es así, el estado o valor de salida quedará inalterado y por tanto se eliminan las pequeñas fluctuaciones.

- *Media armónica.* Este filtro no lineal es adecuado para eliminar ruido gausiano, preservar bordes y a la vez eliminar ruido impulsional de tipo positivo. Considerando que v_{ij} es el vecindario de tamaño L entorno al punto (i,j), podemos definir el filtro así:

$$\hat{f}(i,j) = \frac{L}{\sum_{x,y \in V_{ij}} \frac{1}{g(i+x,j+y)}} \tag{3.15}$$

- *Media contra-armónica.* Como en el caso anterior, este filtro es adecuado para eliminar ruido gausiano, preservar bordes y a la vez eliminar ruido impulsional. Si

el parámetro P es positivo eliminará ruido impulsional negativo y si P es negativo eliminará los impulsos positivos.

$$\hat{f}(i,j) = \frac{\sum_{x,y \in V_{ij}} g(i+x, j+y)^{P+1}}{\sum_{x,y \in V_{ij}} g(i+x, j+y)^{P}} \tag{3.16}$$

- *Media* Y_p. Presenta las mismas características que la media contra-armónica.

$$\hat{f}(i,j) = \left\{ \sum_{x,y \in V_{ij}} \frac{g(i+x, j+y)^{P}}{L} \right\}^{1/P} \tag{3.17}$$

- *Media geométrica.* Es otro tipo de filtros de promedio no lineal que se define como el producto de L puntos pertenecientes al vecindario elevados a $1/L$. Es adecuado para eliminar ruido gausiano y preservar bordes, pero se ve afectado por impulsos negativos.

$$\hat{f}(i,j) = \prod_{x,y \in V_{ij}} g(i+x, j+y)^{1/L} \tag{3.18}$$

Otra gama de técnicas de restauración procede del *análisis por ondículas (wavelets).* Una ondícula o wavelet es una onda de una duración limitada y con una media cero. La transformada de ondículas consiste en realizar una convolución de una ondícula a diferentes escalas y en diferentes posiciones respecto a la señal a transformar. El resultado de esta transformación es un espacio escala-tiempo o escala-posición que contiene los coeficientes que caracterizan a la señal. Cada coeficiente indica lo mucho que se parece una ondícula de una determinada escala a la señal en un determinado lugar de esta. A diferencia de la transformada de Fourier donde no se sabe la localización en el tiempo de las componentes de una determinada frecuencia, o de la representación en el tiempo donde se puede ver la localización pero no la frecuencia de una señal, la transformada de ondículas permite determinar la escala (o frecuencia) que existe en una determinada localización espacial o temporal. Por tanto podemos decir que la transformación permite disponer de la información contenida en la señal de una forma más utilizable y manejable [142].

La idea básica del filtrado utilizando la transformada de ondículas es similar al utilizado en las transformadas de Fourier. Se asume que el ruido tiene componentes de poca energía o amplitud y se da a escalas reducidas, por tanto el método consiste en eliminar los coeficientes de menor amplitud aplicando un umbral en los coeficientes correspondientes a las escalas refinadas de la señal. Posteriormente se realiza la transformada inversa reconstruyendo la señal que ha perdido parte del ruido. El problema del filtrado con ondículas está en que no eliminan ruido impulsional y existe un enorme rango de familias de ondículas, no siendo evidente determinar cual de ellas es la más adecuada para una aplicación [35].

También se han aplicado *redes neuronales* en la eliminación de ruido de imágenes. Greenhill y Davies [69], utilizan una red neuronal de tres capas, con una capa de entrada de tamaño 5×5 y una sola neurona en la salida. La red neuronal se debe mover a través de toda la imagen obteniéndose en cada posición el valor restaurado correspondiente

al pixel actual. Los resultados son bastante dependientes del tipo de imágenes y del tipo de ruido presente en las imágenes utilizadas para el aprendizaje. Además, en casos prácticos en los que las imágenes deseadas no están disponibles, el entrenamiento se hace con imágenes filtradas con otros métodos de restauración, limitando la capacidad de filtrado neuronal a la capacidad de suavizado de la técnica de restauración utilizada para generar las imágenes de referencia. Varona y Villanueva [210], obtienen también resultados semejantes.

3.1.2.3 Suavizado adaptativo

Los filtros adaptativos son aquellos que presentan diferentes comportamientos cuando actúan ante discontinuidades que cuando lo hacen ante secciones continuas. Dentro de los filtros adaptativos, podemos encontrar dos soluciones: aquellos que utilizan métodos iterativos para ir realizando el filtrado y aquellos que solamente utilizan una iteración. A continuación veremos las técnicas de filtrado más relevantes, que solo requieren una iteración y que por tanto son más eficientes computacionalmente:

- *Eliminación de impulsos por contraste (CDOR).* Este filtro mide si el punto a filtrar está demasiado alejado del promedio de un *vecindario de centro borrado* (vecindario sin el punto central). Si esto es así significa que el punto es un impulso indeseable y por tanto el valor restaurado es la media del vecindario de centro borrado $\breve{g}(i,j)$. Si por el contrario hay proximidad entre el punto $g(i,j)$ y $\breve{g}(i,j)$, entonces no se produce ningún filtrado. Esta es la expresión que describe el filtro CDOR (Contrast-dependent outlier removal)

$$\hat{f}(i,j) = \begin{cases} g(i,j) & \text{si } |\frac{g(i,j)-\breve{g}(i,j)}{\breve{\sigma}}| < T \\ \breve{g}(i,j) & \text{en otro caso} \end{cases} \tag{3.19}$$

  donde es necesario seleccionar un umbral definido T y calcular la varianza del vecindario de centro borrado $\breve{\sigma}$.

- *Eliminación suavizada de impulsos por contraste (SCDOR).* Este filtro es una variación del anterior donde la decisión entre no filtrar y filtrar no se hace de una forma discreta sino de una forma suave. Por tanto el valor restaurado es una combinación del valor de entrada $g(i,j)$ y del promedio del vecindario de centro borrado $\breve{g}(i,j)$.

$$\hat{f}(i,j) = \frac{|\frac{g(i,j)-\breve{g}(i,j)}{\breve{\sigma}}|}{|\frac{g(i,j)-\breve{g}(i,j)}{\breve{\sigma}}| + k}\breve{g}(i,j) + \frac{k}{|\frac{g(i,j)-\breve{g}(i,j)}{\breve{\sigma}}| + k}g(i,j) \tag{3.20}$$

- *Pesado inverso al gradiente (GIW).* La idea de este filtro radica en que los puntos del vecindario deberían contribuir en la reconstrucción de un punto en función de lo cerca que se encuentren de él. Para ello se utilizan unos pesos $w(i,j,x,y)$ que miden esta distancia:

$$w(i,j,x,y) = \begin{cases} 1/2 & \text{si } x = 0 \text{ y } y = 0 \\ \frac{0.5/\sum(1/max(0.5,|g(i+x,j+y)-g(i,j)|))}{max\{0.5,|g(i+x,j+y)-g(i,j)|\}} & \text{en otro caso} \end{cases} \tag{3.21}$$

$$\hat{f}(i,j) = \sum_{x,y} w(i,j,x,y) g(i+x, j+y) \tag{3.22}$$

- *Media del vecindario seleccionado (SNA).* Esta técnica, a diferencia de aquellas que utilizan ventanas o vecindarios centrados respecto al punto de la imagen que queremos filtrar, utiliza para realizar el filtrado de un punto todos los posibles vecindarios que contienen a ese punto. Por ejemplo con un vecindario 3×3, dado un punto, tenemos que hay nueve vecindarios 3×3 que lo contienen. Si calculamos la varianza y la media de cada uno de los nueve vecindarios y nos quedamos con el valor medio del vecindario con menor varianza, entonces estamos realizando un filtrado SNA (Selected Neighborhood Averaging). Este filtro tiene la característica de que nunca promediará puntos que estén en lados diferentes de un borde, pues en este caso la varianza es alta y se elegirá otra región de menor varianza para realizar el promediado. Por tanto, esta técnica asume que cada punto en la imagen pertenece a una región homogénea y dicha región puede ser cubierta por un vecindario que está completamente dentro de esta región [73].

- *Filtro sigma.* El filtro sigma mira a los puntos en un vecindario de $g(i,j)$ y realiza un promedio de solo aquellos que están dentro de un intervalo $\pm 2\sigma$ centrado en $g(i,j)$.

$$\hat{f}(i,j) = \bar{g}(i+x, j+y) \; \forall \; \text{x,y} \begin{cases} (i+x, j+y) \in v_{ij} \\ g(i,j) - 2\sigma_n < g(i+x, j+y) < g(i,j) + 2\sigma_n \end{cases} \tag{3.23}$$

  Experimentos realizados por Lee [128] indican que el filtro sigma se comporta mejor ante ruido gausiano que el filtro GIW, la mediana y la media SNA. Este aspecto también lo hemos podido comprobar nosotros [109] aplicando una métrica para medir la calidad de restauración, como presentaremos más adelante en la sección 3.2.

- *MSME.* El filtro MMSE (Minimun Mean Square Error), introducido por Lee [126], hace uso del conocimiento de la varianza local σ_l para determinar si se debe realizar un filtrado promedio o se debe dejar intacta la imagen. Esta capacidad adaptativa queda claramente reflejada en su correspondiente expresión matemática:

$$\hat{f}(i,j) = \frac{\sigma_l}{\sigma_l \pm \sigma_n} g(i,j) + \frac{\sigma_n}{\sigma_l + \sigma_n} \bar{g}(i,j) \tag{3.24}$$

  donde σ_n es la varianza estimada del ruido presente en la imagen y $\bar{g}(i,j)$ es el promedio en torno a un vecindario del punto $g(i,j)$. En las zonas planas de la imagen las dos varianzas deben ser similares con lo cual el filtro tiende a ser un promedio aritmético. Sin embargo en las regiones con discontinuidades σ_l va a ser mucho mayor que σ_n con lo cual el filtro tiende a ser $\hat{f}(i,j) = g(i,j)$, es decir, en los bordes el filtro no actúa, con lo cual se preservan las discontinuidades pero también se deja pasar el ruido.

- *MSME Mejorado.* Lee [127] mejora el método *MSME* presentado por él mismo [126], incrementando el suavizado de los puntos cercanos a las discontinuidades. Para ello se calcula la dirección del borde y se promedia solo con aquellos puntos del vecindario que están al mismo lado del borde que el punto central.

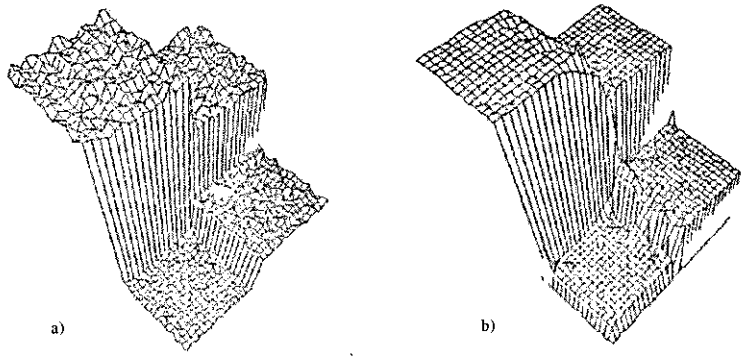


Figura 3.2: Restauración mediante el método de Grimson [70]. a) Superficie degradada b) Superficie restaurada.

- *Ajuste en regiones sin discontinuidades.* Grimson [70] plantea un método de restauración de regiones homogéneas dentro de imágenes de distancia, basado en detectar primero las discontinuidades en la imagen y posteriormente realizar un ajuste individualizado de las regiones aisladas por las discontinuidades. La idea de esta técnica se basa en que no se deben realizar ajustes de funciones sobre imágenes que constan de diversas superficies entre las cuales hay saltos, sino que los ajustes se deben hacer de forma individualizada sobre cada una de las regiones continuas de la imagen. Inicialmente se realiza un ajuste aproximado de toda la imagen y a partir de los residuos generados, que serán mayores en los bordes que en las zonas homogéneas, detecta las discontinuidades. En la imagen de residuos los pasos por cero son utilizados para detectar la correcta localización de los bordes, sin embargo, surgen problemas al aparecer pasos por ceros adicionales debidos a puntos de inflexión en la imagen a restaurar o ruidos significativos comparados con las discontinuidades entre regiones. Además las discontinuidades detectadas no son puntuales y tienen un cierto tamaño, especialmente en los bordes que no son escalones puros, con lo cual surge el problema de cómo conectar las diversas regiones aisladas entre sí. En la figura 3.2 podemos ver que en la conexión entre las cuatro regiones de que consta la imagen restaurada existe una apreciable degradación.

- *Media truncada modificada con doble ventana (DW-MTM).* El filtro DW-MTM (Double Window-Modified Trimmed Mean), fue introducido para superar las dificultades que presenta el filtro $MMSE$ ante ruido impulsional [149]. Se utiliza una nueva estimación de la media utilizando el promedio en un rango limitado entorno a la mediana. La introducción de la mediana elimina el ruido impulsional y por tanto se mejora las características del filtro promedio en presencia de este ruido. Inicialmente se calcula la mediana en un vecindario v_{ij} de tamaño $n \times n$, y a partir de este valor se realiza una media truncada pero diferente que la vista en los filtros de orden, por eso se llama media truncada modificada MTM. La técnica MTM realiza un promedio de los puntos en una ventana de tamaño $q \times q$ $(q > n)$, que están incluidos en un intervalo simétrico en torno al punto estimado

anteriormente mediante la mediana. Este intervalo viene definido por un factor proporcional k a la desviación estándar σ_n del ruido esperado. Matemáticamente podríamos expresar el filtro DW-MTM de la siguiente forma

$$\hat{f}(i,j) = \bar{g}(i+x, j+y) \ \forall \ \text{x,y} \begin{cases} -q/2 < x < q/2 \\ -q/2 < y < q/2 \\ g(i+x, j+y) > med(v_{ij}) - k\sigma_n \\ g(i+x, j+y) < med(v_{ij}) + k\sigma_n \end{cases} \tag{3.25}$$

- *Filtrado escala-espacio.* Witkin [215] introduce la técnica conocida como *filtrado escala-espacio* (scale-space filtering). El método produce una representación de una señal a diferentes escalas, la cual es utilizada para realizar la restauración de la señal conservando las formas más destacables. Inicialmente se realiza la convolución de la señal con una máscara gausiana a diferentes escalas (diferentes σ), obteniéndose una imagen escala-espacio. La detección de los pasos por cero de la segunda derivada en la imagen escala-espacio, permite obtener una imagen con los puntos de inflexión. A partir de esta imagen, se genera un árbol de rectángulos que describe la señal a lo largo del tiempo (o espacio) y a diferentes escalas (fig. 3.3). Los rectángulos largos en el eje de la escala, representan zonas limitadas por dos bordes que persisten al variar la escala, por tanto, son consideradas zonas representativas de la señal. Estas secciones de la señal son aproximadas mediante el ajuste de curvas parabólicas. La detección de los bordes persistentes que delimitan las regiones representativas es el aspecto más delicado del método y varios criterios de estabilidad pueden ser utilizados.

- *Suavizado adaptativo multiresolución (MAS).* Meer [140] presenta la técnica de restauración conocida como *MAS* (Multiresolution Adaptive Smoothing), donde por cada punto se busca la ventana centrada más grande que contenga una región plana y constante. El criterio para determinar la región que cumple esto se obtiene comparando una medida de homogeneidad local con una medida global. Si una ventana es declarada homogénea, al punto central se le asigna el valor promedio de la ventana. Si no se encuentran regiones homogéneas, entonces estamos sobre una discontinuidad y se aplica el método de suavizado adaptativo con preservación de bordes *MSME* de Lee [126]. El trabajo de Meer muestra una comparación del filtro *MAS* con otros métodos iterativos. En particular, realiza la comparación con los métodos iterativos de Saint-Marc y Perona [178, 162] para el caso de una sola iteración, y mediante criterios visuales determina que el filtro *MAS* se comporta apreciablemente mejor. A pesar de los buenos resultados, el método presenta algunas limitaciones ya que al utilizar un filtro *MSME* en presencia de discontinuidades, no hay suavizado en puntos próximos a bordes. Además, la presencia de pendientes en las superficies hace que la máscara de promediado se reduzca mucho en esos puntos, debido a que la varianza local de una región inclinada es mayor que la de una horizontal, generando suavizados pobres en este tipo de superficies.

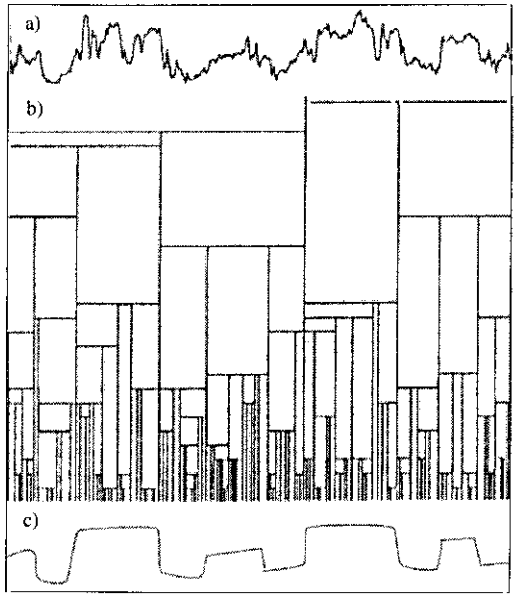


Figura 3.3: Método de restauración espacio-escala de Witkin [215]. a) Perfil degradado, b) Árbol de rectángulos, c) Perfil restaurado.

3.1.2.4 Suavizado adaptativo iterativo

Dentro de los filtros iterativos, uno de los primeros trabajos descritos en la literatura de visión artificial, es el suavizado basado en el modelo facetado de Haralick y Watson [74]. Por cada punto de la imagen, se realiza un conjunto de 9 ajustes de una superficie polinómica a los puntos contenidos en una ventana 3×3 situada en las nueve posibles posiciones conteniendo al pixel bajo consideración. De los nueve ajustes, aquel con menor residuo es utilizado para generar el valor restaurado. Siguiendo esta aproximación, los valores restaurados provienen de ajustes a regiones homogéneas, y de esta forma, en las proximidades de las discontinuidades las zonas correspondientes a saltos no intervienen en la restauración y por tanto los bordes tienden a preservarse.

Saint-Marc [178] calcula el módulo del vector gradiente en cada punto de la imagen, y a partir de este valor, genera los pesos de una máscara de tamaño 3×3 cuyos valores cambian al mover la máscara por la imagen. Los coeficientes de la máscara reflejan el grado de continuidad de la señal. Cuando mayor sea el gradiente, mayor probabilidad existe de que haya una discontinuidad, y por tanto, menor será el peso de la máscara para evitar que ese punto contribuya. Por el contrario, en zonas homogéneas, la máscara obtenida también se homogeneiza con lo cual su acción se traduce en un simple promediado sobre el vecindario. El proceso iterativo converge después de más de 250 iteraciones, requiriendo tiempos de 700 segundos en imágenes 100×100. Las superficies suaves degeneran en planos horizontales, con lo cual si se quiere preservar superficies planas con una cierta pendiente, es necesario trabajar con la derivada de la

imagen.

Yu [218] presenta un método robusto de segmentación y restauración de imágenes conocido como RESC (Residual Consensus). El proceso iterativo va cogiendo conjuntos de puntos de la imagen de forma aleatoria, con ellos, ajustando una superficie a los puntos, construye tantos histogramas de residuos como conjuntos seleccionados. Por cada histograma de residuos detecta aquel ajuste con un mayor número de puntos en el histograma y cuya suma de residuos sea mínima. Si se cumple esta condición significa que hemos cogido un grupo de puntos que difiere poco de la superficie ajustada, con lo cual el ajuste es correcto y podemos coger los parámetros de la función para realizar una restauración, segmentación o una reconstrucción. Este proceso se repite hasta que no quedan más puntos por elegir de forma aleatoria. El método RESC presenta sustanciales mejorías al ser comparado con las técnicas LMS y RANSAC.

Perona y Malik [162], realizan el filtrado basándose en la ecuación de difusión de calor (ec. 3.26), pero utilizando barreras aislantes definidas mediante diferencias de puntos en el vecindario. Estas barreras detienen la difusión del calor (o suavizado de las superficies) en las discontinuidades de la imagen. Sin embargo, es conocido que el promediado iterativo es un proceso equivalente a la difusión de calor, por tanto, la aproximación planteada por Saint-Marc [178] o la de Perona y Malik [162], conducen finalmente a soluciones muy similares. Trucco [202] plantea el problema de restauración como un problema de difusión, resolviendo la ecuación de difusión bajo ligaduras de conservación de la energía en los bordes para preservar las discontinuidades.

$$\frac{\partial calor}{\partial t} = b \cdot \nabla^2 calor \quad \text{donde } calor(i,j,t=0) = g(i,j) \tag{3.26}$$

Umasuthan y Wallace [204], presentan un método de filtrado adaptativo que aplican a la restauración de imágenes de distancia captadas por una técnica de triangulación láser. El método consta de dos etapas, la primera utiliza la técnica robusta de estimación *MLMS* (o mediana del error cuadrático mínima) que permite eliminar el ruido impulsional. La segunda etapa utiliza un suavizado no isotrópico con preservación de bordes, que está inspirado en el trabajo de difusión de Perona [162], utilizando el gradiente de la imagen desfocalizada mediante un filtro gausiano en el cálculo de la presencia de discontinuidades para frenar la difusión. El algoritmo tiene tres parámetros de control: el coeficiente de conducción, el tamaño del núcleo de suavizado y el tiempo de finalización del algoritmo, debiéndose de elegir los valores más adecuados para llegar a un compromiso en el suavizado obtenido, la preservación de bordes y el tiempo de cómputo.

3.2 Evaluación de la calidad de restauración

A pesar de las diferentes técnicas de restauración publicadas, a la hora de elegir una de ellas para utilizarla en una determinada aplicación, surge el problema de cómo seleccionar la mejor de ellas. En las comparaciones que se hacen entre diferentes métodos, en muchos casos la evaluación de los resultados se hace de forma manual observando visualmente los resultados de las distintas técnicas. Debido a que esta evaluación es muy subjetiva e imprecisa, ya que la evaluación visual capta fundamentalmente la inteligibilidad de la imagen pero no su fidelidad respecto a otra considerada como ideal, se hace necesario

la utilización de métricas analíticas que valoren la calidad de una restauración sin recurrir a la inspección visual. Además estas métricas no solo pueden ser utilizadas para evaluar sino también para la fase de diseño de una etapa de preprocesamiento como la restauración.

Entre las principales métricas analíticas podemos destacar las siguientes [169]:

- *Strehl.* Esta métrica es simplemente una medida de la reducción de contraste de una imagen reconstruida $\hat{f}(i,j)$ al compararla con otra ideal $f(i,j)$. La métrica en realidad realiza una evaluación de tipo subjetivo que encaja con una calificación visual.

$$Q = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{F}(u,v)dudv}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F(u,v)dudv} \tag{3.27}$$

- *Correlación cruzada.* Es una medida de la proximidad entre las imágenes $\hat{f}(i,j)$ y $f(i,j)$. Sin embargo, esta medida da buenos resultados incluso cuando la imagen $f(i,j)$ está degradada por un filtro paso bajo y tiene una baja calidad de tipo subjetivo.

$$Q = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F(u,v)\hat{F}^{*}(u,v)dudv}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|F(u,v)|^{2}dudv} \tag{3.28}$$

- *Error absoluto medio.*

$$E = \frac{\sum_i\sum_j|f(i,j)-\hat{f}(i,j)|}{\sum_i\sum_j 1} \tag{3.29}$$

- *Error relativo medio.*

$$E = \frac{\sum_i\sum_j\frac{|f(i,j)-\hat{f}(i,j)|}{f(i,j)}}{\sum_i\sum_j 1} \tag{3.30}$$

- *Error cuadrático medio.*

$$E = \frac{\sum_i\sum_j\{f(i,j)-\hat{f}(i,j)\}^2}{\sum_i\sum_j 1} \tag{3.31}$$

- *Diferencia media de normales [204].*

$$\bar{d\vec{n}} = \frac{\sum_i\sum_j|\arccos(\vec{n}_f\cdot\vec{n}_{\hat{f}})|}{\sum_i\sum_j 1} \tag{3.32}$$

Las métricas analíticas más frecuentemente utilizadas son el error cuadrático medio y el error absoluto. Sin embargo, ambos métodos contemplan solamente la desviación local entre los puntos de la imagen reconstruida y de la imagen ideal, pero no considera la disposición de un punto en relación con los de su vecindario. La figura 3.4 muestra un claro ejemplo donde dos reconstrucciones que no deberían ser consideradas con la misma calidad de restauración, sin embargo, tienen el mismo error cuadrático medio.

La métrica $\bar{d\vec{n}}$ mide que la orientación de las superficies sea la misma en la imagen ideal y la restaurada, y por tanto si la superficie ideal es suave premiará aquellas superficies

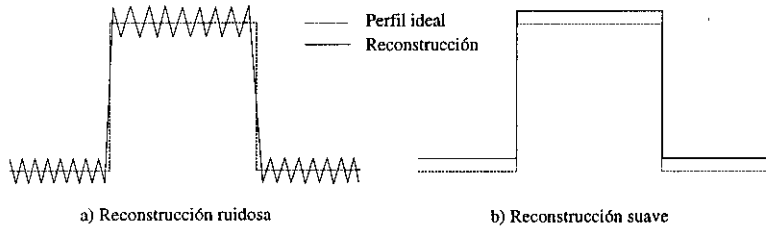


Figura 3.4: Ejemplo de dos reconstrucciones con el mismo error cuadrático medio

que también lo sean. Sin embargo, como norma general no mide la fidelidad en términos de la separación entre la imagen ideal y la restaurada, y dos superficies perfectamente paralelas separadas por un desplazamiento constante no es posible detectarlo con esta métrica de forma aislada.

Otros métodos de evaluación se basan en la observación de los resultados después de pasar la imagen restaurada por alguna etapa posterior de procesamiento, como una extracción de bordes o una votación en un espacio paramétrico de Hough. Sin embargo en este caso se corre el riesgo de elegir un método de restauración que funciona óptimamente siempre y cuando se mantengan fijas las etapas posteriores de procesamiento con las que se realizó la evaluación. Por tanto, si el sistema de procesado se modifica, la etapa de restauración no tiene por qué seguir siendo la más adecuada y por tanto el método está muy fuertemente ligado con una determinada aplicación.

En vista de la falta de estandarización y poca fiabilidad para determinar la fidelidad de una restauración mediante los métodos actuales, vemos la necesidad de definir una métrica analítica que nos dé una medida objetiva que nos establezca un criterio para poder seleccionar el mejor método de restauración dentro de un conjunto de posibles candidatos. A continuación presentaremos una métrica que considera tanto los residuos puntuales como las relaciones entre puntos, para valorar positivamente a aquellas técnicas que no introducen distorsiones y además suavizan las superficies.

3.2.1 Definición de una nueva métrica de calidad de restauración: *GRI*

El *error cuadrático medio* o el *error absoluto medio* son excelentes métricas para medir la desviación de la señal restaurada frente a la señal original. En zonas donde un borde no se preserva o un ruido impulsional no es eliminado aparece un residuo elevado, con lo cual ambas métricas penalizan las distorsiones no eliminadas por el método de restauración. La evaluación visual permite medir de una forma poco fiable las distorsiones en discontinuidades, sin embargo capta mejor la textura física de las superficies continuas. Si las superficies homogéneas contienen aún ruido sin eliminar, la apreciación visual no será positiva. Esta misma valoración de la suavidad se podría hacer utilizando la métrica que mide la diferencia en las normales, sin embargo, en las regiones suaves en las que no se preservan adecuadamente los bordes se produce una fuerte penalización debido a la pérdida de paralelismo entre las superficies restaurada e ideal. Esto hace que esta métrica

no solo mida suavidades sino que también en determinadas regiones mide fidelidades y por tanto es más difícil de interpretar.

La idea principal de la métrica que vamos a presentar, combina las características de las evaluaciones visuales que captan la suavidad o rugosidad, con la de evaluaciones basadas en residuos que focalizan su atención en la fidelidad. Por tanto, nuestra métrica va a tratar de valorar positivamente la fidelidad en la preservación de discontinuidades, y a la vez, la suavidad o ausencia de rugosidad en las zonas que deben ser homogéneas. Para ello vamos a combinar dos métricas que miden de forma aislada e individual la fidelidad y la suavidad.

Para medir la fidelidad en la restauración, y en particular la preservación de discontinuidades, definimos la métrica de fidelidad de restauración *FI* (Fidelity Index) mediante una exponencial del promedio de los errores absolutos entre la imagen ideal $f(i,j)$ y la imagen restaurada $\hat{f}(i,j)$ (ec. 3.33). El motivo de utilizar la función no lineal "exponencial negativa" se debe por un lado a que normaliza el índice de medida entre 0 y 1, y por otro a que aumenta la capacidad de discriminación en los valores próximos a 1 y hace lo contrario en los valores con alto residuo a los cuales les corresponde siempre valores muy próximos a cero, con lo cual se agrupa todas las técnicas con preservaciones malas, muy malas ó pésimas en un mismo grupo que indica que todas ellas son inadecuadas ($FI \simeq 0$).

$$FI = e^{-\left(\sum_{i,j=1}^{N,N} |f(i,j)-\hat{f}(i,j)|\right)/N^2} \tag{3.33}$$

Para medir el suavizado de la imagen, utilizamos la exponencial de un factor de rugosidad ξ, con lo que la métrica *SI* (Smoothing Index) la definimos así:

$$SI = e^{-\xi} \tag{3.34}$$

El término que mide la rugosidad, ξ, se obtiene realizando el promedio de las diferencias entre los incrementos angulares $d\alpha$, correspondientes al cambio de las normales superficiales de dos puntos vecinos tanto para la imagen ideal como para la reconstruida (fig. 3.5). Como consecuencia lo que estamos midiendo con ξ es básicamente la fluctuación media de las normales de la imagen restaurada, ya que en el caso de la imagen ideal esta fluctuación debe ser nula si asumimos que una superficie ideal es suave y por tanto sin rugosidad. Los incrementos angulares de los vecinos en las direcciones perpendiculares x e y , los calculamos mediante las siguientes ecuaciones:

$$d\alpha_{fx} = \arccos\left(\frac{\vec{n_f}(i,j)\cdot\vec{n_f}(i-1,j)}{\|\vec{n_f}(i,j)\|\cdot\|\vec{n_f}(i-1,j)\|}\right) \tag{3.35}$$

$$d\alpha_{fy} = \arccos\left(\frac{\vec{n_f}(i,j)\cdot\vec{n_f}(i,j-1)}{\|\vec{n_f}(i,j)\|\cdot\|\vec{n_f}(i,j-1)\|}\right) \tag{3.36}$$

$$d\alpha_{\hat{f}x} = \arccos\left(\frac{\vec{n_{\hat{f}}}(i,j)\cdot\vec{n_{\hat{f}}}(i-1,j)}{\|\vec{n_{\hat{f}}}(i,j)\|\cdot\|\vec{n_{\hat{f}}}(i-1,j)\|}\right) \tag{3.37}$$

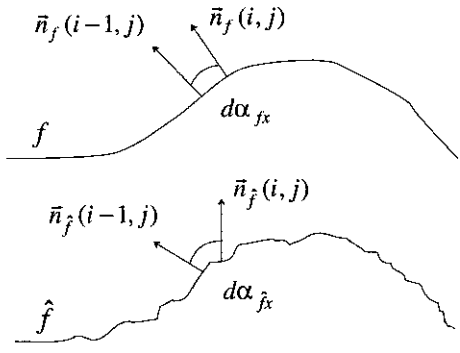


Figura 3.5: Incrementos angulares entre las normales de puntos vecinos en la dirección x para un perfil ideal y el mismo restaurado.

$$d\alpha_{\hat{f}y} = \arccos\left(\frac{\vec{n_{\hat{f}}}(i,j)\cdot\vec{n_{\hat{f}}}(i,j-1)}{\|\vec{n_{\hat{f}}}(i,j)\|\cdot\|\vec{n_{\hat{f}}}(i,j-1)\|}\right) \tag{3.38}$$

De esta forma, definimos el factor de rugosidad ξ, mediante la siguiente expresión:

$$\xi = \left(\sum_{i,j=1}^{N,N}(d\alpha_{\hat{f}x} - d\alpha_{fx})\cdot(d\alpha_{\hat{f}x} \geq d\alpha_{fx}) + (d\alpha_{\hat{f}y} - d\alpha_{fy})\cdot(d\alpha_{\hat{f}y} \geq d\alpha_{fy})\right)/(2N^2) \tag{3.39}$$

en donde los términos condicionales $d\alpha_{\hat{f}x} \geq d\alpha_{fx}$ y $d\alpha_{\hat{f}y} \geq d\alpha_{fy}$ aparecen para no considerar los incrementos angulares en las zonas de transición de la imagen ideal como si fuesen posibles rugosidades, con lo cual la rugosidad en la imagen restaurada solo se calcula en las zonas donde la imagen ideal sea continua.

Nótese que los valores de ambas métricas, *FI* y *SI*, están acotados entre 0 y 1. Correspondiendo un valor de 1 para el caso de fidelidad y suavizado perfectos, mientras que valores próximos a cero indican una mala calidad en la reconstrucción. La integración de las dos métricas en una única que mida la calidad de reconstrucción global contemplando tanto fidelidad como suavizado de una forma conjunta, la podemos obtener aplicando una media geométrica de los índices *FI* y *SI*:

$$GRI = \sqrt{FI \cdot SI} \tag{3.40}$$

donde *GRI* es la métrica que indica la calidad global de restauración (Global Restauration Index), que sigue teniendo un rango de valores posibles entre 0 y 1. Técnicas de restauración con buenas calidades de suavizado, pero mala preservación de bordes presenta un *FI* cercano a cero y un *SI* cercano a 1. Un simple promedio entre *FI* y *SI* daría un valor de 0.5 indicando que el método de restauración es intermedio. Sin embargo, utilizando la integración multiplicativa de la ecuación 3.40, si una de las dos métricas genera valores próximos a cero la penalización en la métrica global *GRI* va a ser

Método de filtrado	*GRI*	*FI*	*SI*
DW-MTM (σ_n = 1, L=3, k=2, q=5)	0.69	0.62	0.76
MAS (σ_n = 1, L=7,5,3)	0.68	0.61	0.76
MMSE (σ_n = 1, L=3)	0.66	0.59	0.74
Sigma (σ_n = 1, L=5)	0.65	0.61	0.70
Mediana (L=3)	0.61	0.57	0.66
Wavelet (Daubechies, 3, Suave)	0.60	0.54	0.67
SNA (L=3)	0.57	0.52	0.64
GIW (L=5)	0.55	0.56	0.54
SCDOR (L=5,k=3)	0.51	0.47	0.55
FIR (L=3)	0.51	0.33	0.79
Media truncada (L=5, M=1)	0.50	0.32	0.79
Media aritmetica (L=3)	0.49	0.37	0.65
Media Geométrica (L=3)	0.49	0.36	0.65
Media Y_p (L=3, P=2)	0.48	0.36	0.65
Gausiano (L=7)	0.48	0.30	0.77
Media armónica (L=3)	0.48	0.35	0.65
Contrarmónico (L=3, P=2)	0.45	0.31	0.66
CDOR (L=5, T=5)	0.44	0.47	0.42
Rango medio (L=5)	0.32	0.15	0.68

Tabla 3.1: Clasificación de métodos clásicos de restauración según métrica *GRI* (ruido gausiano).

significativa. Esto concuerda con el propósito marcado inicialmente, es decir, la métrica va a premiar aquellas técnicas que realizan reconstrucciones fieles, y a la vez, generan superficies suaves. Si alguno de estos factores no se cumple la calidad de la reconstrucción disminuirá considerablemente.

3.2.2 Evaluación de técnicas de restauración mediante la métrica *GRI*

Una vez definida la métrica *GRI* estamos en condiciones de comparar diferentes tipos de técnicas de restauración. Para ello hemos generado un perfil sintético que contiene una muestra variada de diferentes tipos de superficies y discontinuidades (fig. 3.6a). Sobre este perfil hemos añadido ruido gausiano con una desviación estándar de 1 mm (fig. 3.6b). Varias técnicas de restauración no iterativas fueron aplicadas, y la calidad de restauración utilizando la métrica *GRI* fue calculada. Para evitar posibles variaciones de la métrica ante determinadas configuraciones del ruido sobre el perfil, se realizó el promediado de *GRI* para 100 muestras diferentes de perfiles contaminados con ruido gausiano. En la tabla 3.1 se presentan los diferentes métodos de restauración ordenados según el índice de calidad *GRI* que obtuvimos al analizar los perfiles restaurados por cada técnica, mostrándose también las componentes de calidad parciales *FI* y *SI*.

Podemos observar en la tabla 3.1 que los mejores métodos según la métrica *GRI* son la media truncada con doble ventana *DW-MTM*, la técnica de suavizado adaptativo

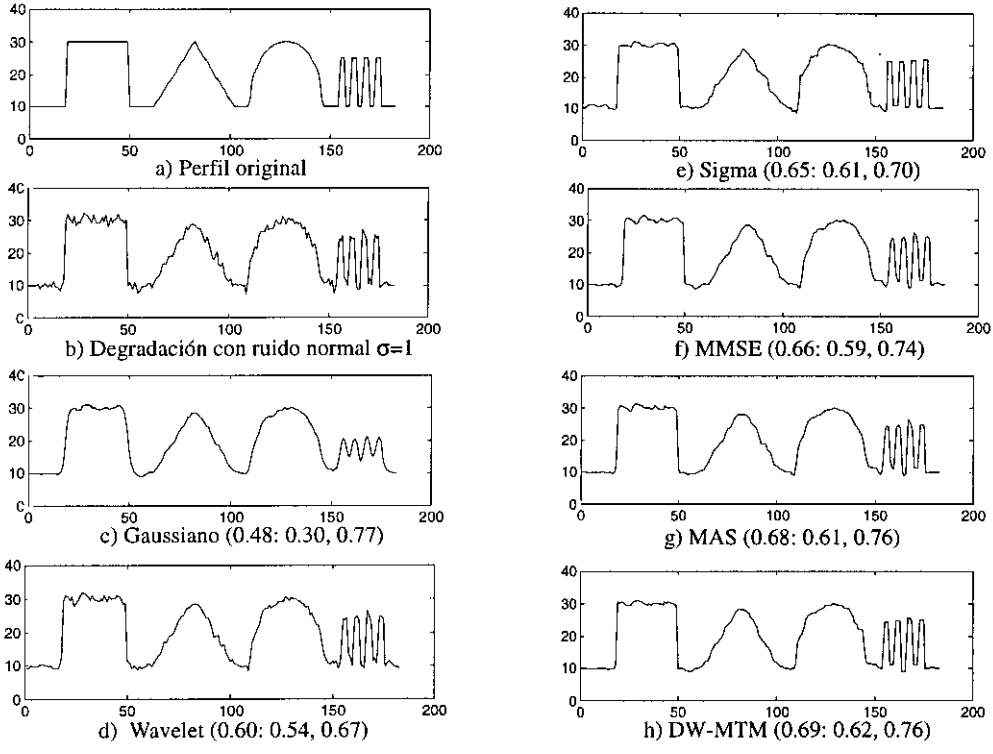


Figura 3.6: Reconstrucción de un perfil contaminado con ruido gausiano mediante algunas de las técnicas de restauración clásicas. Entre paréntesis se indica las métricas *GRI*, *FI* y *SI* respectivamente.

multiresolución MAS, y el filtro de mínimo error cuadrático medio *MMSE*, con índices de 0.69, 0.68 y 0.66 respectivamente. En las figuras 3.6h,g,f se puede apreciar uno de los cien perfiles restaurados por las técnicas *DW-MTM*, *MAS* y *MMSE* respectivamente. Observando los perfiles y sus índices parciales *FI* y *SI*, podemos ver que hay una buena preservación de bordes y una apreciable mejoría en el suavizado de las tres restauraciones. También mostramos en las figuras 3.6e,d,c el perfil correspondiente a un filtrado utilizando el filtro sigma, un máscara gausiana y ondículas Daubechies de orden 3 con umbralización suave. Podemos observar en la tabla 3.1 que la eliminación de ruido mediante ondículas no preserva tan bien los bordes como en las anteriores técnicas y el suavizado tampoco es acentuado, presentando un índice *GRI* de 0.60. El filtrado gausiano realiza un aceptable suavizado ($SI = 0.77$), sin embargo la preservación de bordes no es respetada ($FI = 0.30$), con lo cual el índice de calidad global *GRI* se reduce a 0.48.

Es necesario remarcar que la métrica *GRI* es apropiada para la comparación de técnicas de restauración utilizando las discrepancias *relativas* entre los valores de calidad obtenidos, sin embargo, los valores *absolutos* de la métrica no deben ser considerados.

Esto es así pues los valores obtenidos dependen de la amplitud y tipo de ruido presente en la imagen, y de la propia imagen de prueba utilizada. Por tanto un valor absoluto *GRI* obtenido al restaurar un perfil contaminado con ruido como el de la figura 3.6b, no debe ser comparado con otro valor de *GRI* obtenido al reconstruir otro tipo de imagen o perfil, o con otro tipo de ruido. La forma más adecuada de realizar una comparación consiste en utilizar una imagen de prueba representativa conteniendo un amplio rango de diferentes tipos de discontinuidades y superficies, y con un particular tipo de ruido. Bajo estas condiciones, y analizando las diferencias relativas en *GRI*, la métrica permite clasificar las técnicas de restauración de una forma precisa.

Obsérvese que existe una perfecta correspondencia entre el concepto subjetivo que tenemos de suavidad y los valores de *SI*. Así la restauración gausiana y la generada a partir de ondículas (wavelets) presentan visualmente una clara diferencia en cuanto a la suavidad de sus perfiles, lo cual es puesto de manifiesto por la métrica *SI* al generar los valores 0.77 y 0.67, respectivamente.

Hemos visto que las mejores técnicas de restauración para eliminar ruido gausiano corresponden a filtros adaptativos, como MAS, DW-MTM y MMSE. Por tanto, cualquiera de los tres podría ser utilizado para realizar la restauración de las imágenes que queremos mejorar. Sin embargo, aunque los resultados de los mejores filtros analizados no son malos, se echa en falta una mayor suavidad en los perfiles restaurados. En algunos casos como el MMSE esto se puede explicar ya que el objetivo marcado en su diseño era minimizar el error cuadrático medio, pero no se considera en ningún caso la suavidad o conexión continua entre cada uno de los puntos individuales. Es indudable que cualquier sistema de procesamiento modularizado en una secuencia de etapas, funcionará mejor cuando cada una de ellas entregue los mejores resultados posibles. En nuestro caso, estamos interesados en una etapa de preprocesamiento de imágenes lo mejor posible para facilitar la labor a las etapas posteriores de análisis de las imágenes. Mejores resultados de restauración que los presentados se podrían obtener mediante técnicas iterativas, sin embargo, el alto coste computacional nos impide utilizar estas técnicas para aplicaciones que pretenden operar en tiempo real. Por estos motivos, en la siguiente sección vamos a presentar un nuevo método no iterativo de restauración, que como podremos comprobar mejora, según la métrica *GRI* definida, las técnicas analizadas en la tabla 3.1.

3.3 Nuevo método de restauración adaptativa por ajustes de planos multiresolución con fidelidad-3σ

En esta sección pretendemos definir un filtro adaptativo que sea capaz de restaurar imágenes eliminando la mayor parte del ruido degradante, y a la vez, conservando las discontinuidades presentes. Es decir, queremos eliminar la información que no nos interesa (ruido) y conservar aquella información que nos permita captar la forma de los objetos en una imagen (superficies homogéneas y discontinuidades significativas). Para terminar de definir el problema vamos a especificar lo que entendemos por superficies homogéneas, discontinuidades y ruido:

- Las *superficies homogéneas* de que consta una imagen asumimos que se caracterizan por ser suaves, es decir, no presentan rugosidad, pudiendo ser planas o curvas y

quedando limitadas por discontinuidades.

- Una *discontinuidad* se considera significativa, y por tanto hay que preservarla, si es debida a saltos en la propia escena de mayor amplitud que el ruido presente en la imagen.

- El *ruido* presente es de tipo aditivo pudiendo presentar distribuciones de tipo gausiano, impulsional o una combinación de ambos.

Una vez presentados los objetivos y las asunciones que definen la imagen a procesar, estamos en condiciones de presentar nuestra técnica de restauración.

3.3.1 Definición de la técnica de restauración 3σ-*MPF*

A la hora de diseñar un filtro adaptativo, existen dos grandes problemas:

- *¿Cómo detectar las discontinuidades?.* Por definición un filtro adaptativo trabaja de forma diferente en unos puntos de la imagen que en otros, siendo esta adaptación función de las características locales de la imagen. Una forma de determinar cuando aplicar un modo de filtrado u otro se puede hacer detectando discontinuidades. De esta forma el filtro puede suavizar en las regiones continuas y evitar el suavizado sobre los bordes. En muchos casos la forma de detectar discontinuidades se hace comparando la varianza local calculada en una ventana con la varianza global estimada en toda la imagen. Si la varianza local supera a la varianza global se considera que estamos ante una discontinuidad y en caso contrario se asume que estamos en una región continua. El problema de esta solución estriba en que las estadísticas locales calculadas en regiones pequeñas no son muy fiables con lo cual existe un cierto riesgo de realizar mal la detección de la discontinuidad.

- *¿Cómo determinar el tamaño de la ventana de procesamiento?.* Cuanto mayor sea la ventana del filtro que vamos moviendo a lo largo de la imagen, mejor es el suavizado que se puede conseguir cuando se trata de regiones continuas. Sin embargo, si la ventana es grande la probabilidad de que una discontinuidad esté presente también crece, y por el contrario, la capacidad de detectar la discontinuidad decrece. Por tanto existe un compromiso que limita la elección de la ventana de procesamiento que se suele resolver, en la mayoría de los casos, eligiendo ventanas 3×3.

Nosotros proponemos utilizar un criterio que mida la fidelidad de ajuste de un plano a un conjunto de datos que pertenecen a la imagen, para decidir dónde hay una discontinuidad y determinar el tamaño de la ventana de procesamiento más adecuada. Este criterio de fidelidad del ajuste lo llamamos *fidelidad*-3σ y se puede definir mediante la siguiente expresión

$$fidelidad-3\sigma = \begin{cases} Verdadero & \text{Si } \forall (x,y) \in v_{ij} \ |g(x,y) - (ax+by+c)| \leq 3\sigma \\ Falso & \text{En caso contrario} \end{cases} \quad (3.41)$$

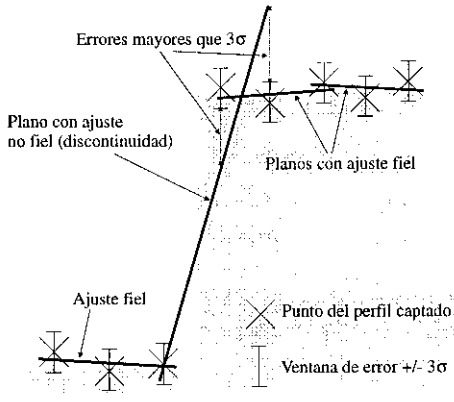


Figura 3.7: Detección de discontinuidades utilizando el ajuste con *fidelidad-3σ*.

siendo σ la desviación estándar del ruido aditivo esperado sobre la imagen, y a, b y c los parámetros del plano ajustado. Es decir, después de realizar un ajuste de unos datos a un plano diremos que se realizó un ajuste con *fidelidad-3σ* si "todos los puntos en la ventana están a una distancia de los respectivos puntos del plano ajustado, menor o igual que 3σ".

Mediante este criterio de ajuste podremos determinar que existe una *discontinuidad* cuando no sea posible realizar un ajuste de *fidelidad-3σ* sobre una ventana de tamaño 3×3 (fig. 3.7). Por el contrario si la región es homogénea, aunque esté contaminada por ruido gausiano, siempre existirá una alta probabilidad de que una región se pueda ajustar de una forma fiel a un plano, incluso aunque la superficie tenga una cierta curvatura. Por otro lado, el criterio de *fidelidad-3σ* también lo utilizaremos para elegir el *tamaño de la ventana de procesamiento*, la cual será variable y se podrá ir adaptando a las características de la imagen.

Teniendo presentes las consideraciones anteriores, el algoritmo *3σ-MPF (3σ-Multiresolution Plane Fitting)* que proponemos se basa en un ajuste multiresolución de planos sobre regiones homogéneas, que evoluciona hacia un método de preservación de bordes en las regiones con discontinuidades. La idea consiste en suavizar tanto como sea posible en las regiones homogéneas donde son factibles los ajustes sobre grandes regiones. Si el ajuste no es satisfactorio, probablemente debido a regiones con curvatura o a la proximidad de bordes, se reducirá la región de aplicación hasta conseguir un ajuste fiel. Finalmente, si después de reducir las regiones de aplicación no se consigue un ajuste con *fidelidad-3σ*, significará que estamos sobre una discontinuidad y se deberá aplicar un método de preservación de bordes.

En la figura 3.8 se muestra el pseudocódigo del algoritmo *3σ-MPF*, donde se puede apreciar que para realizar el ajuste multiresolución utilizamos secuencialmente tres ventanas de tamaños: 7×7, 5×5 y 3×3. Comenzando por la ventana de mayor tamaño, ajustamos un plano a los puntos de la imagen contenidos dentro de dicho ventana y evaluamos el criterio *fidelidad-3σ*. Si hay algún punto que no cumple este criterio, se repite el proceso con la siguiente ventana de menor tamaño. Finalmente, cuando se

```
POR cada punto de la imagen g(i,j)
    σ=estimación_ruido(g,(i,j));
    tamaño_ventana=7x7;
    REPEAT
        a,b,c=ajuste_plano_minimos_cuadrados(g,(i,j),tamaño_ventana);
        test_fidelidad_3σ=fidelidad_3σ(g,(i,j),tamaño_ventana,(a,b,c),σ);
        IF test_fidelidad_3σ==FALSE
            tamaño_ventana=tamaño_ventana-2x2;
    UNTIL (test_fidelidad_3σ==TRUE OR tamaño_ventana<3x3);
    IF test_fidelidad_3σ==TRUE   // región continua encontrada
        f̂(i,j)=c;
    ELSE                         // discontinuidad o impulso detectado
        ref=Mediana(g,(i,j),3x3,conectividad_4);
        f̂(i,j)=media_rango_3σ(g,(i,j),3x3,conectividad_8,|g(x,y)-ref|<3σ)
    END
```

Figura 3.8: Pseudocódigo del método de restauración 3σ-*MPF*.

obtenga un ajuste con *fidelidad*-3σ, el valor de la imagen a filtrar correspondiente al centro de la ventana es sustituido por la media de los puntos contenidos en esa ventana.

En el caso de que la mínima ventana de 3×3 no genere ajustes fieles, aplicamos la técnica MTM (Modified Trimmed Mean) que combina el algoritmo de los filtros de orden con promedio truncado (trimmed mean) y la selección de candidatos a promediar que se usa en el filtro sigma. Mediante el valor central y sus 4 vecinos de conectividad 4, estimamos un valor inicial calculando la mediana de estos 5 valores. Este valor estimado es utilizado para realizar un promediado con aquellos puntos que se encuentren dentro del intervalo $\pm 3 \cdot \sigma$ en una ventana 3×3 de conectividad 8. De esta forma se consigue preservar bordes, reducir ruido gausiano al realizar un promediado y a la vez eliminar ruido de tipo impulsional.

La desviación estándar σ del ruido aditivo esperado sobre la imagen, no necesita ser estimada calculando la varianza en una ventana, cuya fidelidad de estimación se degrada a medida que la ventana se hace más pequeña. Para obtener esta estimación utilizamos el modelo de repetitividad σ_D deducido en el capítulo 2 (ec. 2.28 ó 3.42), que permite obtener una buena aproximación del ruido aditivo presente utilizando la imagen de reflectancia $\Re$. Por comodidad lo repetimos a continuación:

$$\sigma_D = \sqrt{\frac{8.37 \cdot 10^{-3}}{T \cdot 10^{\Re/80.9}} + \left(\frac{D_{max}}{9 \cdot 10^7 T}\right)^2 + 1} \tag{3.42}$$

Debido a que el método de filtrado presentado no pretende ser exclusivo para ser aplicado sobre parejas de imágenes Distancia-Reflectancia, en el caso de no disponer de una imagen de reflectancia $\Re$ para estimar el ruido, la técnica utilizada por Meer [140] parece más adecuada que una simple estimación local de varianzas. El método que

propone Meer calcula la varianza media de toda la imagen mediante la detección del pico del histograma de varianzas que es normalmente unimodal, obteniéndose resultados de estimación bastante robustos.

El filtro que hemos presentado, 3σ-*MPF*, está inspirado en las mejores cualidades de los filtros MAS y DW-MTM. Las debilidades manifestadas en ambos diseños han sido superadas añadiendo la estrategia de fidelidad-3σ que se basa en un test sobre los residuos de un ajuste por mínimos cuadrados de un plano. De esta forma es de esperar que el comportamiento de este nuevo filtro sea superior a los ya reconocidos filtros MAS y DW-MTM.

Una de las novedades del filtro 3σ-*MPF* reside en que no se necesita establecer ningún umbral de forma arbitraria para determinar la presencia de discontinuidades. La definición de ajuste con *fidelidad*-3σ, es el criterio utilizado para determinar si está presente una discontinuidad, no necesitándose realizar ninguna sintonía de parámetros. Además, no es necesario la estimación de ninguna estadística sobre ventanas locales, como suele ser habitual en otras técnicas para determinar la varianza local de una región, siendo esta aproximación poco precisa al estimarse estadísticas sobre un reducido conjunto de datos. El problema de la selección del tamaño de la máscara de filtrado queda también resuelto en el filtro 3σ-*MPF* mediante la aproximación multiresolución al ser aplicada junto con el criterio de ajuste con *fidelidad*-3σ, el cual fuerza a reducir el tamaño de la ventana si no se cumple dicho criterio. De esta forma somos capaces de suavizar utilizando ventanas grandes donde es posible y al mismo tiempo determinar de una forma fiable la presencia de un borde.

3.3.2 Evaluación comparativa de la técnica de restauración 3σ-*MPF* por la métrica *GRI*

Una vez presentada la técnica de restauración 3σ-*MPF*, vamos a evaluarla utilizando la métrica *GRI* presentada en la sección 3.2. En el caso de degradación por ruido gausiano, si aplicamos esta técnica de restauración al perfil contaminado de la figura 3.9b, obtenemos el perfil mostrado en la figura 3.9i, el cual lo podemos comparar con las restauraciones realizadas por los métodos clásicos (fig. 3.9c-h). Podemos apreciar visualmente que el método que proponemos es de calidad superior al resto, especialmente en términos de suavidad, siendo solo comparable en este sentido con el filtrado gausiano. Sin embargo nuestro método también consigue obtener una buena preservación de bordes, como puede verse en la tabla 3.2, donde el índice *FI* es de 0.64. Adicionalmente como el índice de suavidad *SI* es elevado, 0.87, el índice global *GRI* es de 0.75, superando claramente al resto de técnicas presentadas.

Ante ruido impulsional el filtro diseñado también presenta un buen comportamiento aunque no resulta tan efectivo como para eliminar ruido gausiano. En la tabla 3.3 se puede apreciar cómo la calidad de restauración ante este tipo de ruido, medido con *GRI*, permite incluir el método 3σ-*MPF* entre una de las mejores alternativas para eliminar este tipo de ruido. En la figura 3.10 se aprecian algunos de los perfiles restaurados mediante los métodos analizados. Evidentemente, si una imagen está contaminada únicamente con ruido impulsional un método como *CDOR* o un filtro *Mediana* serían los más adecuados, sin embargo, esto no es habitual y lo normal es que además aparezca siempre un fondo

Método de filtrado	*GRI*	*FI*	*SI*
3σ-MPF (σ_n = 1, L=7,5,3)	0.75	0.64	0.87
DW-MTM (σ_n = 1, L=3, k=2, q=5)	0.69	0.62	0.76
MAS (σ_n = 1, L=7,5,3)	0.68	0.61	0.76
MMSE (σ_n = 1, L=3)	0.66	0.59	0.74
Sigma (σ_n = 1, L=5)	0.65	0.61	0.70
Mediana (L=3)	0.61	0.57	0.66
Wavelet (Daubechies, 3, Suave)	0.60	0.54	0.67
SNA (L=3)	0.57	0.52	0.64
GIW (L=5)	0.55	0.56	0.54
SCDOR (L=5,k=3)	0.51	0.47	0.55
FIR (L=3)	0.51	0.33	0.79
Media truncada (L=5, M=1)	0.50	0.32	0.79
Media aritmetica (L=3)	0.49	0.37	0.65
Media Geométrica (L=3)	0.49	0.36	0.65
Media $\bar{Y}_p$ (L=3, P=2)	0.48	0.36	0.65
Gausiano (L=7)	0.48	0.30	0.77
Media armónica (L=3)	0.48	0.35	0.65
Contrarmónico (L=3, P=2)	0.45	0.31	0.66
CDOR (L=5, T=5)	0.44	0.47	0.42
Rango medio (L=5)	0.32	0.15	0.68

Tabla 3.2: El filtro 3σ-*MPF* evaluado según la métrica *GRI*, y comparado con otras técnicas para eliminar ruido *gausiano*.

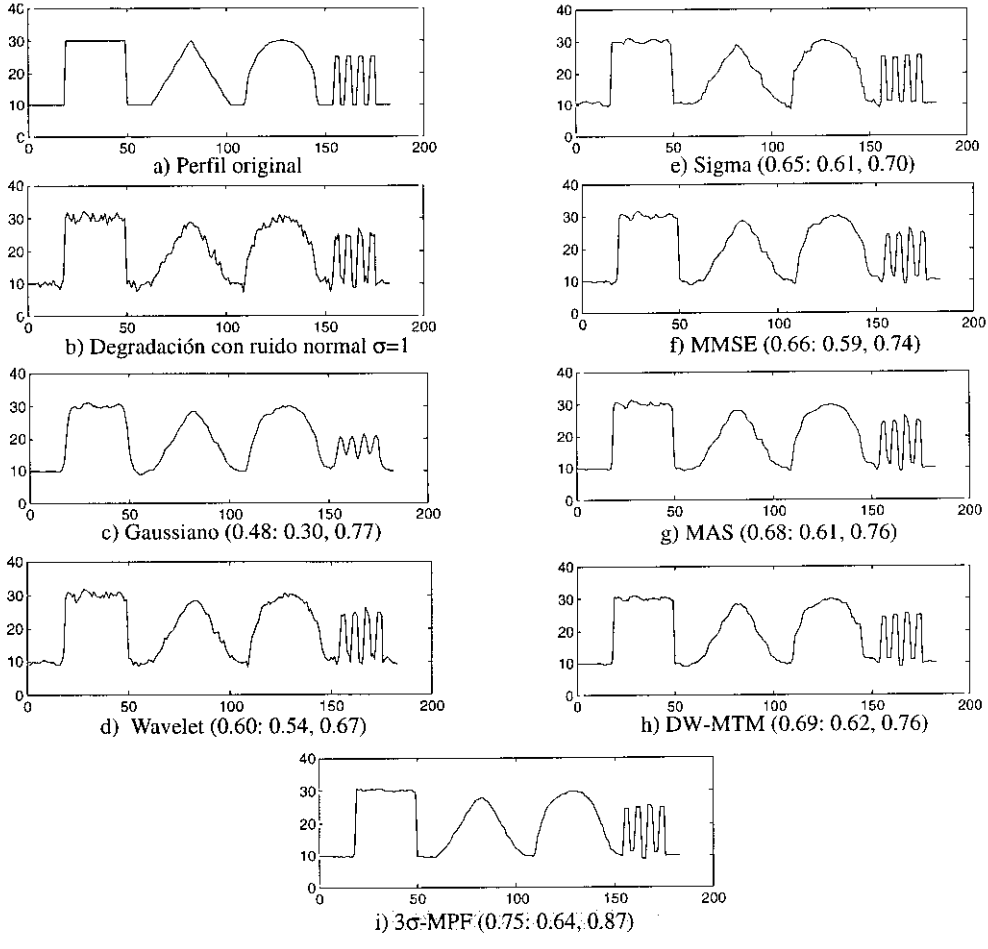


Figura 3.9: Perfiles contaminados con ruido *gausiano* y restaurados mediante 3σ-*MPF* y otras técnicas clásicas. Entre paréntesis se indica las métricas *GRI*, *FI* y *SI* respectivamente.

Método de filtrado	*GRI*	*FI*	*SI*
CDOR (L=5, T=5)	0.97	0.96	0.99
Mediana (L=3)	0.91	0.88	0.94
DW-MTM (σ_n = 1, L=3, k=2, q=5)	0.88	0.83	0.93
3σ-MPF (σ_n = 1, L=7,5,3)	0.84	0.75	0.94
Media truncada (L=5, M=1)	0.59	0.39	0.89
GIW (L=5)	0.52	0.35	0.79
SNA (L=3)	0.51	0.33	0.77
Media Geométrica (L=3)	0.44	0.26	0.73
Wavelet (Daubechies, 3, Suave)	0.33	0.14	0.78
Sigma (σ_n = 1, L=5)	0.32	0.13	0.76
MMSE (σ_n = 1, L=3)	0.32	0.13	0.75
MAS (σ_n = 1, L=7,5,3)	0.31	0.13	0.74
Media Y_p (L=3, P=2)	0.27	0.10	0.73
SCDOR (L=5, k=3)	0.27	0.11	0.66
Media aritmética (L=3)	0.24	0.08	0.71
FIR (L=3)	0.23	0.07	0.74
Gausiano (L=7)	0.22	0.07	0.74
Contrarmónico (L=3, P=2)	0.11	0.02	0.71
Rango medio (L=5)	0.04	0.003	0.73
Media armónica (L=3)	0.04	0.002	0.72

Tabla 3.3: El filtro 3σ-*MPF* evaluado según la métrica *GRI*, y comparado con otras técnicas para eliminar ruido *impulsional*.

de ruido gaussiano o uniforme.

En el caso de imágenes contaminadas con ruido gausiano e impulsional, el método 3σ-*MPF* se muestra de nuevo como el más atractivo de todos. Siendo incluso más recomendable que filtros que operan en dos etapas donde en un primer paso se aplica un filtro *CDOR* o *Mediana* para eliminar la parte impulsional, y a continuación un filtro *MAS*, *DW-MTM* o *MMSE* para eliminar la componente gausiana del ruido (tabla 3.4 y fig. 3.11). En estos casos de dos etapas, la degradación en la preservación de bordes se acumula al pasar por ambas etapas, obteniéndose mejores resultados al aplicar el filtro 3σ-*MPF* el cual consta de una sola etapa.

Para apreciar cualitativamente los resultados de las restauraciones en imágenes completas y no solamente sobre perfiles, vamos a aplicar el método de restauración 3σ-*MPF* sobre imágenes de distancias sintéticas contaminadas con ruido gausiano, ruido impulsional y una combinación de ambos ruidos. En la figura 3.12 se aprecian en la columna de la izquierda y de arriba hacia abajo: Las superficies contaminadas con ruido gausiano, impulsional y gausiano+impulsional. El resultado de la restauración empleando nuestro método se puede observar en la columna de la derecha de la misma figura. Podemos observar cómo las discontinuidades han sido preservadas y al mismo tiempo se ha realizado un suavizado tanto en las regiones continuas como en las proximidades de discontinuidades.

Al aplicar el filtrado 3σ-*MPF* sobre imágenes de distancias conteniendo escenas reales,

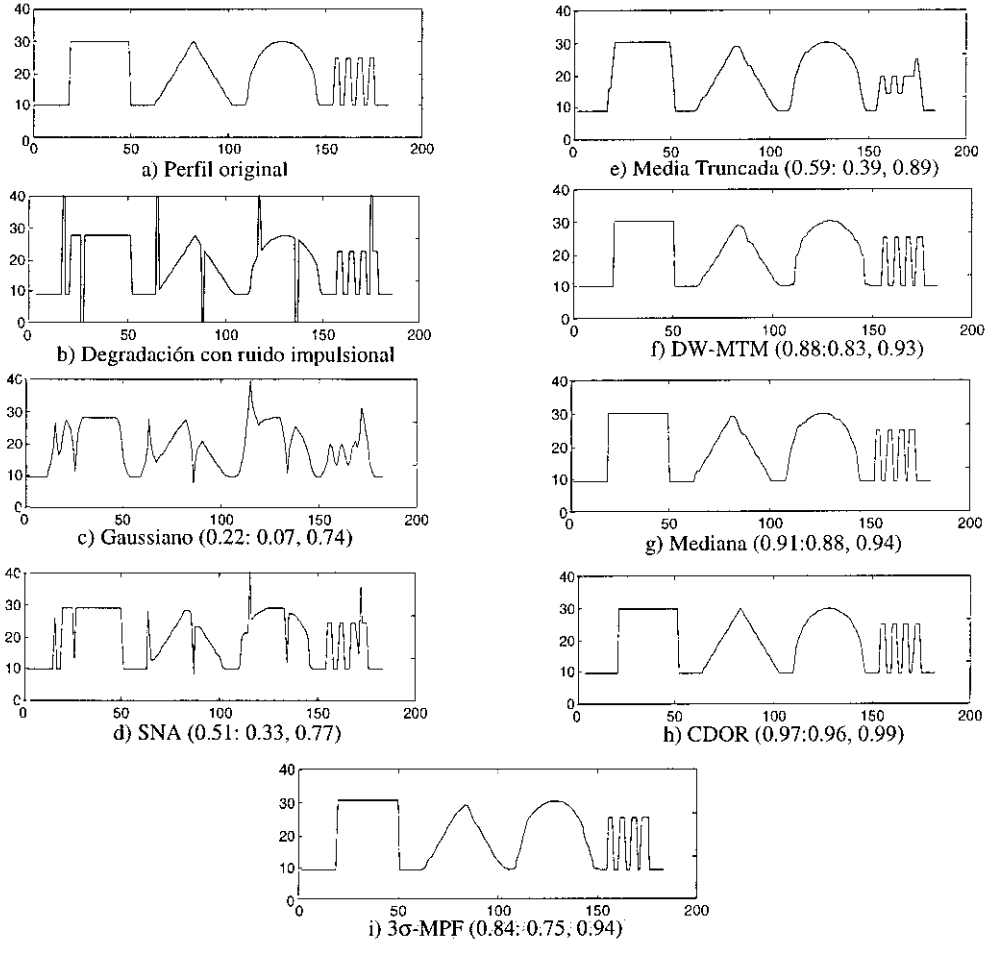


Figura 3.10: Perfiles contaminados con ruido *impulsional* y restaurados mediante 3σ-*MPF* y otras técnicas clásicas. Entre paréntesis se indica las métricas *GRI*, *FI* y *SI* respectivamente.

Método de filtrado	*GRI*	*FI*	*SI*
3σ-MPF (σ_n = 1, L=7,5,3)	0.69	0.58	0.82
DW-MTM (Mediana)	0.68	0.55	0.83
DW-MTM (CDOR)	0.68	0.60	0.77
MAS (Mediana)	0.67	0.55	0.82
MAS (CDOR)	0.67	0.59	0.77
DW-MTM (σ_n = 1, L=3, k=2, q=5)	0.65	0.56	0.75
Mediana (L=3)	0.57	0.51	0.64
Media truncada (L=5, M=1)	0.46	0.29	0.75
CDOR (L=5, T=5)	0.44	0.46	0.42
SNA (L=3)	0.37	0.25	0.56
Media Geométrica (L=3)	0.34	0.19	0.59
GIW (L=5)	0.32	0.21	0.47
MAS (σ_n = 1, L=7,5,3)	0.24	0.09	0.63
MMSE (σ_n = 1, L=3)	0.23	0.09	0.62
Sigma (σ_n = 1, L=5)	0.23	0.09	0.59
Media Y_p (L=3, P=2)	0.21	0.07	0.59
FIR (L=3)	0.20	0.06	0.69
Gausiano (L=7)	0.20	0.06	0.68
Wavelet (Daubechies, 3, Suave)	0.20	0.08	0.49
SCDOR (L=5,k=3)	0.19	0.08	0.45
Media aritmetica (L=3)	0.18	0.06	0.57
Contrarmónico (L=3, P=2)	0.08	0.01	0.57
Rango medio (L=5)	0.04	0.002	0.59
Media armónica (L=3)	0.03	0.001	0.57

Tabla 3.4: El filtro 3σ-*MPF* evaluado según la métrica *GRI*, y comparado con otras técnicas para eliminar ruido *gausiano* e *impulsional*.

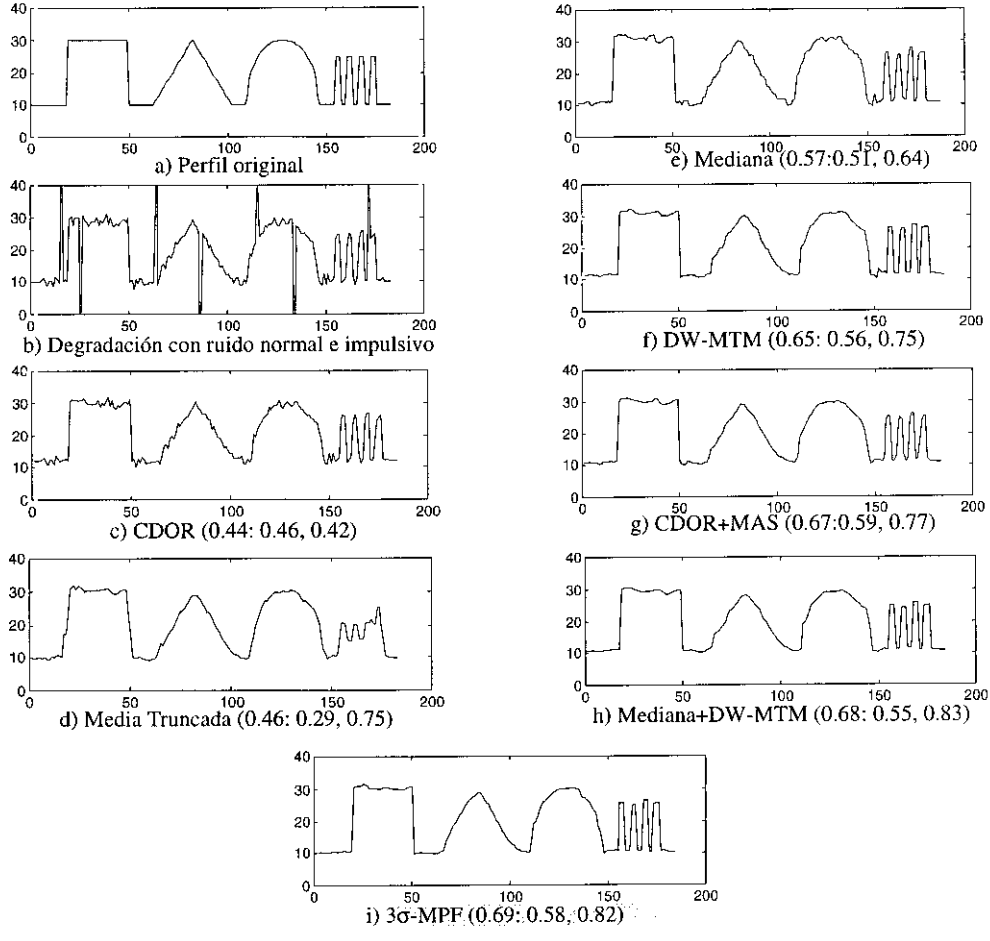


Figura 3.11: Perfiles contaminados con ruido *gausiano e impulsional* y restaurados mediante 3σ-*MPF* y otras técnicas clásicas. Entre paréntesis se indica las métricas *GRI*, *FI* y *SI* respectivamente.

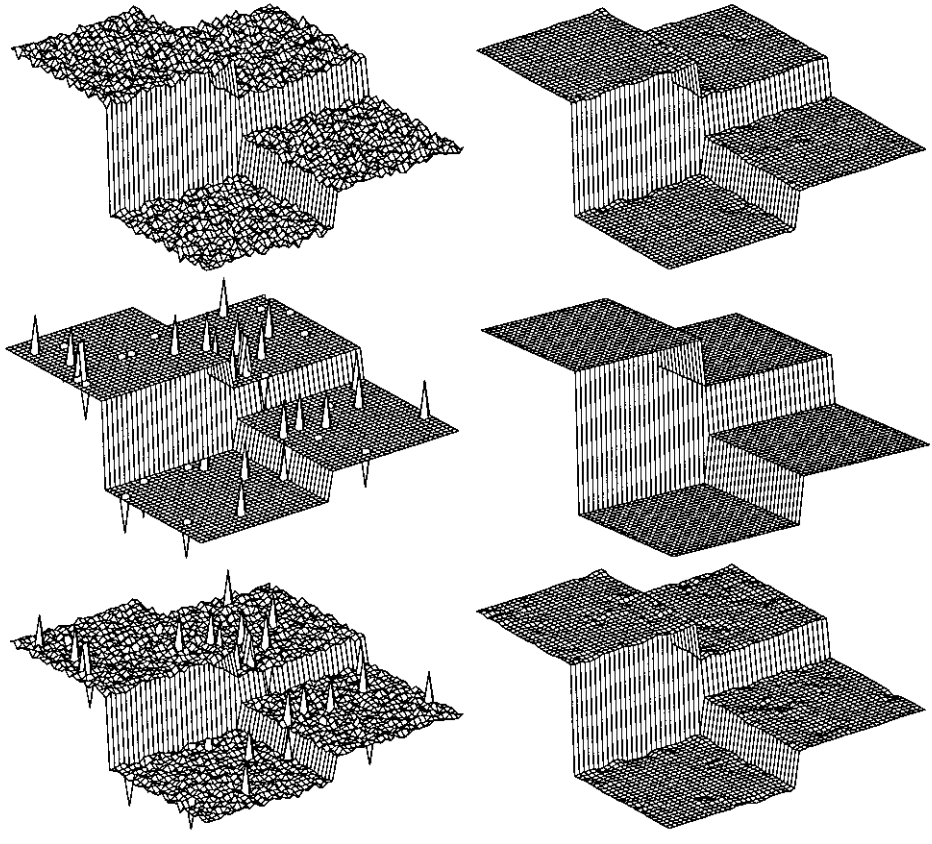


Figura 3.12: Superficies sintéticas contaminadas con ruido gausiano, impulsional y una combinación de ambos (izquierda). Restauración aplicando el filtro 3σ-*MPF* (derecha).

como la mostrada en la figura 3.13 que corresponde a un árbol artificial conteniendo dos naranjas, la mejora conseguida es más difícil de apreciar visualmente debido al mayor rango dinámico en la imagen. Para observar bien la mejora conseguida, en la tercera fila de la figura 3.13 podemos observar el perfil correspondiente a la línea horizontal (en blanco) sobre la imagen de distancia de la segunda fila que pasa por el centro de una de las naranjas. En el perfil se puede apreciar que en aquellos puntos donde el ruido presente está en consonancia con los valores de ruido estimados, se consigue un alto suavizado, y en cualquiera de los casos, la preservación de bordes se mantiene.

Considerando la frecuencia de activación de las diferentes técnicas de filtrado de que consta el filtro adaptativo 3σ-*MPF*, para el caso de la imagen natural de la figura 3.13, el ajuste de planos sobre ventanas 7x7 se activa en un 17% de los casos, un 16% para ventanas 5x5, un 28% para ventanas 3x3 y la técnica de media truncada modificada se aplica en el 39% de los casos. En la figura 3.14 se han representado mediante máscaras las regiones de activación de cada una de las cuatro posibles técnicas en las que puede evolucionar el filtrado 3σ-*MPF*. Como era de esperar los ajustes sobre ventanas grandes 7x7 se aplican en las regiones más homogéneas y la media truncada modificada se aplica en las discontinuidades o en sus proximidades.

3.4 Conclusiones

Los aspectos más destacables tratados a lo largo del presente capítulo han sido estos:

- Hemos presentado una *revisión de las técnicas de preprocesamiento* de imágenes incluyendo tanto métodos de realce como, fundamentalmente, métodos de restauración.
- Se ha propuesto una *nueva métrica* para la medida de la *calidad global de restauración* que hemos denominado *GRI*. Dicha métrica penaliza aquellos métodos de restauración que no preservan las discontinuidades con lo cual se pierde fidelidad entre la imagen restaurada y una referencia ideal, y considera positivamente la suavidad conseguida sobre la superficie de las regiones continuas.
- Se ha propuesto una *nueva técnica de restauración* de imágenes denominada 3σ-*MPF* (3σ-Multiresolution plane fitting), que permite realizar restauraciones con alto grado de suavizado y a la vez preservar las discontinuidades significativas de la imagen. La técnica utiliza una aproximación multiresolución realizando ajustes secuenciales de planos sobre ventanas de la imagen, comenzando por ventanas de mayor tamaño y progresivamente reduciendo su tamaño hasta que se encuentra un ajuste del plano con una fidelidad de tipo 3σ. Si ningún ajuste de este tipo se detecta significa que nos encontramos ante un borde y aplicamos un filtro de preservación de bordes *MTM*.
- Se ha *evaluado el método de restauración* 3σ-*MPF* utilizando la métrica *GRI* y se han comparado los resultados obtenidos frente a aquellos obtenidos a través de otras técnicas de restauración. Hemos encontrado que el método de restauración propuesto 3σ-*MPF* se comporta apreciablemente mejor que las demás técnicas en presencia de ruido gausiano y gausiano-impulsional.

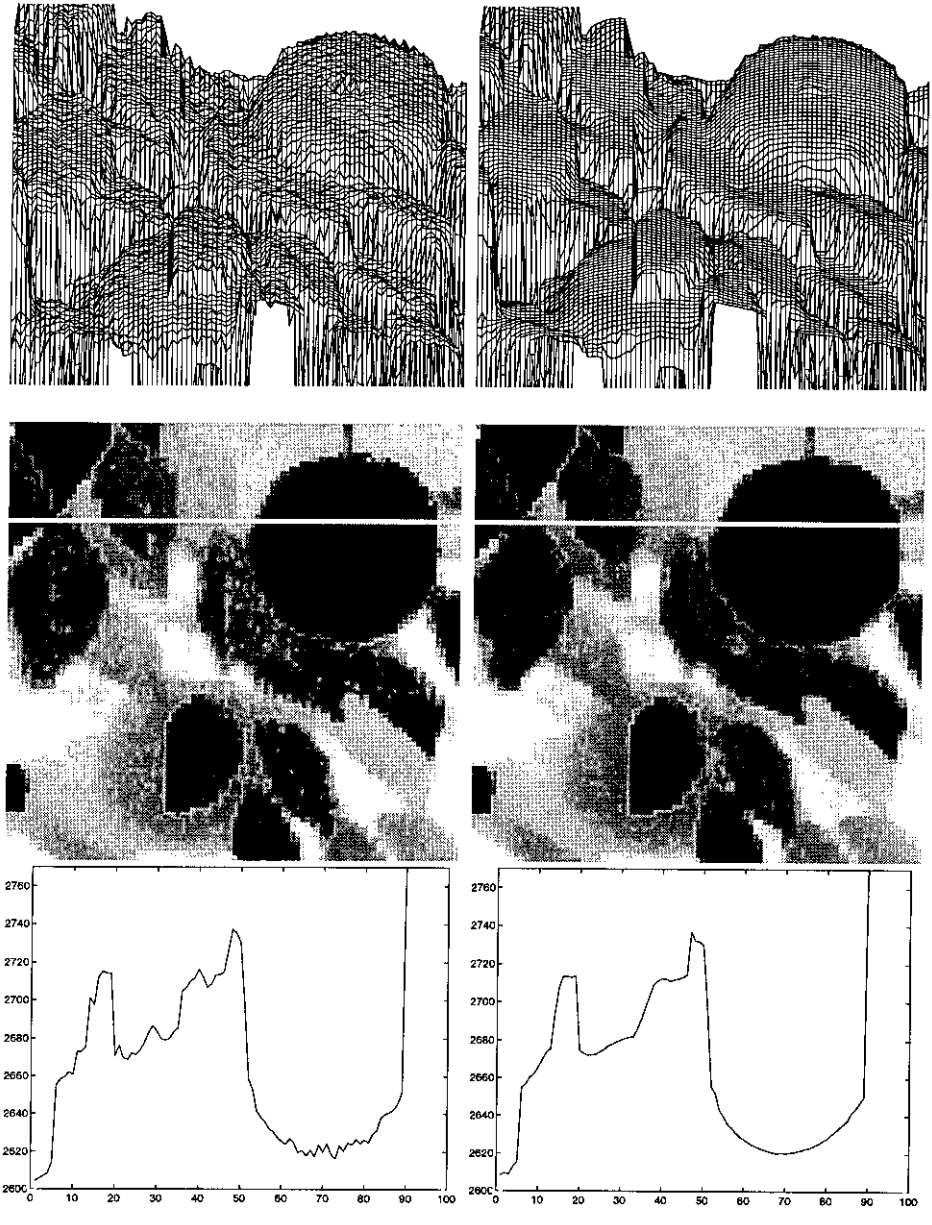


Figura 3.13: Restauración aplicando el filtro 3σ-*MPF* de una imagen de distancias captada con nuestro sensor láser (columna izquierda imagen degradada, columna derecha imagen restaurada).

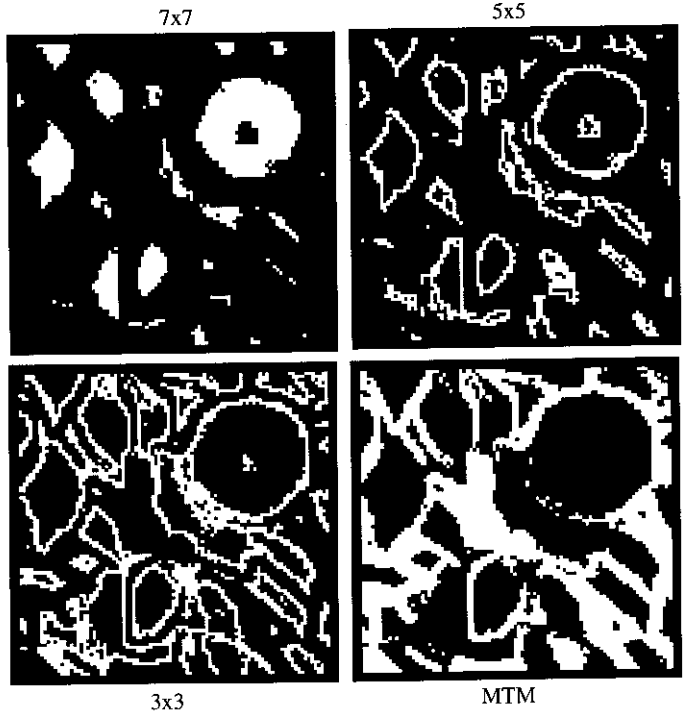


Figura 3.14: Máscaras representando las zonas de actividad del filtro 3σ-*MPF* sobre la imagen considerada en la figura 3.13. De izquierda a derecha y de arriba abajo: ajustes de planos en ventanas 7x7, 5x5, 3x3 y activación de la media truncada modificada.

Capítulo 4

Método de reconocimiento de objetos esféricos: Estrategia modular de acumulación de indicios mediante múltiples primitivas

Resumen. *En este capítulo se presenta la metodología de reconocimiento aplicada sobre imágenes de distancia y reflectancia para la detección, localización 3-D y parametrización de objetos esféricos. Inicialmente se muestran los fundamentos y técnicas de análisis de imágenes estableciéndose tres etapas diferenciadas: segmentación, descripción y clasificación. Seguidamente presentamos nuestro sistema de reconocimiento de objetos esféricos basado en acumulación de indicios que está especialmente indicado para casos con baja estructuración del entorno y oclusiones significativas de los objetos. El método de reconocimiento constituye un proceso modular que realiza la extracción de cuatro tipos de primitivas, permitiendo captar propiedades discriminantes del objeto buscado. A partir de dichas primitivas se realiza una extracción de parámetros de la esfera junto con el grado de confianza de dicha estimación. Finalmente se realiza una integración de hipótesis parciales generando los candidatos definitivos para ser identificados como esferas, que vienen parametrizados por su posición 3-D, radio y reflectividad media. En todo momento, el proceso de reconocimiento se distingue por una integración entre las imágenes de distancia y reflectancia basada en el modelo del sensor láser presentado en el capítulo 2, lo cual redunda en una mejor capacidad de interpretación.*

En el capítulo anterior vimos que mediante una técnica de restauración éramos capaces de obtener imágenes de mejor calidad que las directamente proporcionadas por el sensor láser. Por tanto ahora estamos en condiciones de poder aplicar estrategias de reconocimiento sobre la imagen de una forma más fiable que sin dicho preprocesamiento. Este hecho se podrá comprobar en el próximo capítulo, donde se realiza una evaluación de los métodos de reconocimiento presentados a lo largo del presente capítulo.

La presente tesis está orientada hacia la resolución de un tipo de problemas que contemplan la detección, localización-3D y caracterización de frutos en un árbol; por ello la estrategia de reconocimiento que se presentará se centrará en la satisfacción de este objetivo. Como se manifestó en el capítulo 2, la propiedad fundamental a utilizar para el reconocimiento será la *forma* de los objetos, siendo dicha información deducible tanto a partir del mapa de distancias como de reflectancias. Otro aspecto discriminante utilizado estará ligado a las *propiedades ópticas superficiales* del objeto, en concreto la reflectividad. Finalmente, la *posición tridimensional* de los objetos se utilizará para restringir las detecciones válidas a aquellas que están dentro del volumen de trabajo y para comprobar que se cumplen una serie de reglas de tipo heurístico.

Refiriéndonos a la forma, frutos tales como naranjas, manzanas o melocotones pueden ser aproximadamente modelados como esferas. De esta manera el problema de reconocimiento de frutos por formas lo vamos a plantear como un problema más general de reconocimiento de objetos esféricos en condiciones de oclusión y en entornos de baja estructuración. Así, el sistema de reconocimiento no se limitará solamente a aplicaciones de recolección de frutos en el campo, sino que su rango de aplicación se amplía a cualquier problema que considere el reconocimiento, localización y caracterización de objetos esféricos, donde el resto de objetos no son de interés y por tanto pueden ser ignorados.

Antes de presentar la estrategia de reconocimiento propuesta, analizaremos brevemente las principales técnicas de análisis de imágenes existentes, las cuales nos servirán de base para plantear nuestro método de reconocimiento.

4.1 Técnicas principales de análisis de imágenes

Independientemente del tipo de sensor utilizado, una vez obtenida una imagen o matriz de datos ligada a una determinada propiedad física (p.ej. intensidad, distancia, absorción, etc.), se nos plantea el problema de analizarla e interpretarla. Una imagen suele estar compuesta por una serie de objetos de interés y de un fondo formado por el resto de elementos superfluos. El objetivo de un sistema de análisis de imágenes, en general, es la identificación o reconocimiento de los objetos y la determinación de sus posiciones. Para poder realizar este reconocimiento es necesario aislar el área de la imagen que corresponde a cada objeto. Estas regiones se conocen en el área del procesamiento de imágenes como segmentos, y al proceso de subdividir una imagen en regiones correspondiendo a los objetos presentes, se conoce con el nombre de *segmentación*. Una vez segmentada la imagen, se debe realizar una *descripción* de cada segmento, y finalmente, basándose en la descripción previa, se realiza la *clasificación* de cada objeto. Estas tres etapas: *segmentación*, *descripción* y *clasificación*, son los pasos clásicamente utilizados para realizar el análisis de imágenes. Los tres apartados siguientes presentan una revisión

de las técnicas más destacables en cada una de estas etapas.

4.1.1 Segmentación

La segmentación es una etapa crucial en los sistemas de visión artificial por las dificultades que conlleva y por la importancia de sus resultados. Se han propuesto muchas técnicas de segmentación, sin embargo, todavía no se ha encontrado una solución general a este problema [83, 51]. En la mayoría de los casos, los algoritmos de segmentación operan basándose solamente en la información presente en la imagen. Esta aproximación no se parece a la estrategia utilizada en los seres vivos, donde la información captada directamente por los elementos de visión, es una parte pequeña si la comparamos con el conocimiento adicional utilizado para poder ver objetos e interpretar escenas. Por tanto, no es de extrañar que los resultados obtenidos por las aproximaciones no basadas en el conocimiento, no sean plenamente satisfactorias en algunos casos. Sin embargo, debido al conocimiento limitado de los procesos de almacenamiento y recuperación de la información en los seres vivos, y a las limitaciones de tiempo de proceso de los sistemas de visión experimentales que han intentado aproximarse a su emulación, actualmente no es viable dotar a un sistema de visión de una estrategia para la incorporación de conocimiento semejante a como lo hacen los seres vivos. Esta es un área actual de investigación y los sistemas existentes que incorporan conocimiento están basados en estrategias muy simplificadas.

Existen dos principios básicos para realizar la segmentación, uno está basado en la *detección de discontinuidades* y el otro en la *detección de similitudes*. En el primer caso, se plantea la hipótesis de que dos objetos diferentes deben tener propiedades distintas y por tanto debe existir una frontera entre ellos caracterizada por un cambio pronunciado en algún aspecto [104, 220, 212, 143]. Estas zonas de cambio se conocen como bordes y se pueden detectar analizando un vecindario próximo de cada punto de la imagen. Una vez que los bordes son detectados deben ser marcados y agregados, de forma que se obtengan contornos cerrados que definan los segmentos de la imagen. La estrategia de segmentación mediante similitudes, considera que los puntos que pertenecen al mismo segmento deben tener propiedades semejantes, y por lo tanto, realiza la agrupación de puntos siguiendo un cierto criterio de similitud [105, 82]. Este criterio suele considerar la proximidad espacial y la semejanza de intensidades para realizar la agrupación. Teóricamente, tanto siguiendo el principio basado en discontinuidades como el basado en similitudes, los resultados obtenidos deben ser los mismos. Si se obtienen los bordes y estos son cerrados, mediante un algoritmo de llenado se pueden obtener las regiones, y si obtenemos las regiones podemos calcular los bordes mediante un algoritmo de seguimiento de fronteras. Desafortunadamente, en la práctica esto es muy raro que se obtenga debido al cumplimiento solo parcial de las hipótesis de segmentación presentadas anteriormente.

4.1.1.1 Técnicas para la detección de bordes

Las técnicas de detección de bordes utilizan un pequeño vecindario en torno a un punto para detectarlos. Los bordes obtenidos, son simplemente puntos en la imagen que representan una discontinuidad, pero no hay ninguna relación entre ellos. La agrupación de estos puntos en contornos que delimitan segmentos, se verá en el siguiente apartado.

Las técnicas más habituales para detectar bordes, se basan en la aplicación de filtros paso altos. Los filtros aplicados suelen ser filtros espaciales, que son aproximaciones de filtros frecuenciales, pero con la ventaja de una mayor sencillez de cálculo. Existen fundamentalmente dos tipos de filtros:

- *Filtros paso alto basados en el gradiente*
- *Filtros paso alto basados en la Laplaciana*

Detección de bordes basada en el gradiente. Los bordes de la imagen se obtienen calculando el gradiente en cada punto e identificando aquellos puntos (pixels) con magnitudes de gradiente superiores a un cierto umbral. El gradiente en un punto (x, y) de la imagen $f(x, y)$ es un vector que se define según la ecuación 4.1.

$$\vec{G}[f(x,y)] = \nabla f = (G_x, G_y) = (\partial f/\partial x, \partial f/\partial y) \tag{4.1}$$

Dado el alto número de elementos que intervienen en una imagen es muy importante el tiempo de cálculo; por ello se han utilizado muchas aproximaciones para calcular las derivadas parciales en las direcciones x e y. Además es habitual aproximar la magnitud del gradiente $G = \sqrt{G_x^2 + G_y^2}$, con fines de eficiencia computacional mediante $G = |G_x| + |G_y|$ o $G = Max(|G_x|, |G_y|)$. Según el tipo de aproximación empleada para las derivadas parciales, podemos encontrar los siguientes filtros [51]:

- *Roberts.* Uno de los primeros operadores utilizados fue introducido por Roberts (1965). El operador cruzado de Roberts utiliza dos ventanas 2×2 para aproximar las dos componentes del gradiente (fig. 4.1). Mediante estas ventanas, se calcula la diferencia de los puntos de la diagonal de las ventanas como muestran las ecuaciones 4.2 y 4.3.

$$G_x = \partial f(x,y)/\partial x = f(x, y+1) - f(x-1, y) \tag{4.2}$$

$$G_y = \partial f(x,y)/\partial y = f(x,y) - f(x-1, y+1) \tag{4.3}$$

- *Prewitt, Sobel y Frei-Chen.* El operador de Prewitt (1970) realiza la aproximación utilizando dos ventanas de tamaño 3×3 (fig. 4.1), orientadas para detectar bordes tanto en direcciones verticales como horizontales. El cómputo de las aproximaciones al gradiente, se muestra en las ecuaciones 4.4 y 4.5, tomando $A = 1$. El operador de Sobel (1970) da peso doble a los puntos geométricamente más próximos, y por tanto $A = 2$. La aproximación de Frei-Chei (1977) utiliza $A = \sqrt{2}$.

$$\begin{aligned} G_x = \partial f(x,y)/\partial x \;=\; & (f(x-1,y+1) + Af(x,y+1) + f(x+1,y+1)) - \\ & (f(x-1,y-1) + f(x,y-1) + f(x+1,y-1)) \end{aligned} \tag{4.4}$$

$$\begin{aligned} G_y = \partial f(x,y)/\partial y \;=\; & (f(x-1,y-1) + Af(x-1,y) + f(x-1,y+1)) - \\ & (f(x+1,y-1) + f(x+1,y) + f(x+1,y+1)) \end{aligned} \tag{4.5}$$

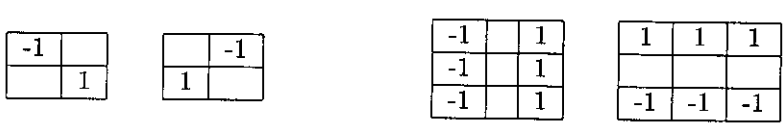


Figura 4.1: Ventanas utilizadas por el operador de Roberts (izquierda) y Prewitt (derecha).

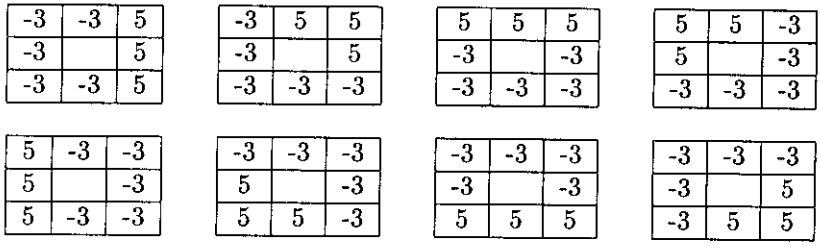


Figura 4.2: Ventanas utilizadas por el operador de Kirsch.

- *Kirsch, Robinson y Nevatia-Babu.* El método de Kirsch (1971), utiliza ocho máscaras de tamaño 3×3, cada una de las cuales da un indicación de la existencia de un borde en una determinada dirección (fig. 4.2). Una vez aplicadas todas las máscaras, la magnitud del gradiente, G, es el máximo del resultado obtenido con cada una de las máscaras. La dirección del gradiente θ es un múltiplo de 45 grados, cuyo valor depende de la máscara con la que se obtuvo el resultado máximo. Un método semejante es utilizado por Robinson (1977), pero sus máscaras utilizan valores menores; solo 0, ± 1 y ± 2. Nevatia-Babu (1980), utilizan 12 máscaras de tamaño 5×5 y los intervalos angulares son de 30 grados.

Los detectores de borde basados en el gradiente, presentan dos problemas principales: El ruido y el grosor de los bordes detectados. Los bordes ruidosos se pueden tratar parcialmente filtrando la imagen original antes de aplicar los operadores de detección de bordes. En este sentido, hay que tener cuidado con el tipo de filtrado que se utiliza, puesto que es bastante frecuente utilizar filtros que distorsionan los bordes o los hacen indetectables. Idealmente, un algoritmo de eliminación de ruido se debe aplicar solo en la dirección del borde, es decir, perpendicular a la dirección del gradiente, de esta forma no se difuminan. El segundo problema se debe a que los bordes detectados suelen tener más de un punto. Esto es debido a que los bordes no son siempre escalones ideales y frecuentemente son rampas o bordes suavizados. Se suele solucionar este problema aplicando técnicas de supresión de puntos no máximos y por tanto conservando sólo los máximos locales en los mapas de gradientes.

Probablemente el mejor detector de bordes escalón es el filtro de *Canny*, debido a su sencillez y a los buenos resultados que se obtienen con él [26]. El operador de Canny,

	1	
1	-4	1
	1	

Figura 4.3: Máscara utilizada para calcular la Laplaciana.

tras suavizar la imagen original con un filtro gausiano, elige como puntos borde aquellos con gradiente máximo. Posteriormente se realiza un posprocesamiento aplicando un umbral en un proceso de histéresis que elimina bordes falsos. Los resultados obtenidos son casi óptimos y por ello es el detector de bordes escalón más aceptado actualmente.

Detección de bordes basado en la Laplaciana. La Laplaciana está definida en términos de segundas derivadas parciales de la forma

$$L[f(x,y)] = \nabla^2 f = \partial^2 f/\partial x^2 + \partial^2 f/\partial y^2 \tag{4.6}$$

Los filtros basados en la Laplaciana, se pueden aproximar como se indica en las ecuaciones 4.7 y 4.8, y por tanto se puede utilizar una máscara 3×3 como la mostrada en la figura 4.3.

$$\partial^2 f/\partial x^2 = f(x+1,y) - 2f(x,y) + f(x-1,y) \tag{4.7}$$

$$\partial^2 f/\partial y^2 = f(x,y+1) - 2f(x,y) + f(x,y-1) \tag{4.8}$$

El hecho de que la Laplaciana se base en las segundas derivadas parciales de la imagen la hace extremadamente sensible al ruido. Es por ello, que no suele ser tan utilizada como los detectores de borde basados en el gradiente. Sin embargo, la detección de pasos por cero una vez aplicada la Laplaciana, permite obtener con mayor precisión la posición del borde, lo que es útil cuando existen bordes anchos y graduales. Mediante la técnica de detección de pasos por ceros en la segunda derivada, lo que hacemos es detectar los máximos locales en un mapa de gradientes, en contraste con la detección por umbral utilizada en los métodos basados en el gradiente.

La alta sensibilidad al ruido de la Laplaciana crea problemas y por ello se suele utilizar la Laplaciana junto con una etapa previa de reducción de ruido. En este sentido Marr y Hildreth proponen un filtro Gausiano para promediar la imagen, dicho filtro se caracteriza por conservar espacialmente las propiedades en la imagen. El método consiste en convolucionar la imagen con una máscara gausiana y posteriormente aplicar la Laplaciana. Debido a las propiedades conmutativa y asociativa de la convolución, aplicar el método anterior es lo mismo que convolucionar directamente la imagen con la función *LOG* o Laplaciana del Gausiano (ec. 4.9).

$$\begin{aligned} \nabla^2(G(x,y) * f(x,y)) &= (\nabla^2 G(x,y)) * f(x,y) = \\ LOG * f(x,y) &= \left((\frac{x^2+y^2-2\sigma^2}{\sigma^4 \quad -}) e^{-\frac{x^2+y^2}{2\sigma^2}} \right) * f(x,y) \end{aligned} \tag{4.9}$$

La variación de σ permite realizar filtrados variables, pudiéndose adaptar el algoritmo a las diferentes dimensiones de los bordes. La solución adoptada por Witkin utiliza un entorno *espacio-escala*, donde el eje de escala se corresponde con σ [215]. Las diferentes escalas de resolución se consiguen al variar σ; cuando su valor crece se detectan solamente los bordes más significativos y al disminuir se detectan incluso los más pequeños.

Existen otras técnicas, también basadas en la segunda derivada, que utilizan el ajuste de funciones sobre pequeñas áreas de la imagen. La idea es obtener una función continua que represente la imagen discreta original, de esta forma se pueden obtener localizaciones con precisión subpunto. En el *modelo facetado* de Haralick [74], el vecindario de un punto es aproximado por una función cúbica , cuyos coeficientes son obtenidos mediante ajuste por mínimos cuadrados. Los puntos que corresponden a un borde son aquellos que cumplen que la primera derivada es distinta de cero y la segunda derivada es igual a cero.

4.1.1.2 Segmentación por agregación y detección de contornos

Los resultados obtenidos utilizando los métodos vistos en la sección anterior, generan normalmente bordes fraccionados que no delimitan regiones. Debido a esto, los algoritmos de detección de bordes son seguidos de métodos de unión de puntos (agregación) o por algoritmos de detección de contornos. De esta forma es posible cerrar los contornos, o bien, obtener una caracterización del contorno fragmentado. A continuación vamos a ver algunos de los métodos utilizados para realizar estas tareas.

Técnicas de agregación mediante análisis local. Una de las formas más directas de realizar la unión de bordes, es analizar las características de los puntos en torno a un pequeño vecindario, utilizando *técnicas de relajación*. Un punto de la imagen, inicialmente no considerado como borde, puede llegar a serlo si cumple una serie de restricciones. Dos posibles restricciones son que la diferencia de gradientes y la diferencia de direcciones del gradiente, entre dos puntos, deben ser inferiores a un cierto umbral. Por tanto, un punto (a, b) de un vecindario de (x, y) se une al grupo de puntos al que pertenece (x, y), si se cumplen los criterios de agrupamiento. Una forma sencilla de marcar los grupos de puntos que definen los diferentes contornos, es asignar un nivel de gris diferente a cada grupo. El método es aplicado moviendo una ventana (3×3 o 5×5) a través de toda la imagen, aunque existen otros algoritmos conocidos como seguidores de bordes, que comienzan a trabajar desde un punto con alta magnitud en el gradiente. En este caso, a partir del punto inicial se va siguiendo el contorno y se va aplicando el criterio de agrupamiento hasta cerrar los bordes [190].

Transformada de Hough. La transformada de Hough es un método para el reconocimiento de líneas rectas y curvas [47, 120, 90, 152, 159]. La estrategia consiste en expresar la línea a reconocer mediante una expresión del tipo: $g(\vec{x}, \vec{c}) = 0$. Donde $\vec{x} = (x, y)$ representa las coordenadas de la imagen, y $\vec{c}$ son las coordenadas en el espacio de la transformación o espacio de parámetros. En el caso de una recta la transformación se define mediante

$$\rho = x \cos\theta + y \sin\theta \tag{4.10}$$

donde ρ y θ, son las variables en el espacio de parámetros, y representan la distancia y el ángulo de un segmento perpendicular a la recta considerada y que pasa por el origen del

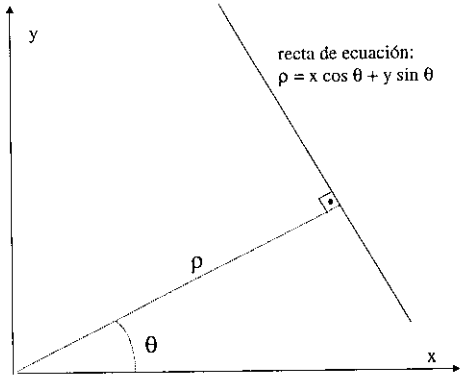


Figura 4.4: Parámetros ρ y θ que definen una recta utilizados en la transformada lineal de Hough.

sistema de referencia (fig. 4.4). Mediante esta transformación por cada punto (x, y) de la imagen, se genera una sinusoide en el espacio $\rho - \theta$. Todos los puntos correspondientes a una recta, generan sinusoides que se intersectan en un solo punto (ρ, θ), y que caracterizan la recta presente en la imagen. La manera de generar estas sinusoides y los puntos de corte, se plantea como un proceso de votación sobre el espacio de parámetros. Este espacio se discretiza en forma de matriz con el objetivo de crear casillas donde poder depositar un voto por cada posible pareja de parámetros. Cada punto que representa un borde en la imagen, realiza la votación incrementando el valor de varias casillas de la matriz de acumuladores. Finalmente, las casillas con mayores votaciones se corresponden a los puntos de corte de sinusoides, y por tanto nos permiten detectar las líneas rectas presentes en la imagen.

La transformada de Hough se puede utilizar también para detectar líneas curvas [3, 158] e incluso cualquier contorno con forma arbitraria [6]. En el caso de búsqueda de circunferencias, los parámetros necesarios son tres: el centro de la circunferencia (c_x, c_y) y el radio R de esta. En este caso la transformación es

$$(x - c_x)^2 + (y - c_y)^2 = R^2 \tag{4.11}$$

El principal problema de la transfomada de Hough, es la larga búsqueda requerida en el espacio de parámetros. Existen formas de aumentar la eficiencia de este método, por ejemplo, utilizando la dirección del gradiente de la curva; de esta forma se limita el número de votaciones que emite cada punto [120]. Otras optimizaciones utilizan la selección de tripletas de forma aleatoria mediante la técnica conocida como RHT (Randomized Hough Transform) [116] o incluso aplican heurísticas basadas en las relaciones espaciales de los puntos de una circunferencia [124, 131, 86].

Ajuste de curvas a los bordes. Dada una imagen conteniendo los bordes fragmentados correspondientes al contorno de un solo objeto, podemos ajustar una curva a esos datos de tal forma que obtengamos los parámetros que definen esa curva y por tanto permitiendo definir la frontera de un objeto. Las técnicas de ajuste aplicables pueden ser diversas

[221]: mínimos cuadrados, mínimos cuadrados de medianas, minimización por otros criterios, ajustes sobre subconjuntos de puntos elegidos aleatoriamente ó división del conjunto de puntos en tantos grupos como incógnitas haya en la curva a calcular [219].

Algunas de estas técnicas son más robustas ante presencia de ruido de tipo impulsional y otras, como el ajuste por mínimos cuadrados, no lo son tanto. En cada circunstancia se aplicará la técnica más adecuada. En cualquier caso el mayor problema se encuentra a la hora de aislar los puntos del borde que pertenecen a un único objeto. Como es de esperar, si estamos tratando de definir la región sobre la cual existe un objeto ¿cómo vamos a separar dentro de una imagen de bordes aquellos puntos que deben ser ajustados conjuntamente por pertenecer al mismo objeto, si no sabemos dónde está dicho objeto?. La respuesta no es evidente y supone un fuerte contratiempo a la hora de aplicar esta estrategia.

Detección de bordes continuos mediante grafos. Un grafo está formado por una serie de nodos unidos mediante arcos. La idea de esta estrategia consiste en representar cada posible componente elemental de un borde mediante un nodo. Los nodos se unen por arcos que llevan asociados un peso. Este peso es un coste, e indica la penalidad que supondría que aceptásemos la existencia de un borde entre los nodos que une el arco, considerando la calidad del contorno resultante. De esta forma, el contorno óptimo se obtiene encontrando un camino dentro del grafo con coste mínimo.

La función de coste utilizada debe generar valores pequeños cuando el camino entre dos puntos de la imagen se mantiene dentro de una zona de transición, indicando que estamos sobre un borde. Si no hay o no permanecemos en la zona de transición, el coste debe crecer debido a que esos dos puntos no representan un borde. Una posible medida de coste $c(a,b)$ entre los puntos a y b, se puede describir en términos del gradiente de esta forma

$$c(a,b) = \frac{(G_{max}(f(x,y)))^2}{G(f(a_x,a_y))\overline{G(f(b_x,b_y))}} \tag{4.12}$$

El problema de encontrar un camino de coste mínimo en un grafo no es trivial, y requiere gran cantidad de cálculo. Se suelen adoptar soluciones más rápidas a costa de obtener soluciones casi mínimas.

4.1.1.3 Segmentación por umbral

La técnica de segmentación por umbral, se basa en la selección de un valor límite, T, que separa las zonas de la imagen que pertenecen a objetos, de las zonas correspondientes al fondo. Este valor T se conoce con el nombre de umbral. La obtención del umbral se puede realizar a partir de características *globales* de la imagen, o a partir de propiedades *locales*. En el caso de que el umbral tenga un valor que depende de cada punto de la imagen, es decir $T = T(x,y)$, se dice que el umbral es *dinámico*.

Dada una imagen $f(x,y)$, la imagen umbralizada $g(x,y)$ se obtiene de la forma siguiente:

$$g(x,y) = \begin{cases} 1 & \text{si } f(x,y) > T \\ 0 & \text{si } f(x,y) \leq T \end{cases} \tag{4.13}$$

La selección del umbral T se puede hacer de diversas formas:

- *Experimentalmente.* Se realizan umbralizaciones a varias imágenes y se selecciona el valor que aparentemente separa mejor los objetos del fondo. Este método no es recomendable debido a que no se adapta a futuros cambios en las imágenes.

- *Promedio de picos en histograma.* Dado un histograma bimodal, es decir, aquel con dos máximos claramente diferenciados, para obtener el umbral T basta detectar los máximos y realizar el promedio.

- *Óptimo.* Consideremos que el histograma $h(z)$ de una imagen se puede aproximar por la suma ponderada de dos funciones de densidad de probabilidad:

$$h(z) = P_1 p_1(z) + P_2 p_2(z) \tag{4.14}$$

donde, P_1 y P_2 representan las probabilidades *a priori* y $p_1(z)$ y $p_2(z)$ son las funciones de densidad. El valor óptimo de umbralización T cumple la igualdad $P_1 p_1(T) = P_2 p_2(T)$. Por tanto, suponiendo funciones de densidad Gausianas para $p_1(z)$ y $p_2(z)$, con medias m_1 y m_2, y considerando dispersiones iguales en ambos casos , es decir $\sigma_1 = \sigma_2 = \sigma$, tenemos que el valor óptimo de T viene dado por

$$T = (m_1 + m_2)/2 + ln(P_2/P_1)(\sigma^2/(m_1 - m_2)) \tag{4.15}$$

Después de obtener la imagen binaria, todos los objetos aparecen en la imagen representados por un "1". Para separar los diferentes objetos, y finalizar la segmentación, se debe aplicar un algoritmo de etiquetado que busque regiones conexas.

En muchos casos, el histograma no presenta una configuración bimodal o existe mucho ruido entre ambos picos. Para "limpiar" el histograma y resaltar los picos se suele aplicar un umbral previo que elimine de la imagen los puntos correspondientes a transiciones entre regiones. Para ello se puede utilizar como métrica el gradiente de cada punto. De esta forma, se eliminan los puntos correspondientes a transiciones, profundizando el valle que separa los picos del histograma y por tanto resaltando estos.

En el caso de que la imagen conste de varias componentes, como en las imágenes a color que tienen 3 componentes, el histograma es tridimensional. Para realizar la segmentación se utilizan técnicas de agrupamiento (clustering) y el concepto de umbral se sustituye por la clasificación de puntos en un espacio de características. En el caso de parejas de imágenes distancia-intensidad, se ha utilizado un tipo especial de histograma que combina ambos tipo de información en un solo histograma conocido como RIH (Range Intensity Histogram). El histograma RIH se forma acumulando por cada distancia la suma de los valores de intensidad, por tanto los picos de este histograma se corresponden a regiones en un mismo plano y con alta intensidad [98].

4.1.1.4 Segmentación basada en características de regiones

En este apartado vamos a ver técnicas de segmentación que obtienen las regiones de una forma directa. Existen tres aproximaciones fundamentales: *agrupamiento en espacio de características*, *crecimiento de regiones* y *división y unión*.

Agrupamiento en espacio características. Mediante este procedimiento, por cada punto de la imagen se extrae una serie de características, como su posición

(i,j), su intensidad $f(i,j)$, el gradiente y su dirección, formando un patrón o vector de características. A continuación se realiza un agrupamiento sobre el espacio de características buscando los grupos naturales, aplicando alguna de las múltiples técnicas de agrupamiento que existen [54]. Cada grupo detectado se corresponderá a un segmento, que normalmente deberá ser mejorado aplicando técnicas de relleno, crecimiento o división de segmentos [97].

Crecimiento de regiones. Es un procedimiento mediante el cual se agrupan puntos o subregiones para formar nuevas regiones mayores. La versión más sencilla consiste en utilizar varios puntos *semilla*, a partir de los cuales se van agregando otros puntos para formar regiones. Para agregar un punto en una región, se debe cumplir un cierto *criterio de similitud*. Por ejemplo, un criterio posible consistiría en incluir un punto en una región si la diferencia de intensidad, entre el punto y la "semilla" correspondiente a esa región, no supera un cierto umbral. Este proceso se repite hasta que no haya más puntos que satisfagan el criterio de similitud.

Dos aspectos críticos en esta técnica son, la selección inicial de las semillas y del criterio para incluir puntos. Una forma de selección de las semillas puede consistir en calcular medidas de similitud sobre todos los puntos de una imagen y ver si aparecen agrupaciones naturales. Los centroides de estos grupos se pueden utilizar como las "semillas" iniciales. En cuanto al criterio utilizado, este debe caracterizar la región, para ello se pueden utilizar descriptores de regiones como los citados en la sección 4.1.2.2 que se verá posteriormente (momentos de área, texturas, intensidades medias, coeficientes de un ajuste [8] ó curvaturas [203]). Besl [14], ajusta *superficies de orden variable* a regiones para caracterizarlas, definiendo de esta forma un criterio para realizar el crecimiento de regiones. Las semillas se obtienen después de un proceso de etiquetado donde, usando la curvatura, se distinguen entre ocho diferentes tipos de superficies. Najman [153], investiga técnicas de segmentación basadas en la representación de la imagen como una superficie topográfica y utilizando un vertido de agua sobre ella, determina los mínimos más significativos allí donde haya mayor acumulación de agua. Otros trabajos utilizan "culebras" (snakes), que son curvas continuas que desde un estado inicial ajustan dinámicamente su posición y forma hasta que se establece un estado de equilibrio con los datos subyacentes en la imagen [171].

División y unión. En esta aproximación, se parte de una única región correspondiente a la totalidad de la imagen. Se utiliza un criterio de similitud P para saber si una región es homogénea, y por tanto, saber si no debe ser dividida. Aplicando el criterio de similitud a una región R, si no se cumple dicho criterio, es decir $P(R) = Falso$, entonces debemos dividir la región de una forma arbitraria. Posteriormente, se realiza la unión de las regiones que cumplen el criterio, es decir, si $P(R_i) \cup P(R_j) = Verdadero$. El proceso anterior se repite iterativamente, hasta que el algoritmo se detiene cuando no es posible realizar más uniones o divisiones.

El proceso de división de una región a varias es problemático, puesto que hay que definir como realizarla. Evidentemente, existen infinitas formas de dividir una región en varias. Una posibilidad consiste en dividir la región de una manera regular, por ejemplo, partiendo la región en cuatro secciones cuadradas. En la figura 4.5 se puede ver un ejemplo del proceso de división y unión, utilizando divisiones uniformes mediante cuadrados.

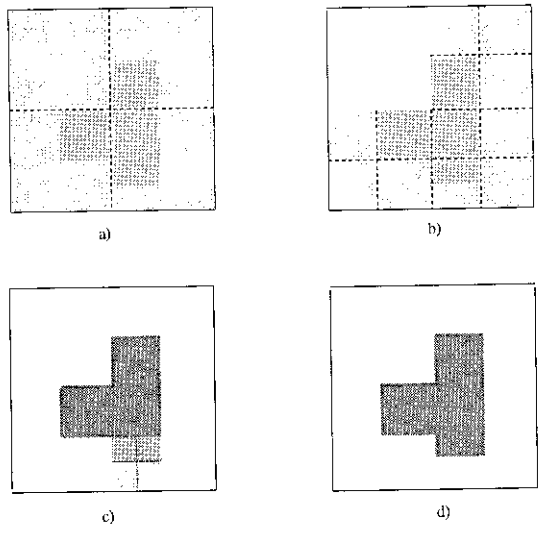


Figura 4.5: Segmentación mediante *división y unión*: a) partición inicial, b) y c) las regiones que cumplen la propiedad de homogeneidad P se marcan con blanco o gris oscuro. Si no se cumple se realizan mas subdivisiones, d) imagen final segmentada.

.

4.1.1.5 Segmentación por movimiento

El movimiento de un objeto respecto a un fondo inmóvil, permite realizar la segmentación del objeto que se mueve [96]. Una técnica inmediata consiste en tomar dos imágenes a diferentes tiempos y restar punto a punto las imágenes. La *imagen diferencia* obtenida contiene el contorno del objeto móvil en la dirección del movimiento. La imagen diferencia $d_{i,j}(x,y)$ entre los instantes i y j se calcula así:

$$d_{i,j}(x,y) = \begin{cases} 1 & \text{si } f(x,y,t_i) - f(x,y,t_j) > Umbral \\ 0 & \text{en otro caso} \end{cases} \tag{4.16}$$

Otro método consiste en generar una *imagen de diferencias acumulativas*. Basta capturar una secuencia de imágenes y restarlas respecto a una imagen referencia (p.ej. la primera de la secuencia), acumulando las diferencias resultantes. En el momento que el objeto móvil sale del área ocupada inicialmente, se obtiene una segmentación perfecta del objeto móvil.

4.1.2 Descripción o extracción de características

Una descripción consiste en proporcionar una serie de características referentes a un objeto que permitan su reconocimiento. Estas características, deben ser suficientemente representativas y discriminantes como para permitir diferenciar entre objetos distintos o detectar los objetos que pertenezcan a la misma clase. Idealmente, los descriptores

deberían ser independientes del tamaño, posición y orientación del objeto en la imagen, puesto que un cambio en estos aspectos nunca debe suponer la modificación de la clase de objeto. La etapa de descripción es muy importante, puesto que el posterior proceso de reconocimiento o clasificación se va a basar en estos descriptores. Si la descripción no es adecuada o suficientemente discriminatoria, la complejidad de los algoritmos de reconocimiento será mayor y difícilmente se va a poder realizar una correcta clasificación. En función de qué entidad se describa, se encuentran tres tipos de descriptores: descriptores de *contorno*, descriptores de *regiones* y otros dedicados al caso especial de imágenes *tridimensionales*.

4.1.2.1 Descriptores de contorno

Descripción de contornos mediante códigos cadena. Los códigos cadena se utilizan para representar un contorno mediante pequeños segmentos rectos en una determinada dirección y de una longitud predefinida. Se suelen emplear dos tipos de código cadena: de 4 direcciones y de 8 direcciones. Si trabajamos con códigos de 4 direcciones, se marcan los segmentos apuntando a la derecha con un 0, hacia arriba con un 1, izquierda con un 2 y abajo con un 3. Para generar un código cadena (fig. 4.6), se recorre el contorno hasta que nos desplazamos una determinada longitud. Se conecta, mediante un segmento recto, el punto final alcanzado con el de partida y se extrae su código cadena al elegir aquel con la dirección más próxima a la del segmento obtenido. Este proceso se repite hasta que se alcanza de nuevo el punto de partida, con lo cual se completa la descripción del contorno cerrado.

El código cadena que se obtiene no depende de la posición del objeto en la imagen, pero depende del punto de comienzo de la descripción, de la orientación del objeto y de su tamaño. Para normalizar la descripción en cuanto al punto de inicio, se suele redefinir este punto de tal forma que el número entero que representa el código cadena tenga una magnitud mínima. El código cadena también es sensible a rotaciones, por ello, en vez de utilizar el código en sí, se puede utilizar la diferencia entre los códigos. Esta diferencia se calcula contando (p.ej. en dirección horaria) el número de direcciones que separan dos códigos adyacentes. La normalización del tamaño puede ser obtenida subdividiendo el contorno de todos los objetos en un número constante de segmentos y por tanto obteniendo siempre códigos cadena de la misma longitud.

Integrando estos criterios de normalización, el *número forma* se define como la primera *diferencia*, obtenida a partir de una descripción con *códigos cadena*, con menor magnitud. Su orden n indica el número de dígitos empleados para su representación.

Descripción de contornos mediante firmas. Una firma es una representación de un contorno mediante una función unidimensional. Una de las firmas más sencillas se obtiene al calcular el centro de un contorno, y tomando este punto como referencia, representar la distancia r entre el centroide y un punto del contorno en función del ángulo de barrido θ. La firma obtenida es la función unidimensional $r(\theta)$ que depende del punto de partida y del tamaño del contorno. La invarianza respecto al tamaño se consigue, por ejemplo, normalizando la función para que tenga un valor máximo determinado. La invarianza ante el punto de partida, se consigue aplicando una estrategia similar a la utilizada para el código cadena.

La representación del ángulo, ϕ, existente entre la tangente a la curva en un punto y

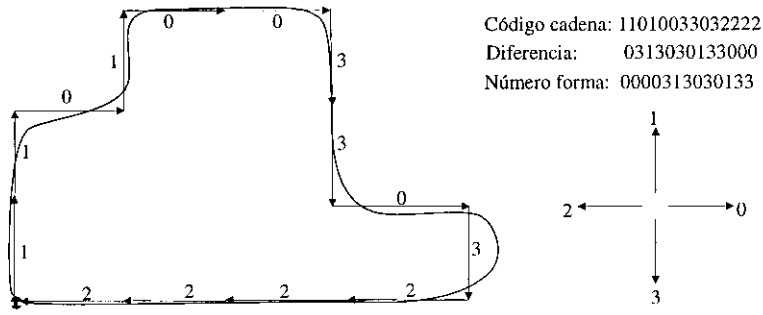


Figura 4.6: Generación del código cadena mediante la subdivisión del contorno en segmentos.

una línea fija de referencia (p.ej. la horizontal), en función del desplazamiento s a lo largo del contorno, es otra forma de firma que se representa mediante la función $\phi(s)$. La representación gráfica de la función muestra las líneas rectas del contorno mediante valores constantes o segmentos horizontales. Por tanto, el histograma de $\phi(s)$ presenta máximos locales cuando existen líneas rectas en el contorno que pretende describir.

Una vez obtenida la firma tenemos una función unidimensional, pero es necesario obtener una descripción más adecuada para la fase de reconocimiento. Debido a que el problema ha pasado de un espacio bidimensional a otro unidimensional, la descripción es más sencilla y en la literatura existen multitud de descriptores utilizados para estos casos. Por ejemplo, se pueden calcular los momentos de la firma, que para sus ordenes más bajos son el promedio y la varianza de la señal.

Descripción de contornos por aproximaciones poligonales. Un contorno puede ser aproximado mediante un polígono. El objetivo de esta aproximación es capturar la esencia de la forma del contorno, pero utilizando el mínimo número de segmentos posibles. Existe un método, que data de 1972, que encuentra polígonos con perímetro mínimo (fig. 4.7a). El proceso comienza colocando celdas cuadradas sobre el contorno, de tal forma que dicho contorno queda encerrado entre las paredes interiores y exteriores de las celdas concatenadas. Si consideramos el contorno como una goma elástica y esta se contrae, obtenemos una aproximación al contorno inicial mediante segmentos rectos.

Otras soluciones se basan en técnicas de agrupamiento de puntos según algún criterio. Los puntos a lo largo de un contorno se van agrupando, hasta que el error acumulado en un ajuste por mínimos cuadrados a una recta supera un cierto umbral. En este último caso, se finaliza la aproximación de un segmento y se comienza otra agrupación en búsqueda del segmento siguiente.

Otra técnica consiste en dividir sucesivamente un segmento recto en dos partes, hasta que se alcanza un determinado criterio (fig. 4.7b). Si la máxima distancia (siguiendo la perpendicular a un segmento recto que une dos puntos del contorno) a un punto del contorno, supera un determinado umbral, este punto lejano del contorno se convierte en un vértice más de la aproximación poligonal. Para un contorno cerrado la mejor pareja de puntos para comenzar, es la formada por los dos puntos más alejados que pertenecen al contorno.

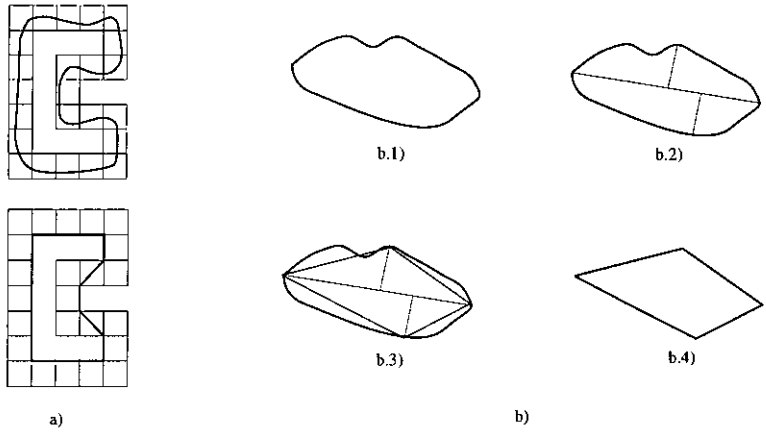


Figura 4.7: a) Método de descripción mediante polígonos de perímetro mínimo b) Aproximación mediante divisiones sucesivas de segmentos: b.1) contorno original, b.2) puntos de máxima separación respecto al segmento central, b.3) división del segmento central, b.4) polígono final.

Descriptores de Fourier. Si los puntos correspondientes a un contorno, los vemos como si estuviesen situados en el plano complejo, tenemos que cada punto bidimensional (x, y) se reduce a un número complejo $x + jy$. Si aplicamos la transformada de Fourier a la secuencia de puntos del contorno, obtenemos la respuesta espectral $F(u)$. Utilizando las primeras componentes de baja frecuencia de $F(u)$, se pueden distinguir curvas que sean relativamente distintas en su forma. La normalización de la transformada de Fourier ante diferentes tamaños y rotaciones, se consigue multiplicando $F(u)$ por una constante y por $e^{j\theta}$, respectivamente.

Como ejemplo práctico de la utilización de los descriptores de Fourier podemos presentar el trabajo de Ghazanfari [62], en donde se presenta la clasificación de pistachos en la categoría de cerrados o abiertos. La característica discriminante de un pistacho abierto es la aparición de discontinuidades en su contorno. Para captar esta circunstancia el autor selecciona los siete armónicos de Fourier más discriminantes; con ellos, y un clasificador neuronal consigue clasificaciones con precisiones del orden del 95%.

4.1.2.2 Descriptores de regiones.

Descriptores básicos de regiones. En las aplicaciones de tiempo-real, donde debido a las limitaciones de tiempo, los algoritmos desarrollados deben ser computacionalmente eficientes, se utilizan mucho descriptores sencillos de calcular. El uso de estos descriptores, se limita a casos en los que los objetos a detectar son fácilmente distinguibles y solo se requiere un conjunto limitado de ellos. A continuación se citan algunos de estos descriptores o características:

- *Momento de área.* El momento de área de orden pq, se define así

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y) \tag{4.17}$$

- *Área.* El área de una región se define como el número de puntos que contiene. Utilizando la definición de momento de área, tenemos que área $= m_{00}$.
- *Orientación.* El ángulo θ de orientación de una región, se define considerando el eje de menor momento de inercia, de esta manera

$$\theta = 0.5 \arctan \left[\frac{2(m_{00}m_{11} - m_{10}m_{01})}{(m_{00}m_{20} - m_{10}^2) - (m_{00}m_{02} - m_{01}^2)} \right] \tag{4.18}$$

- *Excentricidad.* Es la relación entre las longitudes de los ejes mayor y menor del área.
- *Perímetro.* Es el número de puntos en el contorno del área.
- *Compacidad.* Se define como el perímetro^2/área.
- *Número de Euler.* Es el número de regiones conexas, menos el número de agujeros en la región.

Descripción de regiones por textura. No existe una definición formal de lo que es textura, pero un descriptor de textura debe dar una medida cuantitativa de la suavidad, rugosidad y regularidad de una superficie. Existen dos aproximaciones principales a la caracterización de la textura, mediante estudios estadísticos y estructurales. Las descripciones estadísticas tratan de medir el grado de suavidad, rugosidad o granularidad de las regiones, y las descripciones estructurales se basan en la distribución espacial regular de determinados patrones.

Una aproximación sencilla para describir la textura, es el uso de los *momentos del histograma de intensidad* de la imagen o de una región. Si representamos por i a los diferentes niveles de intensidad , la media de las intensidades por m y el histograma de intensidad por $h(i)$, el momento μ_n de orden n, se define según la ecuación 4.19.

$$\mu_n = \sum_{i=0}^{N} (i-m)^n h(i) \tag{4.19}$$

Se deduce según la ecuación 4.19, que $\mu_0 = 1$ y $\mu_1 = 0$. Sin embargo el momento de orden 2, se corresponde a la varianza del histograma y es una medida del contraste, o si trabajamos con su inversa obtenemos una medida de la suavidad de la imagen. El tercer momento, μ_3, es una medida de la distorsión , y μ_4 es una medida de la planitud. Sin embargo, estas medidas (debido a que trabajan con el histograma) carecen de la información correspondiente a la posición relativa de cada punto dentro de la imagen.

Una forma de conservar la información espacial, es la utilización de *matrices de coocurrencia.* Si definimos un operador de posición P, que dé una salida binaria indicando si se cumple o no una determinada relación espacial entre dos puntos de intensidad i y j,

podemos generar una matriz A donde cada elemento a_{ij} indica el número de veces que se ha cumplido en la imagen el operador espacial P entre dos puntos con intensidad i y j. Un ejemplo de operador espacial P puede ser el siguiente: "El punto a la derecha de uno con intensidad i debe tener intensidad j". La matriz A será siempre cuadrada y tendrá tantas filas como número de niveles de intensidad. Para obtener la matriz de coocurrencia C normalizamos la matriz A mediante un factor que es igual a la suma de todos sus elementos. Haralick propone los siguientes descriptores a partir de la matriz C [73]:

- Probabilidad máxima: $max_{i,j}(C_{ij})$
- Momento de diferencia de elementos: $\sum_{i,j}(i-j)^n C_{ij}$
- Momento de diferencia inverso: $(\sum_{i,j} C_{ij})/(i-j)^n$
- Entropia: $-\sum_{i,j} C_{ij} \log C_{ij}$
- Uniformidad: $\sum_{i,j} C_{ij}^2$

Por otro lado las propuestas estructurales, consideran que una simple *primitiva de textura*, puede ser utilizada para formar patrones más complejos, mediante la aplicación de una serie de reglas de generación de patrones. Mediante técnicas de reconocimiento estructural de patrones, se pueden detectar las primitivas y en función de la estructura en la que estén dispuestas, se puede determinar la clase de patrón al que pertenece una región de la imagen.

Descripción de regiones mediante esqueletos. Una manera de representar la forma estructural de una región, es mediante *esqueletos*, que se obtienen aplicando algoritmos de adelgazamiento a dicha región. El algoritmo propuesto por Blum, se conoce como *transformación de eje medio* (MAT-Medial Axis Transformation). Consiste en representar la estructura básica de la región mediante ejes. Para obtener los ejes se recorren todos los puntos de la región, y por cada uno de ellos se comprueba si existen dos puntos del contorno de la región que están a la misma distancia del punto que actualmente se está visitando. Si la condición anterior se cumple, entonces ese punto visitado pertenece al eje medio. Sin embargo, este algoritmo es prohibitivo desde un punto de vista de eficacia computacional.

Existen algoritmos más eficaces, como el propuesto por Naccache en 1984. Utiliza un conjunto de 4 máscaras de tamaño 3×3, con las cuales determina si un punto interior a una región no tiene características de esqueleto, con lo cual debe ser marcado. Una vez aplicado el algoritmo iterativamente por todos los puntos de la imagen, los marcados se eliminan y los restantes constituyen el esqueleto de la región bajo análisis.

Descripción de regiones por momentos invariantes. Existen unos descriptores que son invariantes a cambios en el tamaño, orientación y translación de la región segmentada en la imagen, a estos momentos se les conoce como *momentos invariantes*. Para obtener su expresión de una forma abreviada, se suelen definir en función de otros momentos no invariantes como el momento central μ_{pq} (ec. 4.20) y momento central normalizado η_{pq} (ec. 4.21).

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \tag{4.20}$$

$$\eta_{pq} = \mu_{pq} / \mu_{00}^{(p+q)/2+1} \tag{4.21}$$

De esta forma, se definen los siguientes momentos invariantes:

$$\phi_1 = \eta_{20} + \eta_{02} \tag{4.22}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \tag{4.23}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{4.24}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{4.25}$$

4.1.2.3 Descripción de estructuras tridimensionales

Es bastante aceptado que la utilización de información tridimensional, juega un importante papel a la hora de desarrollar un sistema de visión versátil en entornos no estructurados. Sin embargo, y a pesar que este área de investigación data de hace más de 20 años, los sistemas de visión tridimensional no son muy frecuentes en la industria. Factores como el coste, la complejidad y la velocidad han limitado su uso.

La información tridimensional se puede obtener directamente a través de sensores que captan distancia, como telémetros láser, sensores táctiles, ultrasonidos, etc., o mediante inferencias a partir de imágenes de intensidad puramente bidimensionales, como en el caso de la visión esteroscópica o mediante el estudio de los gradientes de iluminación. En cualquiera de los casos se suele hablar de imágenes de $2^{1/2}D$ debido a que la información tridimensional es incompleta, al solamente captarse datos de una parte de la superficie de los objetos, ya que la zona no visible no es digitalizada a no ser que el objeto sea rotado.

Una vez que disponemos de la información tridimensional, la imagen consta de un conjunto de puntos (x, y, z), cada uno de ellos con tres coordenadas. También existe la posibilidad de organizar la información de una forma similar a cuando tenemos imágenes de intensidad, es decir, utilizando la función bidimensional $f(x, y)$, pero en este caso el valor de f ya no es la intensidad sino una distancia. En cualquiera de las representaciones, los algoritmos de descripción de superficies tridimensionales que vamos a ver a continuación son válidos.

Ajuste de planos a la superficie. Una de las formas más directas de describir, y al mismo tiempo segmentar una imagen tridimensional consiste en ajustar pequeñas áreas de la imagen mediante planos y, posteriormente, combinar esas regiones en entidades de superficie mayores. Primeramente, se ajusta un grupo pequeño de puntos a un plano y se calcula, de cada grupo, el vector unitario normal al plano. Seguidamente se van uniendo los grupos que tengan unas direcciones próximas y sean adyacentes. Estas regiones resultantes, se clasifican p.ej. como curvas, planas o indefinidas, y finalmente, dichas regiones clasificadas, se ensamblan al agrupar regiones adyacentes del mismo tipo. El resultado final es una imagen segmentada y con un descriptor asociado a cada segmento (en este caso segmento curvo o plano) [8, 48].

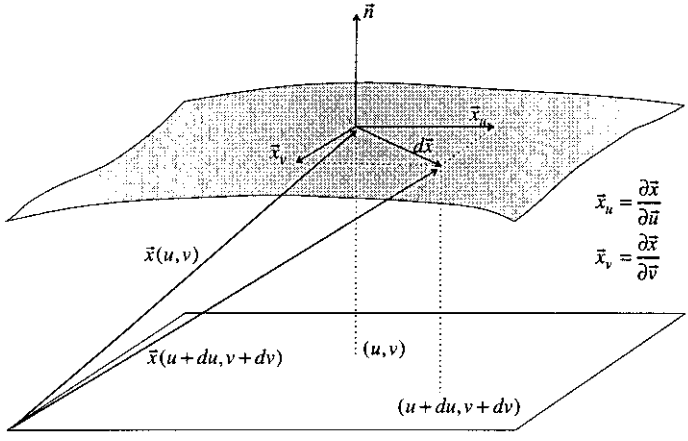


Figura 4.8: Representación de los componentes que definen la normal a una superficie.

Gradiente, normal y curvatura. Este tipo de parámetros pueden ser utilizados para caracterizar superficies, dando información que permita realizar una clasificación inicial grosera. Como ya se vio en la sección 4.1.1, el *gradiente* es un vector cuya dirección es la de máximo cambio y la magnitud es proporcional a dicho cambio. La *normal* a una superficie en un punto (u, v) (fig. 4.8), viene dada según la geometría diferencial, por la ecuación

$$\vec{n}(u,v) = \frac{\vec{x}_u \times \vec{x}_v}{\| \vec{x}_u \times \vec{x}_v \|} \tag{4.26}$$

La *curvatura* k de una superficie es un escalar, y mide el cambio que se produce en la normal de la superficie cuando nos desplazamos por ella. Curvaturas con magnitud distintas de cero indican que en torno a un punto las normales a la superficie cambian, mientras que el signo asociado a la curvatura indica si la superficie es cóncava (positiva) o convexa (negativa). La curvatura normal, k_{normal}, se define de esta forma

$$k_{normal} = \frac{-d\vec{x}d\vec{n}}{\| d\vec{x} \|^2} \tag{4.27}$$

Existen otras versiones de curvatura que no contienen toda la información tridimensional de la superficie, pero sí la esencial. Nos referimos a la curvatura media H y a la curvatura Gausiana K. A partir de ellas, y considerando solo los signos, se pueden derivar 8 tipos de superficies [15]. Otras formas habituales de estimar la curvatura de una superficie emplean ajustes por mínimos cuadrados de polinomios [125] ó utilizan métodos como el introducido recientemente por Matas y conocido como *diferenciado por filtro mediana* [139]. Una completa revisión de diversos métodos para estimar la curvatura ha sido presentada por Worring [216].

Etiquetado de líneas y vértices. Dada una imagen tridimensional, podemos representar los bordes tipo escalón mediante líneas, las cuales en los puntos de intersección forman vértices. Analizando las superficies a ambos lados de las líneas, podemos deducir

si el borde que representan es convexo o cóncavo, asignando una etiqueta a cada tipo de línea. Por otra parte, se crea un diccionario de posibles vértices, los cuales permiten hacer un análisis posterior basado en reglas heurísticas que nos posibilitan clasificar las distintas superficies, o incluso, deducir si varias de ellas pertenecen a un determinado objeto tridimensional (un cubo, esfera, paralelepípedo, etc.). Por ejemplo, si se detecta un vértice formado por la intersección de tres líneas convexas, existe evidencia de que las tres superficies implicadas sean las caras visibles de un cubo. Normalmente estos algoritmos funcionan bien para escenas muy sencillas, formadas por objetos geométricos clásicos y sin la presencia de otros elementos perturbadores que dificulten su visión.

Conos generalizados. Con el objetivo de obtener representaciones de objetos tridimensionales válidas para aplicarlas al reconocimiento mediante emparejamiento de modelos, los *conos generalizados* son bastante adecuados. Un cono generalizado es un volumen que se obtiene al trasladar una sección plana, a lo largo de una curva arbitraria, mientras el ángulo a la curva se mantiene constante y la sección se transforma de acuerdo a una regla de barrido. Por ejemplo, un cilindro se obtiene al desplazar una sección circular a lo largo de un eje recto. Si la sección circular aumenta de diámetro mientras se avanza a lo largo del eje, entonces obtenemos un tronco de cono.

Cuando tenemos un conjunto de puntos, que representan un objeto susceptible de ser representado mediante conos generalizados, la primera tarea consiste en obtener el eje central del objeto y posteriormente la sección que mejor se ajusta a los datos. De esta forma es posible realizar el reconocimiento al comparar la descripción obtenida del objeto desconocido, con el conjunto de descripciones almacenadas en la base de conocimiento de objetos válidos.

4.1.3 Reconocimiento o clasificación

Un algoritmo de reconocimiento o clasificación, esencialmente, debe identificar los objetos segmentados en una escena y asignarles una etiqueta. Para poder llevar a cabo esta identificación, los algoritmos de reconocimiento utilizan las descripciones obtenidas previamente. La etiqueta que se asigna a cada objeto, suele ser un nombre indicando el tipo de objeto de que se trata (p.ej cubo, silla, esfera, gato, etc.), y opcionalmente, se suelen dar algunos datos descriptivos adicionales, como las dimensiones, la posición y orientación del objeto. Por tanto, podemos concluir que el reconocimiento es básicamente un proceso de etiquetado.

Existen dos categorías principales de reconocimiento: los métodos basados en la *teoría de decisiones* y los métodos *estructurales*. Las aproximaciones que se engloban bajo el concepto genérico de *teoría de decisiones*, están basados en descripciones cuantitativas. A diferencia, los *métodos estructurales* utilizan descriptores simbólicos y las relaciones existentes entre ellos.

4.1.3.1 Teoría de decisiones

Clasificación clásica mediante funciones de decisión. Después de la etapa de descripción, se suele tener un grupo de descriptores o características que representan al objeto o segmento bajo análisis. Si este grupo de características, son variables numéricas o las asimilamos a valores, podemos expresarlas en una forma más compacta

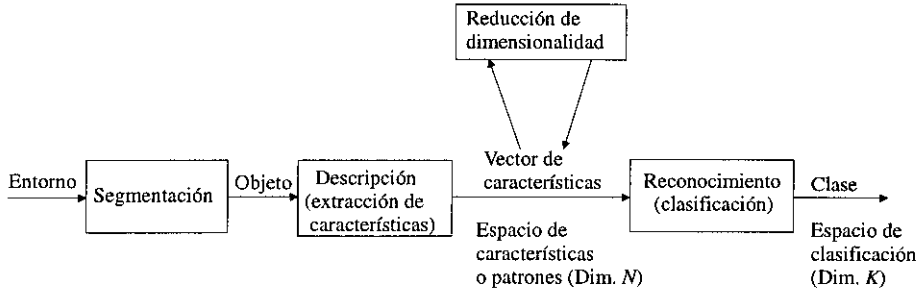


Figura 4.9: Diagrama de bloques del proceso de reconocimiento mediante extracción de características.

mediante un vector de características $\vec{x}$. Este vector contiene la descripción del objeto que pretendemos clasificar. Por otro lado, tenemos un conjunto finito de K clases $S_k(k = 1..K)$, a las que puede pertenecer el objeto. Para poder averiguar a que clase pertenece el objeto desconocido, se utilizan unas funciones discriminantes o de decisión $g_k(\vec{x})$. Cada una de las funciones discriminantes, miden el grado de similitud del objeto desconocido, representado por $\vec{x}$, con la clase bajo comparación S_k. El objeto desconocido se asignará a la clase cuya función discriminante sea máxima. De esta forma el objeto dejará de ser desconocido y se le asignará la etiqueta correspondiente a la clase ganadora.

Algunos autores, en el proceso de reconocimiento, distinguen entre tres espacios: espacio de patrones de dimensión R, espacio de características (*N-dimensional*) y el espacio de clasificación que tiene K dimensiones. El proceso mediante el cual se pasa del espacio de patrones al de características, consiste en una reducción de dimensionalidad con el objetivo de seleccionar solamente las características más discriminantes, y por tanto reducir la complejidad del sistema. En esta breve descripción que estamos haciendo, no vamos a hacer distinción entre el espacio de patrones y características, con lo cual hablaremos independientemente de uno o de otro refiriéndonos al mismo concepto (fig. 4.9).

La selección del espacio de características, es el aspecto más importante de un sistema de reconocimiento basado en estos principios. Si las características elegidas no son suficientemente discriminatorias entre ellas, difícilmente vamos a ser capaces de averiguar la clase de los objetos. Incluso usando la función discriminante más sofisticada, si las características no son discriminantes, no existirá ninguna posibilidad de obtener clasificaciones correctas. A pesar de ello, la mayor parte de la investigación se centra en el análisis y mejora de las funciones discriminantes, mientras la tarea de selección de las características óptimas ha recibido una menor atención. Dicha selección, suele realizarse de forma manual, pareciéndose más a un arte que a una ciencia. Bien es cierto, que existe un conjunto de transformaciones matemáticas que permiten reducir la dimensionalidad del espacio de características, pero no siempre de una forma completamente satisfactoria.

Para obtener un sistema de reconocimiento satisfactorio, es condición indispensable que se cumplan las siguientes propiedades:

- Debe haber una pequeña varianza dentro de cada clase S_k y la separación entre clases diferentes debe ser lo mayor posible. Es decir, los vectores de la misma clase deben ser muy parecidos y los que corresponden a clases diferentes deben ser lo más distintos que se pueda.

- Las características utilizadas para formar el espacio de patrones, deben ser insensibles a cambios en el tamaño, la orientación o la posición del objeto a clasificar en la imagen. Si esto no es así, las clases ocuparán un mayor volumen en el espacio de patrones y la posibilidad de solapamiento entre clases crecerá, con lo cual el desempeño del sistema se verá notablemente afectado.

- La dimensionalidad N del espacio de características debe ser tan pequeña como sea posible.

Una vez que el espacio de características es el adecuado, y por tanto, las clases son visiblemente discernibles, llega el momento de elegir las funciones de decisión adecuadas. Existen muchos tipos de funciones discriminantes, pero todas ellas tratan de medir el grado de similitud entre un vector $\vec{x}$, que representa a un objeto desconocido, y una clase de pertenencia S_k. La clase S_k suele constar de un conjunto de M_k muestras de vectores $\vec{y}_m^{(k)}$, que corresponden a objetos conocidos (ya que se sabe que pertenecen a S_k) utilizados para crear un modelo de la clase de pertenencia. La definición de las funciones de decisión dependen del tipo de modelo utilizado para representar las clases. Existen versiones deterministas, que asignan un vector como modelo representativo de una clase, y versiones estadísticas, que trabajan con la probabilidad de que un vector pertenezca a una clase determinada. Dentro de los modelos estadísticos, se suele hablar de clasificación paramétrica (asume una determinada función de densidad de probabilidad y se estiman sus parámetros) y no paramétrica (la función de densidad no es conocida). Sin embargo, esta distinción puede llegar a ser un tanto engañosa puesto que tanto en los modelos deterministas como en las dos versiones probabilísticas, se trabaja con parámetros.

Tanto en un caso determinista como en el probabilístico, la forma de obtener estos modelos se puede hacer manualmente o mediante aprendizaje. En ambos casos, se habla de métodos supervisados, puesto que debe haber un experto supervisando el proceso de formación del modelo. Existen métodos de generación de modelos no supervisados, donde a priori, ni el propio diseñador sabe cuantas clases existen y es el propio sistema de aprendizaje, el que deduce las clases predominantes mediante técnicas de agrupamiento (clustering). Algunas de las funciones de decisión más utilizadas en la literatura son las siguientes:

- *Los n vecinos más cercanos.* Dada una observación $\vec{x}$, decimos que pertenece a la clase S_k, si los n vecinos más cercanos pertenecen en su mayoría a la clase S_k. Para determinar el concepto de cercanía se puede utilizar la distancia euclídea. La distancia entre un vector $\vec{x}$ y la muestra m de la clase S_k, se representa por $d(\vec{x}, \vec{y}_m^{(k)})$ y se define así:

$$d(\vec{x}, \vec{y}_m^{(k)}) = \sqrt{\sum_{i=1}^{N} (x_i - y_{mi}^{(k)})^2} \tag{4.28}$$

- *El vecino más cercano.* Es un caso particular del caso anterior, considerando $n = 1$. Por tanto basta con tomar como clase de $\vec{x}$, la clase a la que pertenezca el vector $\vec{y}$ más próximo.

- *Regla de Bayes.* Asumiendo una función de densidad normal o Gausiana y una probabilidad a priori igual para todas las clases, se puede deducir la ecuación 4.29, donde Φ_k es la matriz de covarianza y $\vec{\mu}_k$ es el vector promedio correspondiente a la clase S_k. El valor mínimo de $g_k(\vec{x})$, permite averiguar la clase más probable.

$$g_k(\vec{x}) = log(|\Phi_k|) + (\vec{x} - \vec{\mu}_k)\Phi_k^{-1}(\vec{x} - \vec{\mu}_k) \tag{4.29}$$

- *Distancia media ponderada.* Cada clase viene representada por un vector promedio $\vec{\mu}_k$, que es utilizado para calcular la distancia con el vector del objeto desconocido $\vec{x}$. Sin embargo esta distancia está ponderada al tener en cuenta la desviación estándar $\sigma_i^{(k)}$ en cada componente del espacio de la clase. Esta distancia ponderada se expresa en la ecuación 4.30.

$$d(\vec{x}, S_k) = \sum_{i=1}^{N} \left\{ \frac{|x_i - E(y_i^{(k)})|}{\sigma_i^{(k)}} \right\} \tag{4.30}$$

Las funciones discriminantes tienen la propiedad de que parten el espacio de características o de patrones en volúmenes mutuamente excluyentes, mediante hipersuperficies de dimensión $N-1$ (fig. 4.10). Cada uno de estos volúmenes constituyen una clase, siendo también posible que esté formada por varios subvolúmenes no contiguos. Las funciones discriminantes más sencillas son las lineales, que generan hiperplanos como superficies de separación entre clases. Las funciones cuadráticas generan hipersuperficies que delimitan mejor que los hiperplanos la frontera de las clases.

Comparación de plantillas (Template matching). El reconocimiento de objetos requiere la comparación de descripciones con modelos de objetos conocidos. Como vimos anteriormente, hay dos clases fundamentales de reconocimiento: la que trabaja con descripciones cuantitativas y la que utiliza descriptores simbólicos y sus relaciones. A parte de esta diferenciación, también podemos clasificar las técnicas de reconocimiento, según la manera en que proceden, en dos grupos: 1) Técnicas de búsqueda de la presencia de un objeto conocido, y 2) Técnicas de detección de la clase a la que pertenece un objeto desconocido. A este segundo caso pertenecen la mayoría de las técnicas; se trata de comparar un objeto desconocido con diferentes modelos almacenados, para ver con cual se obtiene la mejor correspondencia. Por el contrario, en el primer caso se busca un determinado objeto a lo largo de toda la imagen. En esta categoría recae la técnica conocida como *comparación de plantillas* (template matching), y que vamos a describir a continuación.

La técnica consiste en mover una plantilla a lo largo de toda la imagen. La plantilla es una pequeña matriz, normalmente de menos de 64×64 puntos, que suele contener la imagen del objeto que se busca. Mientras se va moviendo la plantilla por la imagen, para poder medir el grado de similitud en la zona de superposición, se utilizan unas métricas. A continuación se describen las métricas más comúnmente utilizadas, donde $\psi(i,j)$ es la plantilla y $f(i,j)$ es la imagen:

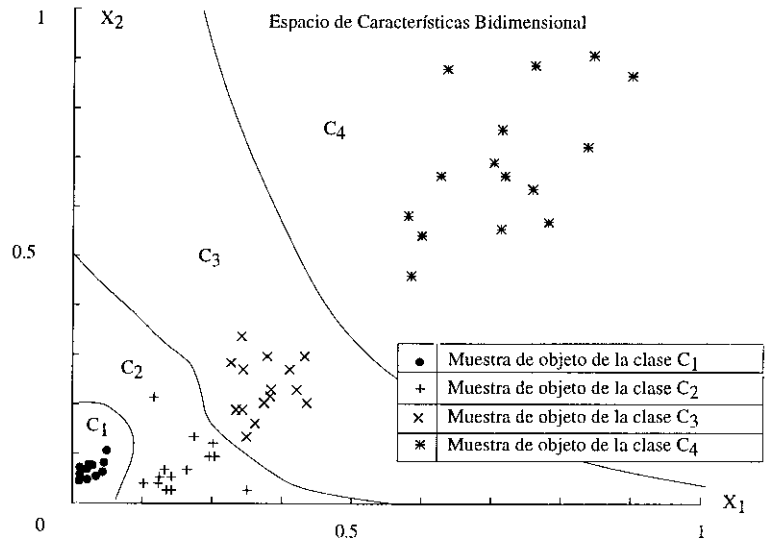


Figura 4.10: Espacio de características donde se aprecia como objetos pertenecientes a la misma clase tienden a agruparse, pudiéndose utilizar funciones discriminantes para realizar la clasificación.

- *Suma de diferencias y de cuadrados de diferencias.* Las métricas de las ecuaciones 4.31 y 4.32, acumulan los errores que se producen al comparar la plantilla con la región correspondiente en la imagen. Sin embargo, requieren que las ventanas a comparar tengan niveles de intensidad semejantes. Además, no aceptan cambios en escala y en rotación.

$$e(x,y) = \sum_i \sum_j |f(i,j) - \psi(i,j)| \tag{4.31}$$

$$e(x,y) = \sum_i \sum_j (f(i,j) - \psi(i,j))^2 \tag{4.32}$$

- *Coeficiente de correlación normalizado.* Esta técnica trata de maximizar la relación señal ruido (SNR-Signal to Noise Ratio) definida como el cociente entre la respuesta a la plantilla buscada dividida por la respuesta al ruido de fondo. La correlación es un caso particular del "Matched filter" cuando el ruido considerado es de tipo gausiano. Los principales aspectos negativos a destacar son su respuesta significativa en las proximidades de la posición óptima de detección, con lo cual los picos de correlación son suavizados; y que se vé afectado fuertemente por la oclusión. Para que la métrica sea insensible a los valores absolutos de intensidad, se debe utilizar el coeficiente de correlación normalizado, que se define según la ecuación 4.33. Los valores de γ, variarán entre 1 y -1, siendo 1 el valor correspondiente a un ajuste perfecto. Si además se quiere trabajar con objetos a diferentes escalas y orientaciones, es necesario extender el conjunto de plantillas

para que se adapten a estos cambios.

$$\gamma(x,y) = \frac{\sum_i \sum_j (f(i,j) - \bar{f}(i,j))(\psi(i,j) - \bar{\psi}(i,j))}{(\sum_i \sum_j (f(i,j) - \bar{f}(i,j))^2 \cdot \sum_i \sum_j \psi(i,j) - \bar{\psi}(i,j)^2)^{1/2}} \tag{4.33}$$

- *Expansion Template Matching (EXM) [172].* Esta técnica está basada en aplicar un filtro que optimiza una medida de similitud conocida como relación señal ruido driscriminante (*DSNR*-Discriminant Signal to Noise Ratio). *DSNR* se define como el cociente entre la respuesta a la plantilla centrada y la respuesta a cualquier otra situación incluyendo un emparejamiento no centrado con la plantilla. De esta forma la respuesta ideal buscada es un impulso en el centro de la plantilla, lo cual diferencia esta técnica de la tradicional correlación que generaba máximos suavizados lo cual acarrea problemas de detección ante oclusión y superposición de varios objetos iguales. El emparejamiento *DSNR* se corresponde con una expansión en un espacio no ortogonal de la imagen de entrada usando funciones básicas que son versiones desplazadas de la propia plantilla. Los filtros diseñados bajo este criterio son discriminantes en zonas con formas significativas de la plantilla como esquinas o zonas con alta curvatura; por tanto concentran su atención fundamentalmente en estos puntos. La expresión que nos permite obtener el filtro en el dominio de la frecuencia es:

$$EXM(w) = \frac{\Psi^*(w)}{S_\psi(w) + S_n(w)} \tag{4.34}$$

donde $\Psi(w)$ es la transformada de Fourier de la plantilla $\psi(x,y)$, y $S_\psi(w)$ y $S_n(w)$ son la densidad espectral de la plantilla y del ruido respectivamente.

Es de destacar que esta técnica de reconocimiento trabaja directamente sobre los datos presentes en la imagen, y no requiere la habitual fase de descripción. A diferencia, existen técnicas de comparación, que se verán en la sección de métodos estructurales, que trabajan con conceptos y primitivas creados en la fase de descripción.

Las técnicas de comparación de plantillas utilizadas para averiguar la clase de un objeto desconocido, que debe pertenece a un modelo dentro de una gran base de modelos, son un caso más complejo debido a su alta explosión combinatoria. Sin embargo, constituyen un método más flexible ya que, en teoría, posibilita almacenar un amplio rango de objetos en su base de conocimiento, permitiendo realizar interpretaciones del entorno completas y no restringiéndose a aplicaciones donde el número de objetos es muy limitado. Un ejemplo significativo es el sistema de reconocimiento MULTI-HASH [115] donde se plantea un proceso de generación de varias hipótesis y mediante una verificación posterior se selecciona el candidato más probable. Para extraer las hipótesis iniciales se utiliza una representación incompleta pero con un cómputo de comparación muy eficiente. Esto es posible ya que se utiliza un conjunto de características locales unidas mediante relaciones y un almacenamiento en tablas HASH. En la fase de verificación, al trabajar con un conjunto de hipótesis reducido, se aplica una comparación más completa mediante una esfera de Gauss mejorada.

Clasificación mediante redes neuronales artificiales. Las redes neuronales artificiales están inspiradas en la versión natural que la naturaleza ha proporcionado al cerebro de los animales y del ser humano. Una red neuronal consta de un conjunto de

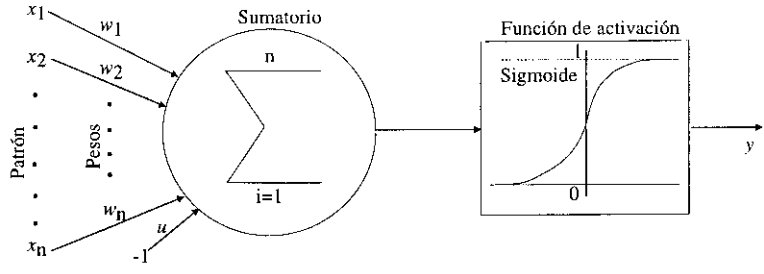


Figura 4.11: Modelo matemático de una neurona.

unidades de procesamiento muy sencillas, llamadas neuronas, que se conectan entre sí. Se caracterizan por el poder de cálculo masivamente paralelo, la capacidad de aprendizaje, la habilidad de generalización y adaptabilidad, la tolerancia a fallos y la posibilidad de cómputo y almacenamiento de los conocimientos de forma distribuida.

Una neurona se suele modelar constituida por varias entradas, un bloque de procesamiento sencillo y de una salida. El modelo matemático de una neurona es muy sencillo (ec. 4.35 y fig. 4.11), realiza básicamente una suma ponderada de sus entradas y luego aplica una función de activación g, que suele dar un valor binario que depende del resultado de la suma. Las funciones de activación más usadas son la escalón, la lineal, la sigmoide y la Gausiana.

$$y = g(\sum_{j=1}^{N} w_j x_j - u) \tag{4.35}$$

Las redes neuronales son agregados de neuronas que se suelen distribuir formando una capa de entrada, otra de salida y opcionalmente otras capas intermedias. Dependiendo de la manera en que se interconecten las neuronas entre las diferentes capas, se distinguen dos tipos de arquitecturas: lazo abierto (no hay lazos y las conexiones van de las capas de la entrada hacia las de salida), y recurrentes (existen realimentaciones). La característica más importante de las redes neuronales radica en la capacidad de aprendizaje. El aprendizaje se puede ver como el mecanismo mediante el cual los pesos w_j de cada neurona se van actualizando, hasta que finalmente la red es capaz de realizar la tarea para la cual estaba diseñada. Entre los algoritmos de aprendizaje más comunes están los siguientes: perceptrón, propagación hacia atrás, Boltzman, regla de Hebb y métodos competitivos como LVQ y Kohonen.

Las redes neuronales se pueden aplicar al reconocimiento o clasificación de patrones, de una forma similar a como trabajaban las funciones discriminantes vistas anteriormente. Si suministramos como entrada de una red neuronal un vector de características representando un objeto, y esta red consta de tantas neuronas en la última capa como clases posibles, estas redes se suelen entrenar para que una única neurona de la última capa se active, indicando la clase a la que corresponde el patrón desconocido. Por tanto, una red neuronal puede trabajar como un clasificador tradicional, y análogamente, trabaja dividiendo el espacio de características en subvolúmenes disjuntos, asociando a

cada uno de ellos una clase.

Las redes neuronales se aplican como clasificadores en muchos campos [107]. Debido a la propiedad de autoaprendizaje o aprendizaje no supervisado, también se aplican en problemas de agrupamiento (clustering). Otras áreas donde son de utilidad incluyen la aproximación de funciones, la predicción de variables, la optimización, el filtrado, el diseño de memorias direccionables por contenido y la teoría de control.

4.1.3.2 Métodos estructurales

Las técnicas discutidas en la sección anterior trabajan con patrones cuantitativos. A diferencia, los métodos estructurales analizan *patrones simbólicos*, formados por un conjunto de primitivas y las relaciones que existen entre ellas. Una primitiva es un símbolo que representa componentes sencillas de un objeto (p.ej. un segmento recto de longitud constante de un contorno se puede representar mediante el símbolo "a", y un segmento curvo por "b"). La idea consiste en descomponer un objeto en una lista de primitivas y en las relaciones que existen entre ellas. El caso más sencillo de relación es la concatenación, donde dos primitivas consecutivas indican que una parte del objeto consiste en la concatenación de las componentes representadas por las primitivas. Una vez que se obtiene la descripción en términos de primitivas, se debe realizar su análisis. Los métodos fundamentalmente utilizados son dos: la comparación de patrones simbólicos y el análisis sintáctico o estructural.

Comparación de patrones simbólicos. Un ejemplo de descripciones simbólicas son los código cadena y los números de forma. En estos casos, para averiguar si dos descripciones representan al mismo objeto, se deben utilizar *medidas de similitud* [192]. Dos posibles medidas de similitud son las siguientes:

- Dadas dos descripciones simbólicas A y B, podemos utilizar como medida de similitud el mayor orden n para el cual existe coincidencia entre cada una de las primitivas. Es decir, $A(1) = B(1), \ldots, A(n) = B(n), A(n+1) \neq B(n+1)$.
- Si representamos por α al número de coincidencias locales entre las descripciones A y B, y por β el número de veces que no coinciden, podemos definir la medida de similitud γ como $\gamma = \alpha/\beta$.

Análisis sintáctico. Cuando los objetos a describir son complejos, no es adecuado realizar una correspondencia directa entre las descripciones simbólicas como vimos anteriormente, pues pequeñas diferencias entre objetos de la misma clase, generarán medidas de similitud muy bajas y el reconocimiento será fallido. En este caso, el proceso de análisis adecuado es el jerárquico, mediante el cual subpatrones de primitivas se agrupan en primitivas de mayor nivel de abstracción, y estas a su vez se agrupan en otras de mayor nivel. Finalmente, se representa el objeto mediante una sola primitiva de alto nivel que es la etiqueta o el nombre del objeto reconocido. Este análisis jerárquico se realiza mediante los conocidos *métodos sintácticos*.

La estructura de un sistema de reconocimiento de patrones sintáctico se puede ver en la figura 4.12. Se pueden observar dos fases diferenciadas: diseño y reconocimiento. En la fase de diseño se seleccionan las primitivas que se van a utilizar y el conjunto de reglas que definen los objetos válidos. En la fase de reconocimiento, se comienza

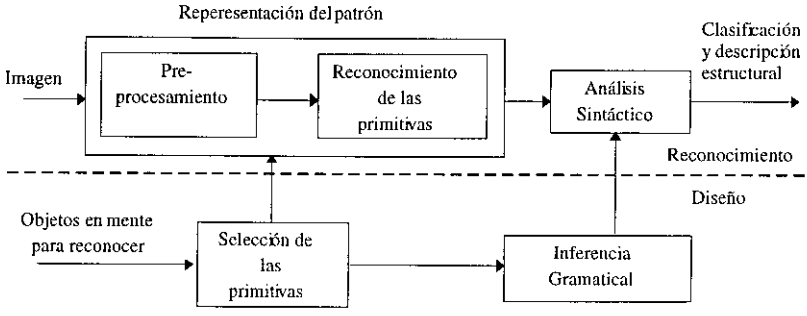


Figura 4.12: Diagrama de bloques de un sistema de reconocimiento de patrones sintáctico.

identificando las primitivas que describen al objeto, las cuales son suficientemente simples como para poderse reconocer mediante algoritmos sencillos. Posteriormente se realiza un análisis sintáctico del patrón de primitivas, determinándose si es sintácticamente correcto con respecto a una serie de gramáticas, y por tanto reconociendo el objeto. Como consecuencia del análisis, se genera una descripción estructural del objeto, normalmente, en forma de árbol, mediante el cual se aprecia el proceso de razonamiento seguido para deducir la clase a la que pertenece el objeto.

En la figura 4.13 se puede ver una imagen compuesta por dos objetos (un cuadrado y un rectángulo). Después de un reconocimiento de primitivas, se asignan los símbolos "a" y "b" en las apariciones de segmentos rectos y curvos respectivamente. De toda la imagen, obtenemos un vector de símbolos (hojas del árbol), a partir del cual y mediante abstracciones sucesivas se realiza el reconocimiento de los objetos en la imagen. Las reglas aplicadas y presentes en la gramática indican que un rectángulo se compone de una esquina ("b") seguida de un lado, repetido cuatro veces, y además los lados opuestos deben tener igual longitud. Un lado consta de un segmento recto ("a") o de varios encadenados. Y un cuadrado es un rectángulo con todos los lados iguales.

Cada gramática, mediante un conjunto de reglas, expresa una clase de objetos válidos o reconocibles, y también el proceso lógico mediante el cual se pasa de las primitivas elementales a la etiqueta final. Para aplicar este análisis en el problema de reconocimiento, se deben utilizar tantas gramáticas como objetos válidos consideremos. Cada gramática representa el conjunto posible de patrones simbólicos aceptables que serán interpretados como un objeto determinado.

Formalmente las gramáticas se definen como una 4-tupla $G = (V_n, V_t, P, S)$, donde V_n son elementos no terminales (denotados por $A, B, \ldots$), V_t son las primitivas o elementos terminales (denotados por $a, b, \ldots$), P son las reglas y S es el símbolo final o etiqueta del objeto a reconocer. En función de el tipo de las reglas permitidas se distingue entre diferentes gramáticas:

- *Gramáticas sensibles al contexto.* Tienen reglas de la forma $\zeta_1 A \zeta_2 \rightarrow \zeta_1 A \zeta_2$.

- *Gramáticas libres de contexto.* Cada elemento no terminal A puede ser sustituido por una cadena β, independientemente del contexto en que aparezca A. En general

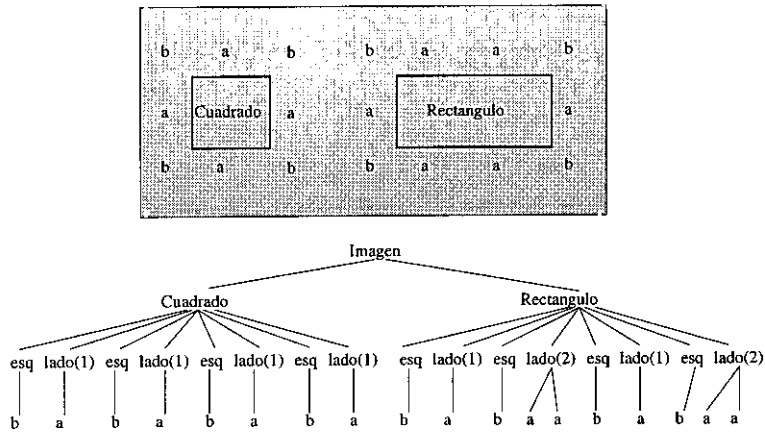


Figura 4.13: Proceso de asignación de primitivas, reconocimiento y descripción estructural de una imagen mediante técnicas sintácticas.

$A \rightarrow \beta$.

- *Gramáticas regulares o de estados finitos.* Las reglas siempre son de estas dos posibles formas: $A \rightarrow aB$ y $A \rightarrow b$. Tienen la ventaja de que el algoritmo de análisis de los patrones puede ser realizado mediante un sencillo autómata de estados finitos, aunque la expresividad del lenguaje es limitada.

4.2 Estrategia de reconocimiento propuesta

4.2.1 Consideraciones iniciales sobre métodos de reconocimiento

Hemos revisado en los apartados anteriores los fundamentos y las técnicas de análisis más extendidas en la visión artificial. El proceso tradicional se divide en las etapas de segmentación, descripción de cada uno de los segmentos, y finalmente, en base a esos descriptores, se realiza una clasificación o reconocimiento asignando una etiqueta y unos parámetros a los objetos detectados. Estas estrategias funcionan bastante bien cuando las imágenes y los objetos a reconocer son bastante ideales. Sin embargo, en un caso real como es el que pretende resolver esta tesis, es frecuente encontrarse con situaciones en las que la captación tiene un alto contenido de ruido, la escena no está apenas estructurada y aparecen problemas de oclusiones parciales de los objetos. Bajo estas condiciones la primera etapa del proceso de reconocimiento, que busca obtener segmentaciones ideales, se hace muy compleja ya que tiende a aparecer una segmentación excesiva caracterizada porque un objeto no viene representado por un segmento, sino que aparecen múltiples, con lo cual se requiere aplicar otras estrategias que agrupando diversos segmentos y almacenando las relaciones entre ellos, traten de determinar la presencia del objeto u objetos buscados. Existen soluciones parciales a esta problemática, pero en ningún caso de tipo general, y a consta de incrementar enormemente la complejidad de los algoritmos.

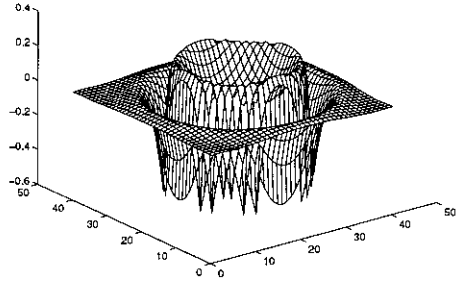


Figura 4.14: Fitro diseñado para detectar esferas aplicando la técnica EXM.

Otro método más directo que es especialmente interesante cuando el universo de objetos a buscar es reducido, es el emparejamiento de plantillas. En nuestro caso, donde se pretende realizar un sistema de reconocimiento de objetos cuasi-esféricos para reconocer, entre otras aplicaciones, objetos tales como fruta, basta crear un modelo de objeto esférico e irlo desplazando por la imagen hasta que se detecte un alto grado de correlación. En un caso general, la búsqueda del emparejamiento con el modelo habría que hacerlo para diferentes tamaños y diferentes orientaciones de este. Sin embargo, en nuestro caso al haber simetría esférica, no se requiere realizar la búsqueda realizando cambios en la orientación, y solo se necesita variar el radio del modelo. Esto hace que el método de reconocimiento por emparejamiento con una esfera sea muy directo e incluso eficiente. A pesar de estos aspectos positivos, también existen otros factores que degradan los resultados obtenibles idealmente, como son la oclusión de los objetos buscados y la presencia de objetos diferentes al buscado pero con un conjunto de características que generan respuestas ante una correlación similares a las del objeto buscado. Llamaremos a estos objetos *seudosimilares*, ya que dan una respuesta similar al aplicar un reconocimiento basado en emparejamiento de plantillas, a pesar de diferir del objeto buscado. Como consecuencia una detección de picos en los mapas de correlación puede generar muchos errores de interpretación al detectarse tanto los objetos buscados como los *seudosimilares*.

Para ilustrar lo que acabamos de exponer vamos a realizar unos ensayos de correlación de un conjunto de imágenes de prueba conteniendo esferas y objetos *seudosimilares* a estas (p.ej discos u hojas). Aplicaremos dos de las técnicas de emparejamiento vistas anteriormente en este capítulo: correlación y filtrado EXM. A partir de un modelo de objeto esférico hemos creado una plantilla esférica de $2^{1/2}$ dimensiones para realizar la correlación, y esta misma plantilla la hemos utilizado para diseñar el filtro mostrado en la figura 4.14 para realizar el filtrado EXM.

Las imágenes de prueba utilizadas se muestran en la columna de la izquierda de la figura 4.15 y son: 1) una esfera aislada, 2) la misma esfera semiocluída por una plancha, 3) la imagen anterior con la adición de un objeto *seudosimilar* a una esfera, y 4) una escena de un naranjo conteniendo un fondo de hojas y dos naranjas, una de ellas muy visible y la otra parcialmente ocluida. Según los resultados de aplicar la correlación y el filtrado EXM sobre las imágenes (columna central y derecha de la figura 4.15 respectivamente)

podemos obtener las siguientes conclusiones:

- La respuesta del filtro EXM ante el patrón buscado es mucho más impulsiva que en el caso de la correlación, como la teoría predice [172].
- La oclusión afecta reduciendo la repuesta y generando respuestas laterales debido a objetos *seudosimilares*. Como vemos la simple plancha de oclusión genera respuestas en ambos casos, especialmente para la correlación donde el rango de objetos *seudosimilares* es mucho mayor que en el filtrado EXM que es más selectivo.
- Objetos como una hoja o un disco plano con un contorno de curvatura similar a la del objeto buscado son fuertemente *seudosimilares* a una esfera, tanto para una correlación como para el filtrado EXM, lo cual se puede ver claramente en la tercera imagen de prueba.
- Las zonas discriminantes de una esfera son los contornos o los bordes, y es precisamente aquí donde el filtro EXM centra su atención como puede verse en la forma del propio filtro (fig. 4.14). En el centro el valor del filtro es próximo a cero, por tanto solamente se realiza una convolución efectiva en busca del emparejamiento correcto justo en los contornos de la esfera. Este aspecto permite definir cuales son los objetos *seudosimilares* de una esfera aplicando EXM: "cualquier objeto con contorno curvo de igual radio que la esfera y separado del fondo". Se puede dar el caso de que estos objetos den más respuesta que la propia esfera si la separación del objeto *pseudosimilar* del fondo es mayor que la separación de la esfera con dicho fondo.
- Las escenas naturales de un árbol frutal contienen oclusiones y objetos *seudosimilares* que hacen del proceso de detección mediante esta técnica un proceso no viable.

Las técnicas de comparación de plantillas que acabamos de analizar, constituyen un proceso lineal que centran su atención en características que pueden no llegar a ser suficientemente discriminantes para realizar una correcta detección. Esto sugiere la necesidad de aplicar técnicas no lineales con capacidad de centrar la atención en características realmente discriminantes, y de esta forma evitar los problemas de ambigüedades y poder llegar a realizar reconocimientos robustos y con una mayor certeza sobre los resultados de clasificación obtenidos. Teniendo presentes estas líneas de actuación en la siguiente sección presentamos la estrategia de reconocimiento que proponemos para este tipo de problemas.

4.2.2 Estrategia de reconocimiento basada en la extracción de primitivas y en la acumulación de indicios

Cuando una persona observa una imagen donde aparecen objetos que son conocidos y no existen dificultades provocadas por escasez de iluminación o oclusiones, el reconocimiento de estos objetos se hace de una forma inmediata. No somos conscientes de haber realizando ningún esfuerzo explícito para percibir una escena de este tipo. El análisis se realiza mediante un procesamiento y transmisión de señales electro-químicas a un nivel

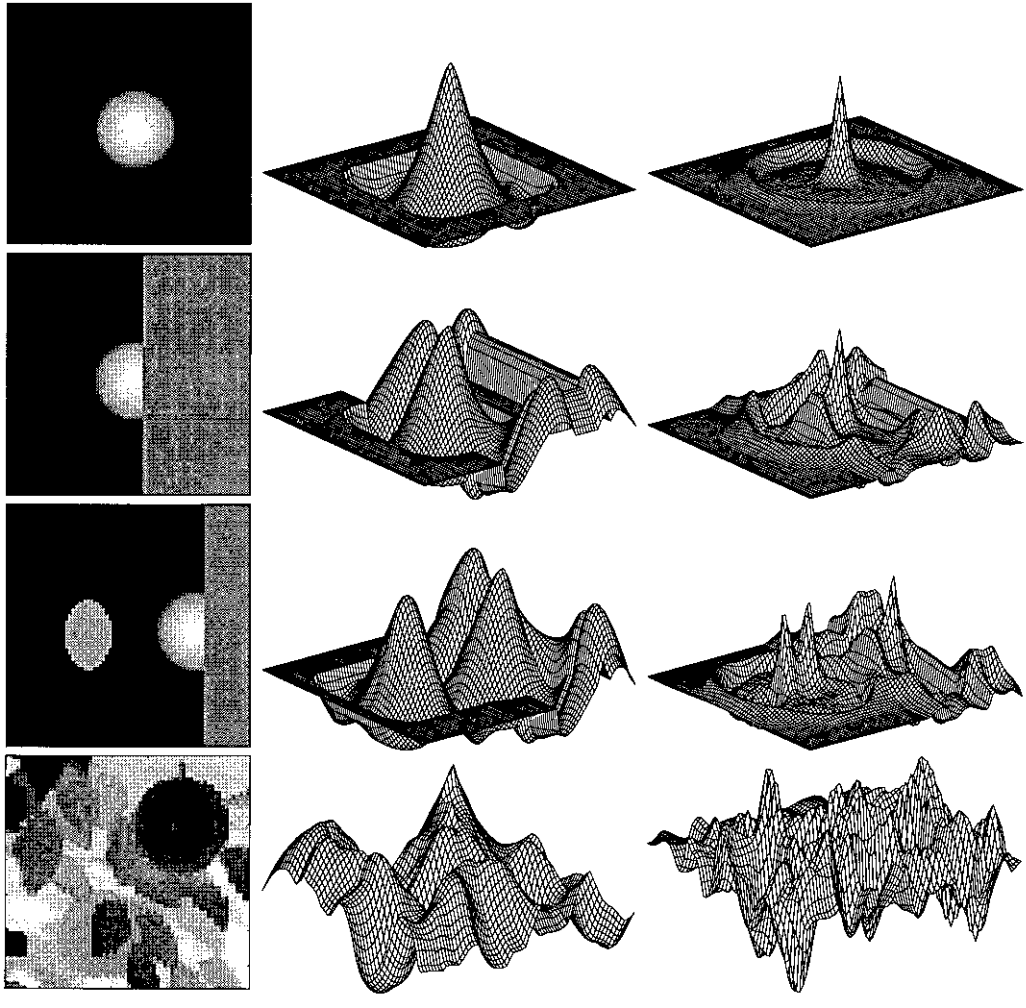


Figura 4.15: Reconocimiento mediante técnicas de emparejamiento de plantillas. Columna izquierda: imágenes de prueba sobre las cuales queremos detectar objetos esféricos; Columna central: resultado de aplicar una correlación con una plantilla esférica; Columna derecha: resultado de aplicar un filtrado EXM utilizando el filtro de la figura 4.14.

neuronal y finalmente cerebral. Sin embargo, cuando las escenas observadas son más complejas no es posible realizar una interpretación automática y el ser humano recurre a otras estrategias. Este análisis requiere un periodo temporal perceptible, durante el cual la persona es consciente de estar realizando un análisis mental para determinar el tipo de objeto que tiene ante sí. Este estudio es un proceso de análisis de características, generación de hipótesis, y finalmente, verificación de la validez de las hipótesis planteadas. El proceso es iterativo y en él se tienen en cuenta tantos parámetros característicos como sean necesarios para poder considerar una determinada hipótesis como la más probable dentro de un universo de posibles objetos. Se trata por tanto de un proceso de acumulación de indicios sobre la posibilidad que el objeto presente en la escena corresponda a una determinada clase, siendo el conjunto de objetos que hemos visto y reconocido a lo largo de nuestra vida el *universo de objetos* reconocibles sin necesitar más aprendizaje.

En el caso que nos ocupa, el *universo de objetos* válidos y susceptibles de ser reconocidos, se limita solo a un único objeto: la esfera. La estrategia de reconocimiento que vamos a presentar se basa en la definición de un conjunto de *primitivas* o componentes básicos, que son por sí solos representativos de esferas o al menos tienen una alta probabilidad de que correspondan a zonas pertenecientes a superficies esféricas. Este conjunto de primitivas constituyen un conjunto de "pistas" que mediante un proceso de generación/verificación de hipótesis, que contempla los indicios generados por cada una de ellas, permite etiquetar como objetos esféricos a aquellos que acumulan un grado de evidencia suficiente [108].

Para ofrecer una idea general de la estrategia de reconocimiento que proponemos, la figura 4.16 muestra un diagrama de bloques donde se pueden ver tres grandes etapas de procesamiento de datos:

1. Generación de primitivas.
2. Estimación de parámetros e indicios.
3. Generación de hipótesis y verificación.

La primera etapa se caracteriza por ser diferente a otras estrategias de reconocimiento que se centran en segmentaciones regionales, de contornos o en combinaciones de ambos casos para mejorar los segmentos. En nuestra aproximación no estamos interesados en segmentaciones perfectas, sino en la acumulación de "pistas" de reconocimiento (primitivas). En el presente trabajo se han definido cuatro tipos de primitivas suficientemente representativas de una esfera, pero la estrategia es modular y se podrían añadir tantas primitivas características del objeto como se estimasen necesarias. De las cuatro primitivas, dos de ellas son primitivas locales y las otras dos son regionales. Las primitivas locales se obtienen analizando un pequeño vecindario alrededor de cada punto, mientras que las regionales se obtienen a partir de segmentos homogéneos extensos.

En la segunda etapa, se estiman los parámetros de la esfera y el conjunto de indicios que indican lo fiable que es la estimación anterior. La estimación de parámetros es posible hacerla a partir de una única primitiva de tipo regional, sin embargo, en el caso de primitivas puntuales no se genera evidencia directa sobre la presencia de una esfera a partir de cada primitiva, y es necesario detectar distribuciones de ellas formando agrupaciones circulares. La definición local de estas primitivas puntuales permite que

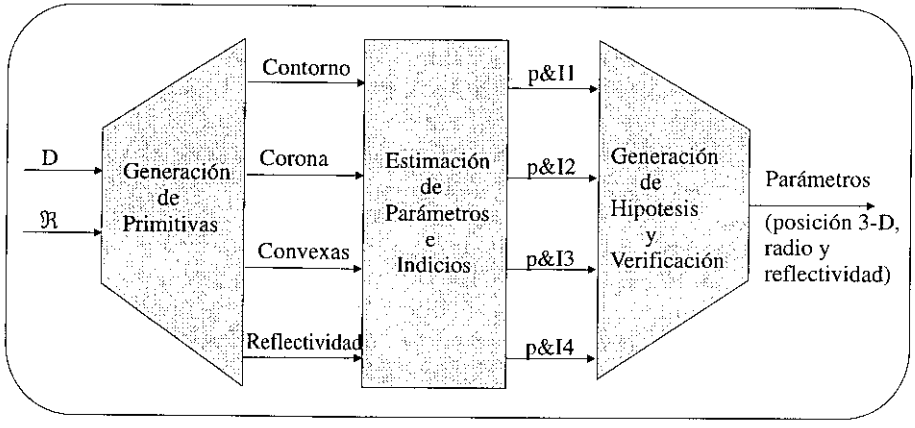


Figura 4.16: Diagrama de bloques de la estrategia general de reconocimiento.

sean eficaces ante problemas de oclusión. Aunque la oclusión reducirá el grado de evidencia generado, la capacidad de detección permanece inalterada, siempre y cuando se detecten suficientes primitivas como para generar hipótesis consistentes.

Finalmente, en la tercera etapa, los cuatro conjuntos de estimaciones parciales serán integrados de forma que las cuatro fuentes de indicios se apoyen mutuamente para generar hipótesis finales más estables y fiables. Las hipótesis generadas que sean consistentes se integrarán realizando un promedio ponderado de sus parámetros y acumulando los indicios. Las hipótesis finales con suficientes indicios serán aceptadas y las que no alcancen un umbral mínimo serán eliminadas.

En las siguientes secciones del presente capítulo se describe de una forma detallada cada una de las tres etapas de procesamiento de que consta la estrategia general de procesamiento propuesta.

4.3 Definición y generación de primitivas

El objetivo normalmente buscado en los procesos de etiquetado de puntos (pixels), se centra en la segmentación de imágenes. En unos casos se pretende obtener el contorno cerrado de objetos, delimitando de esta forma la región en la que aparece el objeto. En otros casos los puntos son etiquetados con la intención de obtener regiones correspondientes a secciones de un objeto. En cualquiera de los casos, el objetivo es realizar una segmentación, es decir, separar los objetos del fondo. Tras la segmentación, una posterior parametrización y clasificación permitiría finalizar el proceso completo de reconocimiento. Sin embargo, una segmentación perfecta es casi siempre una tarea complicada ya que muchas veces, especialmente en entornos no estructurados, los objetos están solapados no pudiéndose separar unos objetos de otros, o en ocasiones, son el fondo y los objetos los que no son separables.

A diferencia de las estrategias habituales de segmentación, nosotros planteamos una

estrategia basada en la generación de un conjunto de primitivas. El objetivo que nos marcamos a la hora de presentar esta generación múltiple de primitivas, consiste en extraer puntos y regiones características o con alta probabilidad de pertenecer a un objeto determinado, evitando al mismo tiempo la necesidad de obtener segmentaciones perfectas. Estas primitivas, de forma individual o mediante su colaboración con otras, generan un conjunto de indicios acerca de la existencia de un objeto. En nuestro caso, el objeto a considerar es la "esfera" y por tanto las primitivas discriminantes seleccionadas captan una serie de propiedades presentes en objetos esféricos.

Una estrategia que también plantea la extracción de primitivas como base para la posterior estimación de parámetros y que no necesita obtener contornos cerrados, fue presentada por Seitz introduciendo el concepto de *ejes locales de simetría* [185]. Mediante esta técnica se marcan aquellos puntos que tienen una fuerte simetría analizando solo un pequeño vecindario en su entorno. Estos puntos serán las primitivas, estando cada uno de ellos parametrizados con el ángulo θ de su eje de simetría. A partir de ellos la estimación de los parámetros del objeto buscado se hace aplicando una estrategia parecida en concepto a la transformada de Hough, pero donde la posición de los puntos no tiene apenas importancia y sí la orientación de las fronteras de los objetos que delimitan. Por ello la estrategia es robusta ante distorsiones, como una dilatación, pero existen muchos problemas especialmente a la hora de asignar las primitivas que contribuyen a un determinado objeto. Esto se debe a que las primitivas se definen mediante un concepto genérico, no habiendo sido definidas de forma específica teniendo presente el tipo de objeto a detectar. Por ello los puntos considerados como primitivas son muy numerosos ya que no solamente surgen de los ejes de simetría de los objetos buscados, sino que surgen a partir de otras entidades visibles con lo cual al haber exceso de primitivas no es evidente derivar interpretaciones inmediatas.

A diferencia del caso anterior, las primitivas que vamos a presentar son exclusivas del objeto buscado con lo cual se facilitará su posterior interpretación. Como puede apreciarse en la figura 4.16, son cuatro las primitivas propuestas en el presente trabajo. Los nombres asignados a cada una de estas primitivas son:

- *primitivas puntuales contorno*
- *primitivas puntuales corona*
- *primitivas regionales convexas*
- *primitivas regionales reflectividad*

El nombre que reciben explican por un lado la extensión espacial de estas, denominando *primitivas puntuales* a aquellos puntos de la imagen que de forma individual tienen unas ciertas propiedades que las hacen discriminantes; por *primitivas regionales* nos referimos a aquellos grupos de puntos conexos que de forma conjunta cumplen otra serie de propiedades discriminantes. Por otro lado las denominaciones de *contorno*, *corona*, *convexo* y *reflectividad* indican la propiedad fundamental que deben cumplir los puntos o regiones para ser considerados como primitivas aceptables. Las primitivas puntuales contorno y corona son puntos que pertenecen a arcos circulares sobre el contorno y corona de una esfera, respectivamente. Las primitivas regionales convexas y reflectividad son

regiones o grupos de puntos que tienen curvatura de tipo convexa y reflectividades medias propias de la superficie del objeto buscado, respectivamente.

El hecho de elegir este tipo de primitivas se justifica por los problemas de oclusión parcial que existe en los ambientes poco estructurados como los agrícolas. En una escena típica de un árbol frutal existen múltiples tipos de oclusión que afectan a la visibilidad de la fruta (p. ej. la creada por las hojas, por otros frutos o por ramas), todas ellas causando la reducción de la superficie visible de la fruta o la partición en varios segmentos de esta. Debido a que estas regiones visibles pueden corresponder tanto a zonas periféricas como interiores del fruto, las primitivas a definir van a tratar de captar propiedades discriminantes del objeto en diferentes puntos de su superficie, con el objetivo de que la oclusión parcial afecte lo mínimo posible. Por ello las primitivas contorno, que captan solo los bordes o el perfil de la esfera, serán apropiadas cuando la superficie central de la esfera este oculta pero no lo esté una sección de su contorno. Igualmente, pero en un anillo más interior que las primitivas contorno, las primitivas corona generarán indicios de esfericidad cuando no sea visible ni el borde ni el centro de la esfera, pero sí lo sea una corona sobre su superficie. En el caso de las primitivas regionales se centrará la atención en las áreas interiores de la esfera, captando la curvatura propia de una esfera mediante las primitivas convexidad o bien captando propiedades ópticas propias del objeto buscado lo cual se hace con las primitivas reflectividad. En definitiva se han contemplado estas cuatro primitivas por considerarse suficientemente significativas y complementarias entre sí como para permitir la detección de objetos esféricos ante diferentes configuraciones de visibilidad.

A modo ilustrativo, y para presentar gráficamente lo que entendemos por cada tipo de primitiva, en la figura 4.17 adelantamos los resultados de la extracción de los cuatro tipos de primitivas. Hemos elegido una escena (fig. 4.17a) que contiene dos naranjas y un fondo de hojas y ramas. Las correspondientes imágenes de distancia y reflectancia se muestran en las figuras 4.17b y 4.17c. En la figura 4.17d vemos las primitivas puntuales contorno que están agrupadas delimitando los bordes de los dos objetos esféricos. En la figura 4.17e vemos las correspondientes primitivas corona que se agrupan formando semiarcos de radio siempre menor que los formados por las primitivas contorno. Las figuras 4.17f y 4.17g muestran las primitivas regionales convexas y reflectividad que han sido derivadas por ser regiones con adecuada convexidad y adecuada reflectividad, respectivamente. En el caso de la escena captada, las imágenes de distancia y reflectancia obtenidas son lo bastante ideales como para que se generen indicios claros en cada una de las primitivas, con lo cual el proceso de detección será redundante, siendo esto importante para dar robustez al sistema.

El proceso de reconocimiento que presentamos es esencialmente una estrategia modular ya que permite la utilización del número de primitivas que se estimen oportunas. Esto quiere decir que incluso utilizando solamente una de ellas, se podría seguir generando indicios suficientes para conseguir la detección de los objetos. Sin embargo, en este caso, en el momento de que por algún motivo no seamos capaces de captar dichas primitivas sobre el objeto, el reconocimiento no será posible al no generarse ningún indicio. Es por ello que se hace necesario la utilización de diferentes primitivas que sean complementarias entre sí, para permitir seguir generando indicios incluso en el caso de que alguna de las otras primitivas no se pongan de manifiesto.

En el caso que queramos reconocer objetos diferentes a una esfera, la estrategia general de

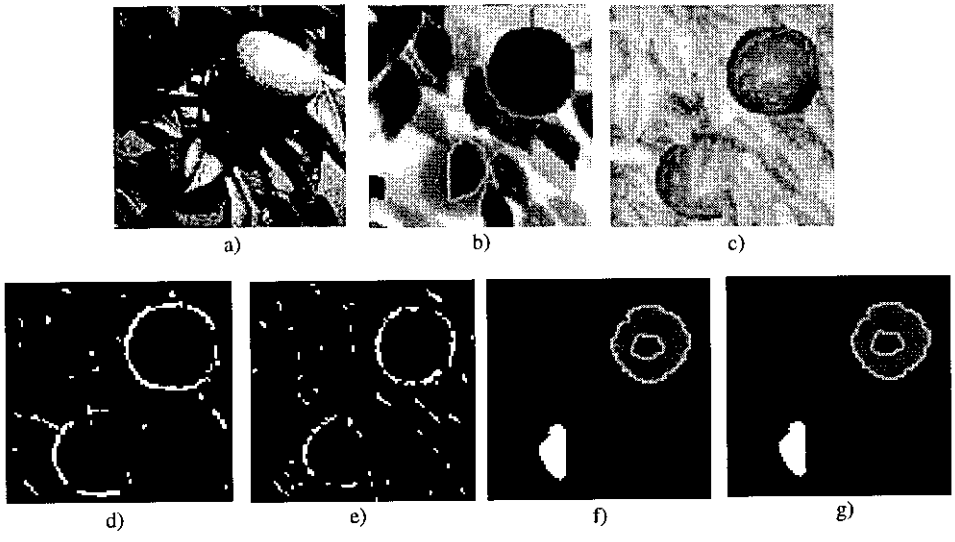


Figura 4.17: Ejemplo de extracción de primitivas: (a) escena fotografiada, (b) mapa de distancias, (c) mapa de reflectancia, (d) primitivas puntuales contorno, (e) primitivas corona, (f) primitivas regionales convexas, y (g) primitivas reflectividad.

reconocimiento propuesta seguiría siendo aplicable. Sin embargo, las primitivas utilizadas no tendrían por qué ser las mismas, tanto en su concepto o semántica como en el aspecto cuantitativo.

En los siguientes apartados pasamos a describir en detalle todos los aspectos contemplados en el planteamiento y diseño de cada una de las cuatro primitivas empleadas.

4.3.1 Primitivas puntuales contorno

El contorno de una esfera es un anillo del mismo radio que esta. En el caso que dicha esfera esté parcialmente oculta, el resultado de la extracción de su contorno es un semiarco circular. La obtención de este tipo de semiarcos es, sin duda, un claro indicio, que convenientemente tratado genera una clara evidencia de presencia de objetos esféricos. Cada uno de los puntos de estos semiarcos, son unidades básicas que generan la información local que permite detectar las esferas que los originan. Estas unidades básicas son las primitivas que pretendemos extraer y que hemos denominado *primitivas puntuales contorno.*

Hay que destacar que estas primitivas no son simplemente los puntos activos tras una etapa de extracción de contornos tradicional (p.ej. Sobel o Canny). Lo que pretendemos realizar al extraer esta primitiva es obtener un subconjunto de los puntos contorno extraídos aplicando los métodos tradicionales. Este subconjunto de puntos incluye aquellos generados por objetos esféricos (o discontinuidades que a uno de sus lados pertenecen a una región convexa), pero excluye el resto de contornos que generan otro

tipo de objetos.

La clara ventaja al trabajar con las *primitivas contorno*, en lugar de imágenes obtenidas tras aplicar una extracción de bordes clásica, radica en que la imagen de *primitivas contorno* es mucho más limpia al no contener los bordes correspondientes a otro tipo de objetos. Por tanto los arcos circulares se pueden apreciar de una forma mucho más clara y consecuentemente su detección se simplifica notablemente.

La ventaja de la extracción de las *primitivas contorno* se pone aún más claramente de manifiesto al considerar las escenas naturales agrícolas conteniendo frutos y hojas como fondo. Bajo estas circunstancias, una técnica de detección de bordes tradicional extrae una gran cantidad de arcos semicirculares, unos que corresponden a los frutos (arcos deseables) y otros generados por las hojas, que también tienen contornos circulares (arcos indeseables). Una posterior etapa de detección de arcos circulares, provocaría la generación de falsas evidencias allí donde se detecten arcos generados por presencia de hojas. Esto significa que bajo estas condiciones se hace absolutamente imprescindible una técnica de extracción de contornos selectiva como la que vamos a presentar ahora.

Para ilustrar el fenómeno de como una técnica clásica de extracción de bordes no discrimina entre los diferentes tipos de contornos, y sin embargo las *primitivas contorno* captan el subconjunto de los bordes que corresponden a "discontinuidades esféricas", obsérvese la figura 4.18 donde se muestra ambos procesamientos aplicados a imágenes sintéticas. En la columna de la izquierda de la figura 4.18, aparece en la parte superior una imagen de distancias donde se ha incluido una esfera y un disco plano. Ambos objetos se encuentran parcialmente ocluidos por una plancha que los cubre, pudiéndose visualizar el 50% de ellos. La esfera viene a representar un fruto parcialmente oculto y el disco plano podría representar una posible hoja de un árbol. Adicionalmente, la imagen sintética contiene una zona escalonada y una marca cuadrada que no es posible apreciar en la imagen de distancias pero sí en la de reflectancias, que se muestra debajo de la correspondiente imagen de distancias. Al aplicar un filtro de Sobel sobre cada una de las imágenes, obtenemos los bordes que se muestran en la columna central. Es posible apreciar que no hay ninguna discriminación en su extracción y los semiarcos correspondientes a los bordes de la esfera y del disco generan indicios de circularidad del mismo orden. Sin embargo, aplicando la técnica de extracción de *primitivas contorno* que describiremos en breve, obtenemos una discriminación entre los diferentes tipos de bordes existentes y por tanto sólo mantenemos los puntos del contorno que provienen de "discontinuidades esféricas" (columna derecha).

Para obtener nuestro objetivo debemos realizar un sistema de clasificación de bordes. En la literatura se pueden encontrar diversos trabajos que contemplan la clasificación de bordes en clases como: salto, cresta y rampa. Los primeros tipos pueden ser detectados aplicando operadores generales de detección de bordes o métodos basados en los residuos entre la imagen original y su versión suavizada [2] ó ajustada mediante un polinomio [118]. Otros clasificadores utilizan operadores morfológicos aunque solamente sirven para detectar una, o a lo sumo, dos clases de bordes. Las técnicas basadas en la covarianza permiten obtener descriptores de forma invariantes a movimientos y permiten detectar bordes salto y cresta [12, 81], al igual que utilizando técnicas basadas en momentos de Zernike [63]. Aplicando la técnica conocida por *aproximación de línea de barrido* (Scan Line Aproximation) se ha podido clasificar entre saltos y bordes cresta [104] aunque los resultados son bastante sensibles al ruido.

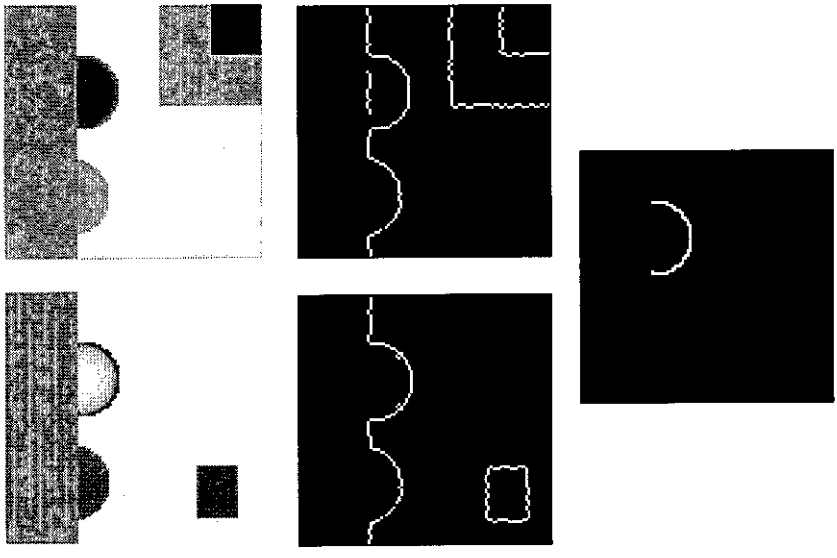


Figura 4.18: Extracción de contornos sobre imágenes sintéticas de distancia y reflectancia (columna de la izquierda), utilizando un filtro de Sobel (columna central) y generando primitivas puntuales contorno (columna derecha)

En trabajos de Nadabar [150] y Zhang-Wallace [220] se presenta una clasificación de bordes utilizando la integración de imágenes en distancia e intensidad. Este último plantea clasificar siete tipos de bordes incluyendo *bordes esféricos* (equivalentes a nuestras primitivas contorno), pero finalmente solo se clasifican: saltos, crestas, marcas y no bordes. Se renuncia a clasificar los bordes esféricos (que ellos llaman *Extremal*) por ser muy similares a los bordes salto. El procedimiento utilizado consta de dos etapas:

1. Clasificación bayesiana utilizando distancia, orientación e intensidad como características.

2. Refinamiento de la clasificación mediante un proceso iterativo de relajación que de forma concurrente realiza una restauración de las imágenes.

Otros trabajos que utilizan imágenes de distancia para clasificar bordes son los de Wani [212] y Mintz [141]; sin embargo en ambos casos la clasificación se reduce a dos tipos de bordes, saltos y crestas. Lo mismo ocurre con el trabajo de Gil [64] donde se combinan los bordes del mapa de distancia e intensidades utilizando operaciones "Y" lógicas locales y globales.

Como vemos en ninguno de los trabajos revisados se clasifican bordes de tipo esférico, que es la clase en la cual nosotros estamos especialmente interesados, por ello plantearemos nuestro propio método de clasificación.

El sistema de clasificación de puntos que proponemos se basa, al igual que en los trabajos de Zhang-Wallace [220], Nadabar [150] y Gil [64], en la integración de la información

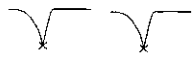
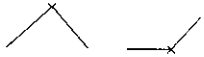
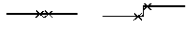
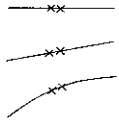
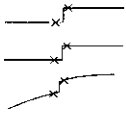
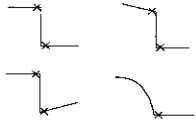
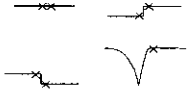
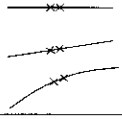
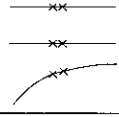
Tipo Punto	Distancia	Reflectancia	$\|K_n\|>u_K_n$	$mD>u_mD$	$dm\Re>u_dm\Re$	$m\Re>u_m\Re$
Borde Esférico (primitiva contorno)			--	SI	SI	--
Borde Cresta			SI	NO	--	--
Borde Marca			NO	NO	--	SI
Otros bordes			--	SI	NO	--
Punto Suave			NO	NO	--	NO

Figura 4.19: Clasificación geométrica de diferentes tipos de puntos utilizando información de distancia y reflectancia.

presente tanto en las imágenes de distancia como en las de reflectancia. Mediante esta estrategia vamos a poder distinguir entre cinco clases de puntos: *bordes esféricos*, *bordes cresta*, *bordes marca*, *otros bordes* y *puntos suaves*. En la figura 4.19 se puede ver de una forma gráfica qué es lo que entendemos por cada una de estas denominaciones.

Basándonos en las descripciones geométricas de estos cinco tipos de puntos podemos establecer de forma inicial las propiedades o características que nos permitirán clasificarlos. De forma general, los *bordes esféricos* se caracterizan por presentar un salto en el mapa de distancias y al mismo tiempo generar una depresión o concavidad abrupta en el mapa de reflectancias. Este último fenómeno se debe a la baja señal de reflexión captada que se produce al incidir sobre superficies muy oblicuas, como indica la ley de Lambert, siendo esta una característica altamente discriminante. Los *bordes cresta* se caracterizan por presentar una alta curvatura tanto positiva como negativa en el mapa de distancias, siendo totalmente irrelevante la información de reflectancia. Los *bordes marca* son cambios en la tonalidad o reflectividad de las superficies, por tanto aparecen cuando hay cambios en el mapa de reflectancias habiendo continuidad en la imagen de distancias. En la categoría de *otros bordes* incluimos el resto de posibles tipos de bordes presentes en una escena. Son por tanto los puntos que presentan discontinuidad

en el mapa de distancias pero que no encajan en ninguna de las categorías anteriores. Finalmente, los *puntos suaves* corresponden a regiones con continuidad o suavidad en la imagen de distancias y que además no presentan marcas, no apareciendo por tanto discontinuidades en el mapa de reflectancia.

Analizando las propiedades que acabamos de utilizar para describir los diferentes tipos de puntos, hemos propuesto un conjunto de cuatro parámetros que nos va a permitir discriminar entre ellos, y por tanto, clasificar cada punto de la imagen de una forma muy directa al aplicar funciones discriminantes lineales y cuadráticas sobre un espacio de características de cuatro dimensiones. Este conjunto de parámetros contiene la pendiente en el mapa de distancias mD, la curvatura media en distancias K_m, la pendiente en reflectancias $m\Re$ y el diferencial de pendientes en reflectancias $dm\Re$.

4.3.1.1 Características

El cálculo de la pendiente en el mapa de distancias, mD, y en el de reflectancias, $m\Re$, lo realizamos utilizando una ventana de tamaño 3×3. Realizamos primeramente una descomposición en las direcciones horizontal y vertical para calcular las pendientes direccionales en D: mD_x, mD_y, y en $\Re$: $m\Re_x$, $m\Re_y$. Posteriormente, la contribución de ambas direcciones es tenida en cuenta para obtener el valor final de las pendientes de la siguiente forma:

$$\begin{aligned} mD &= \sqrt{(mD_x)^2 + (mD_y)^2} \\ m\Re &= \sqrt{(m\Re_x)^2 + (m\Re_y)^2} \end{aligned} \tag{4.36}$$

donde

$$\begin{aligned} mD_x &= \frac{2D_{i,j+1} + D_{i-1,j+1} + D_{i+1,j+1} - (2D_{i,j-1} + D_{i-1,j-1} + D_{i+1,j-1})}{8 \cdot s} \\ mD_y &= \frac{2D_{i+1,j} + D_{i+1,j-1} + D_{i+1,j+1} - (2D_{i-1,j} + D_{i-1,j-1} + D_{i-1,j+1})}{8 \cdot s} \\ m\Re_x &= \frac{2\Re_{i,j+1} + \Re_{i-1,j+1} + \Re_{i+1,j+1} - (2\Re_{i,j-1} + \Re_{i-1,j-1} + \Re_{i+1,j-1})}{8 \cdot s} \\ m\Re_y &= \frac{2\Re_{i+1,j} + \Re_{i+1,j-1} + \Re_{i+1,j+1} - (2\Re_{i-1,j} + \Re_{i-1,j-1} + \Re_{i-1,j+1})}{8 \cdot s} \end{aligned} \tag{4.37}$$

Hay que resaltar que en el cómputo de estas pendientes se considera la resolución espacial s de la imagen capturada, indicando los milímetros de separación entre cada pareja de puntos contiguos. Este factor, por tanto, contribuye a que los parámetros aquí definidos sean invariantes a posibles cambios en la resolución de la imagen, y a hacer que la pendiente calculada sea adimensional.

Para ver las variaciones locales en el mapa de reflectancia $dm\Re$ nos apoyamos en la información que nos da el mapa de distancia para detectar las direcciones de los gradientes. De esta forma, al detectar primeramente la dirección de la transición más representativa, el cálculo de la diferencia de pendientes en el mapa de reflectancia se reduce al cálculo de dos gradientes a lo largo de la línea orientada según la dirección del gradiente. A la hora de calcularlo, el rango angular continuo de 360 grados ha

sido discretizado en un conjunto de 12 secciones angulares, barriendo los 360 grados a intervalos de 30. En base a estas secciones hemos definido dos grupos de gradientes, uno de ellos abarcando líneas de tres puntos en el sentido indicado por la sección angular correspondiente, $grad^{l3}_{seccion}$, y el otro empleando solo dos puntos en el sentido opuesto, $grad^{l2}_{seccion}$. Estas son sus definiciones:

$$grad^{l3}_{seccion}(i,j) = \begin{cases} grad^{l3}_{0}(i,j) = & (\Re_{i,j+3} - \Re_{i,j})/(3 * s_{i,j}) \\ grad^{l3}_{30}(i,j) = & (\Re_{i-2,j+3} - \Re_{i,j})/(3 * s_{i,j}) \\ grad^{l3}_{60}(i,j) = & (\Re_{i-3,j+2} - \Re_{i,j})/(3 * s_{i,j}) \\ grad^{l3}_{90}(i,j) = & (\Re_{i-3,j} - \Re_{i,j})/(3 * s_{i,j}) \\ grad^{l3}_{120}(i,j) = & (\Re_{i-3,j-2} - \Re_{i,j})/(3 * s_{i,j}) \\ grad^{l3}_{150}(i,j) = & (\Re_{i-2,j-3} - \Re_{i,j})/(3 * s_{i,j}) \\ grad^{l3}_{180}(i,j) = & (\Re_{i,j-3} - \Re_{i,j})/(3 * s_{i,j}) \\ grad^{l3}_{210}(i,j) = & (\Re_{i+2,j-3} - \Re_{i,j})/(3 * s_{i,j}) \\ grad^{l3}_{240}(i,j) = & (\Re_{i+3,j-2} - \Re_{i,j})/(3 * s_{i,j}) \\ grad^{l3}_{270}(i,j) = & (\Re_{i+3,j} - \Re_{i,j})/(3 * s_{i,j}) \\ grad^{l3}_{300}(i,j) = & (\Re_{i+3,j+2} - \Re_{i,j})/(3 * s_{i,j}) \\ grad^{l3}_{330}(i,j) = & (\Re_{i+2,j+3} - \Re_{i,j})/(3 * s_{i,j}) \end{cases} \tag{4.38}$$

$$grad^{l2}_{seccion}(i,j) = \begin{cases} grad^{l2}_{0}(i,j) = & (\Re_{i,j} - \Re_{i,j-2})/(2 * s_{i,j}) \\ grad^{l2}_{30}(i,j) = & (\Re_{i,j} - \Re_{i+1,j-2})/(2 * s_{i,j}) \\ grad^{l2}_{60}(i,j) = & (\Re_{i,j} - \Re_{i+2,j-1})/(2 * s_{i,j}) \\ grad^{l2}_{90}(i,j) = & (\Re_{i,j} - \Re_{i+2,j})/(2 * s_{i,j}) \\ grad^{l2}_{120}(i,j) = & (\Re_{i,j} - \Re_{i+2,j+1})/(2 * s_{i,j}) \\ grad^{l2}_{150}(i,j) = & (\Re_{i,j} - \Re_{i+1,j+2})/(2 * s_{i,j}) \\ grad^{l2}_{180}(i,j) = & (\Re_{i,j} - \Re_{i,j+2})/(2 * s_{i,j}) \\ grad^{l2}_{210}(i,j) = & (\Re_{i,j} - \Re_{i-1,j+2})/(2 * s_{i,j}) \\ grad^{l2}_{240}(i,j) = & (\Re_{i,j} - \Re_{i-2,j+1})/(2 * s_{i,j}) \\ grad^{l2}_{270}(i,j) = & (\Re_{i,j} - \Re_{i-2,j})/(2 * s_{i,j}) \\ grad^{l2}_{300}(i,j) = & (\Re_{i,j} - \Re_{i-2,j-1})/(2 * s_{i,j}) \\ grad^{l2}_{330}(i,j) = & (\Re_{i,j} - \Re_{i-1,j-2})/(2 * s_{i,j}) \end{cases} \tag{4.39}$$

De acuerdo a la sección angular activa, la cual se seleccionó al discretizar la dirección del gradiente en el mapa de distancias, solamente la pareja ($grad^{l3}_{seccion}$,$grad^{l2}_{seccion}$) será contemplada a la hora de efectuar la clasificación. En concreto, la diferencia entre ambas componentes será la característica de clasificación (ec. 4.40), aunque sujeta a unas ligaduras locales como veremos más adelante en el apartado de funciones discriminantes.

$$dm\Re = grad^{l3}_{seccion} - grad^{l2}_{seccion} \tag{4.40}$$

La cuarta de las características utilizadas es la curvatura media K_m. Existen definiciones bastantes sofisticadas de lo que se entiende por curvatura [15], sin embargo, una buena definición de este concepto se puede obtener utilizando el concepto de curvatura $K(p,q)$ para una pareja (p,q) de puntos [82]. Dados los vectores unitarios, $\vec{n_p}$ y $\vec{n_q}$, normales a una superficie en los puntos p y q, definimos la curvatura $K(p,q)$ como la razón entre las distancias de sus cabezas y sus orígenes. La siguiente ecuación expresa este concepto de una manera formal:

$$K(p,q) = \frac{\|\vec{n_p} - \vec{n_q}\|}{\|\vec{p} - \vec{q}\|} \cdot s(p,q) \tag{4.41}$$

donde $s(p,q)$ es un término que vale 1 o -1 dependiendo de si la superficie es convexa o cóncava, respectivamente.

$$s(p,q) = \begin{cases} 1 & \frac{\|(\vec{p}+\vec{n_p})-(\vec{q}+\vec{n_q})\|}{\|\vec{p}-\vec{q}\|} > 1 \\ -1 & \text{caso contrario} \end{cases} \tag{4.42}$$

Finalmente, el parámetro utilizado para la clasificación es la curvatura media $K_m(p)$ en un punto p, que se obtiene al promediar las curvaturas entre parejas de puntos del vecindario que contienen al punto p. Es decir:

$$K_m(p) = \frac{\sum_{i=1}^{n} K(p, q(i))}{n} \tag{4.43}$$

4.3.1.2 Funciones discriminantes

Las cuatro características que acabamos de definir forman un espacio de cuatro dimensiones que nos permiten realizar una clasificación atendiendo a una serie de condiciones. Las condiciones que deben satisfacer cada uno de los puntos para ser asignados a la correspondiente clase se esquematiza en la tabla condicional de la figura 4.19. Esta tabla condicional, obtenida a partir de las definiciones gráficas de cada tipo de punto, nos permite realizar una clasificación sin ninguna ambigüedad, como puede apreciarse al no existir ninguna configuración idéntica. El único aspecto que queda por considerar es la elección de las funciones de decisión o discriminantes. Dichas funciones van a consistir en hiperplanos o hipersuperficies cuadráticas, como veremos a continuación, obteniéndose sus parámetros mediante una serie de umbrales deducidos de forma automática atendiendo a propiedades geométricas. Este aspecto es importante pues significa que no es necesario ajustar los umbrales experimentalmente, con la consiguiente ganancia en flexibilidad.

Condición y umbral discriminante para mD. El umbral u_mD correspondiente a la característica mD, lo definimos a partir del modelo geométrico de una esfera de radio R. Consideramos que puntos situados a una distancia del extremo ecuatorial de una esfera menor o igual que el paso entre puntos, s, tienen una pendiente mD considerablemente elevada y por tanto son candidatos a bordes con discontinuidad (fig. 4.20a). El umbral u_mD se fija de esta forma:

Si una circunferencia de radio R obedece al modelo $y = \sqrt{R^2 - x^2}$, tenemos que la pendiente en cada punto de una esfera viene dada por:

$$mD|_{esfera} = dy/dx|_{esfera} = x/\sqrt{R^2 - x^2} \tag{4.44}$$

Vamos a definir h como la fracción entre una distancia x y el radio R de la esfera.

$$h = x/R \tag{4.45}$$

siendo x la distancia entre el centro de la esfera y la proyección sobre el eje de abcisas del punto bajo estudio situado sobre la esfera.

De esta forma podemos definir $mD|_{esfera}$ en términos de h de la siguiente manera

$$mD|_{esfera}(h) = h/\sqrt{1-h^2} \tag{4.46}$$

Así, el valor de h que corresponde a un punto con proyección x que difiere de R en una distancia igual al paso entre puntos viene dado por

$$h_{u_mD} = \frac{x}{R} = \frac{R - 1 \cdot s}{R} \tag{4.47}$$

y por tanto el umbral correspondiente u_mD, que se utilizará para realizar la clasificación de puntos, viene dado por

$$u_mD = mD|_{esfera}(h_{u_mD}) = \frac{h_{u_mD}}{\sqrt{1-h_{u_mD}^2}} \tag{4.48}$$

Condición y umbral discriminante para $dm\Re$. De igual forma, la característica $dm\Re$ nos permitirá discriminar una vez que deduzcamos su correspondiente umbral $u_dm\Re$. Sin embargo, en este caso no solamente se contemplará la influencia aislada de $u_dm\Re$ sino que además serán considerados sus componentes $grad^{3l}$ y $grad^{2l}$, para asegurar que el incremento total es el resultado de la contribución de dos pendientes apreciables y con signos diferentes. Esto se hace así para asegurar la detección fiable de las concavidades abruptas en el mapa de reflectancia, evitando elevados valores de $dm\Re$ provocados por simples transiciones entre superficies con diferentes reflectancias, las cuales son de tipo escalón.

Así, la condición discriminante a satisfacer, relacionada con los incrementos de las pendientes en el mapa de reflectancia, se traduce en

$$(dm\Re \geq u_dm\Re)\ Y\ (grad^{l3} > u_dm\Re/4)\ Y\ (-grad^{l2} > u_dm\Re/4) \tag{4.49}$$

En relación al cálculo del umbral $u_dm\Re$, si consideramos el modelo de una esfera, podemos deducir que el ángulo θ que formaría un supuesto haz láser con respecto a la normal de la superficie esférica sobre la que incide, viene dado por

$$\theta = \arctan(dy/dx) = \arctan(\frac{-x}{\sqrt{R^2-x^2}}) \tag{4.50}$$

De esta forma, utilizando el modelo ideal de interacción láser deducido en el capitulo 2 y que viene dado por la ecuación 2.17, podemos expresar la reflectancia $\Re$ en función de x y R.

$$\begin{aligned} \Re &= 80.9 \log(cte1 \cdot cos\theta) = \\ &= cte2 - 80.9 \log\left[\cos\left(\arctan(x/\sqrt{R^2-x^2})\right)\right] \end{aligned} \tag{4.51}$$

A partir de la expresión anterior, derivándola, podemos obtener la pendiente $m\Re$ en el mapa de reflectancia que le corresponde a un punto x de una esfera de radio R.

$$m\Re|_{esfera}(x) = \frac{d\Re}{dx}|_{esfera} =$$

$$\frac{80.9}{ln(10)\cdot\cos\left(\tan^{-1}(x/\sqrt{R^2-x^2})\right)}\frac{-0.5\left(\frac{2x}{R^2-x^2}+\frac{2x^3}{(R^2-x^2)^2}\right)}{\left(1+\frac{x^2}{R^2-x^2}\right)^{1.5}} = \ldots$$

$$\frac{80.9}{ln(10)}\frac{x}{x^2-R^2} \tag{4.52}$$

Y expresándolo de forma normalizada utilizando el término $h = x/R$, tenemos

$$m\Re|_{esfera}(h) = \frac{-80.9}{ln(10)}\frac{h}{R\cdot(1-h^2)} \tag{4.53}$$

En la figura 4.20b podemos ver en el mapa de reflectancia, la concavidad abrupta correspondiente al contorno de una esfera. A ambos lados de este valle existen dos laderas con pendientes pronunciadas. Denominaremos ladera interna aquella que está próxima al centro de la esfera, y ladera externa a la que está más alejada. Si calculamos la pendiente de la ladera interna utilizando máscaras 3 × 3, tenemos que la mejor estimación de esa pendiente se obtiene dos puntos dentro del radio de la esfera. Por tanto el valor de h correspondiente al umbral límite viene dado por:

$$h_{u_dm\Re} = \frac{x}{R} = \frac{R-2\cdot s}{R} \tag{4.54}$$

La ladera externa de la concavidad también tiene una pendiente considerable aunque de signo contrario. En términos absolutos dicha pendiente externa debe ser muy semejante a la pendiente de la cara interna puesto que debe haber una recuperación de reflectancia rápida asumiendo una superficie colindante con reflectividad típica. Por tanto, el umbral utilizado debe ser aproximadamente el doble de la pendiente en el punto de la esfera considerado, aunque si aplicamos un margen de incertidumbre del 0.75 tenemos la siguiente frontera de clasificación:

$$u_dm\Re = 2\cdot 0.75\cdot m\Re|_{esfera}(h_{u_dm\Re}) = 1.5\frac{80.9}{ln(10)}\frac{h_{u_dm\Re}}{R(1-h^2_{u_dm\Re})} \tag{4.55}$$

Condición y umbral discriminante para $m\Re$. Como podemos apreciar en la tabla condicional de la figura 4.19, la característica $m\Re$ es útil para distinguir entre los *puntos suaves* y los *bordes marca*. Estos últimos se caracterizan por presentar una discontinuidad en el mapa de reflectancia, mientras que los *puntos suaves* pertenecen a regiones continuas tanto en distancia como en reflectancia. De una forma arbitraria vamos a definir como *bordes marcas* a aquellos puntos que teniendo continuidad en distancias tienen una discontinuidad en reflectancia mayor o igual a 2 en un intervalo de 1 milímetro.

La anterior función discriminante permite clasificar correctamente las marcas cuando entran en juego superficies planas. Sin embargo, no es adecuada para detectar marcas

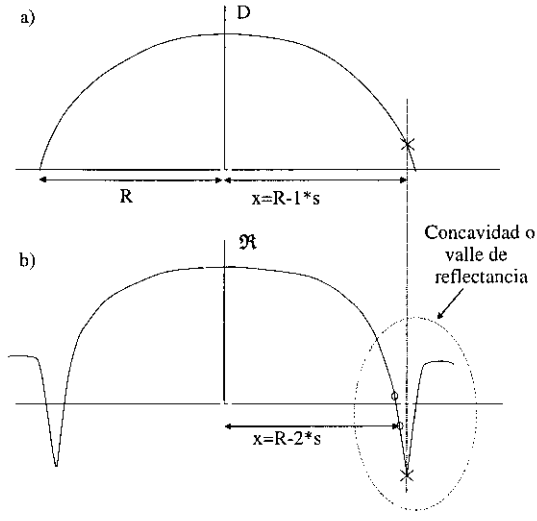


Figura 4.20: Puntos críticos sobre una esfera para el cálculo de los umbrales u_mD y $u_dm\Re$: a) perfil en distancia, b) perfil de reflectancia.

en superficies con curvatura, puesto que en estos casos existe un gradiente en el mapa de reflectancia. Sin embargo este gradiente no es debido a cambios de reflectividad de las superficies sino a la variación gradual en reflectancia correspondiente a la ley del coseno de Lambert. Por tanto, sobre regiones curvas es necesario añadir un término que evite clasificar estos puntos curvos como marcas. Dicho término es una función que varía entre 0 e infinito y que permite elevar el umbral a medida que aumenta el gradiente en distancias. El incremento aplicado corresponde al gradiente de reflectancia esperado suponiendo que estuviésemos sobre un punto de una esfera para una determinada pendiente en distancias.

$$u_m\Re = \quad 2 + m\Re|_{esfera} = \quad 2 + \frac{-80.9}{ln(10)} \frac{h}{R(1-h^2)} = \tag{4.56}$$

$$\left(como\ h = \frac{mD}{\sqrt{1+mD^2}}\right) = \quad 2 + \frac{80.9}{ln(10)} \frac{mD}{R} \sqrt{1+mD^2}$$

Condición y umbral discriminante para k_m. La curvatura media K_m es un parámetro que se utiliza fundamentalmente para discriminar *bordes cresta* frente a otro tipo de puntos sobre superficies continuas. Los *bordes cresta* se caracterizan por no presentar una discontinuidad o salto en distancias, pero sin embargo manifiestan un punto de inflexión. Esta inflexión puede ser detectada como un cambio en las normales a la superficie, y por tanto, como un pico en un mapa de curvaturas. De esta forma evitamos contemplar estos puntos como pertenecientes a regiones continuas.

El umbral que permite decidir si una curvatura es suficientemente grande como para clasificar un punto como *borde cresta*, lo podemos definir teniendo en cuenta que la curvatura correspondiente a la esfera media que se pretende detectar, esté suficientemente alejada del umbral a elegir. En principio, para nuestro caso, denominamos borde cresta

a aquellos puntos que tengan una curvatura, tanto cóncava como convexa, superior a 3 veces la curvatura de la esfera media esperada que es $1/R$.

$$u_K_m = 3/R \tag{4.57}$$

$$|K_m| > u_K_m \tag{4.58}$$

4.3.1.3 Clasificación de los puntos imagen

Una vez definidos los tipos de puntos a clasificar, las características discriminantes y las funciones de decisión, estamos en condiciones de aplicar el método de clasificación de puntos a imágenes sintéticas y reales. Hay que destacar que este método de clasificación realiza una integración de imágenes de reflectancia y distancia, y por tanto, el método está restringido a aplicaciones donde se dispone de ambos tipos de información. La selección de los umbrales de las fronteras de decisión depende del tipo de sensor láser utilizado, sin embargo, en el caso de utilizar otro sensor basta obtener su modelo y a partir de él se pueden derivar los nuevos umbrales siguiendo un procedimiento idéntico al presentado en la subsección anterior.

En la figura 4.21 se puede ver el resultado de clasificar la pareja de imágenes sintéticas de la columna de la izquierda. En la fig.4.21c vemos los puntos clasificados como *bordes esféricos*, los cuales se corresponden con las *primitivas puntuales contorno* que utilizaremos para generar indicios de esfericidad. En la figuras 4.21d,e,f se muestran los *bordes marca*, *otros bordes* y los *puntos suaves*, respectivamente. En este caso no existe ningún borde de tipo *cresta*. Para tener una idea de la distribución de los vectores de características asociados a cada pixel de la imagen, presentamos una muestra aleatoria de 30 puntos de cada clase al proyectarlos sobre dos planos de características: $mD - dm\Re$ (fig.4.22) y $mD - m\Re$ (fig.4.23). Podemos apreciar que las agrupaciones no son siempre unimodales y que tampoco siguen distribuciones gausianas. Se podrían haber aplicado otras técnicas de clasificación, pero en este caso la detallada descripción geométrica de cada tipo de pixel, permite obtener de una forma directa unos umbrales que son discriminantes, como se mostró en la sección anterior. Además la carga computacional requerida para asignar una etiqueta a cada punto, una vez extraídas sus características, es muy baja.

En la figura 4.24a,b,c podemos ver una escena de un árbol artificial con cuatro frutos y sus correspondientes imágenes en distancia y en reflectancia, respectivamente. Si aplicamos la clasificación de puntos a la pareja de imágenes distancia-reflectancia, obtenemos los resultados mostrados en la figura 4.24d,e,f,g,h, donde se muestran los *bordes esféricos*, *cresta*, *marcas*, *otros bordes* y *puntos suaves*, respectivamente. Podemos apreciar que de estas cinco imágenes las dos más representativas a la hora de suministrar indicios de esfericidad, son las correspondientes a *bordes esféricos* (cosa que era de esperar al ser nuestras primitivas puntuales contorno), y a *puntos suaves* puesto que captan las regiones continuas que existen sobre la superficie de las frutas. De hecho y como veremos más adelante, este último tipo será utilizado para generar las *primitivas regionales convexas* y *reflectividad*.

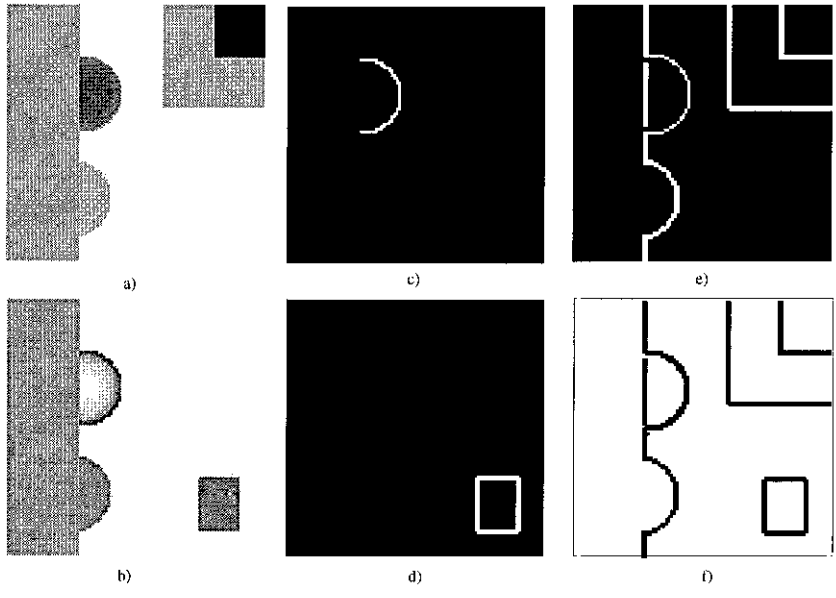


Figura 4.21: Clasificación de puntos mediante el método propuesto: (a) y (b) imágenes sintéticas en distancia y reflectancia, (c) bordes esféricos (primitivas puntuales contorno), (d) bordes marca, (e) otros bordes y (f) puntos suaves.

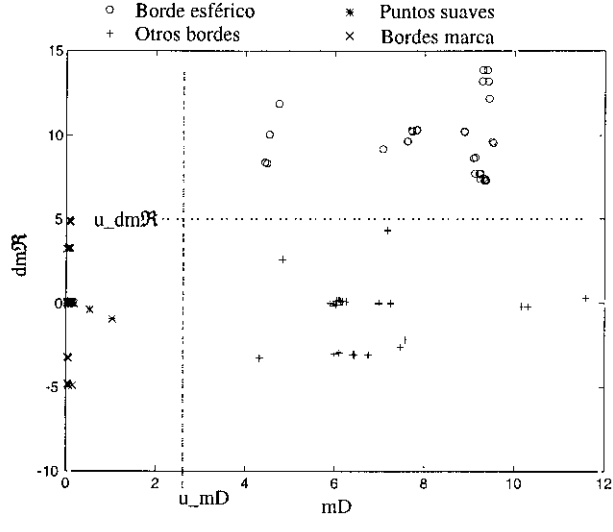


Figura 4.22: Proyección sobre el plano $mD - dm\Re$ de una muestra de puntos correspondiente a la clasificación de la figura 4.21. Obsérvese como los umbrales definidos permiten discriminar los bordes esféricos (primitivas contorno) del resto de bordes.

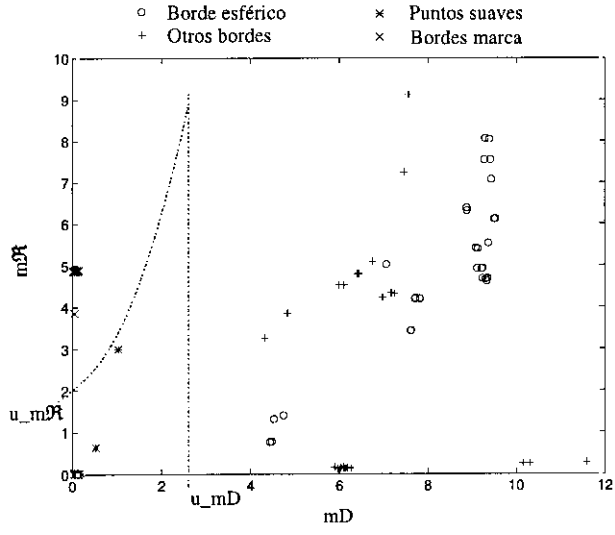


Figura 4.23: Proyección sobre el plano $mD - m\Re$ de una muestra de puntos correspondiente a la clasificación de la figura 4.21. Podemos observar como los umbrales deducidos permiten discriminar entre los puntos suaves y los puntos marca.

4.3.2 Primitivas puntuales corona

Las primitivas puntuales *corona*, al igual que las primitivas *contorno* presentadas anteriormente, son otra forma de extraer indicios de esfericidad mediante la obtención de arcos circulares. Con el hecho de utilizar más primitivas para detectar de nuevo arcos esféricos se pretende complementar los indicios obtenidos anteriormente mediante las primitivas *contorno*, es decir, estamos utilizando una técnica de detección basada en otros principios de tal forma que las diversas circunstancias que puedan alejar nuestras imágenes de los modelos ideales planteados afecten lo menos posible al proceso de reconocimiento.

Un fenómeno no contemplado hasta ahora y que puede afectar a la detección de las primitivas *contorno* se puede dar si los objetos esféricos no son buenos difusores de la luz, y cerca de sus contornos existen superficies orientadas de tal forma que cuando el haz láser incide sobre el borde de una esfera estas superficies devuelven al sensor la energía reflejada especularmente sobre ellas (fig. 4.25). Este fenómeno hace que la amplitud de la señal en los contornos de algunas esferas no sea tan débil como el modelo de reflectancia indica. Esta circunstancia es un ejemplo de fenómenos que se pueden dar y que afectan a la fiabilidad en la detección de primitivas, es por ello necesario complementar la extracción de primitivas mediante otras estrategias que permitan generar más indicios de esfericidad para obtener detecciones suficientemente robustas.

Una forma de obtener arcos circulares centrados sobre objetos esféricos se consigue al marcar aquellos puntos que tienen igual pendiente en un mapa de distancias. El resultado obtenido es una corona circular con un radio que es función de la pendiente seleccionada.

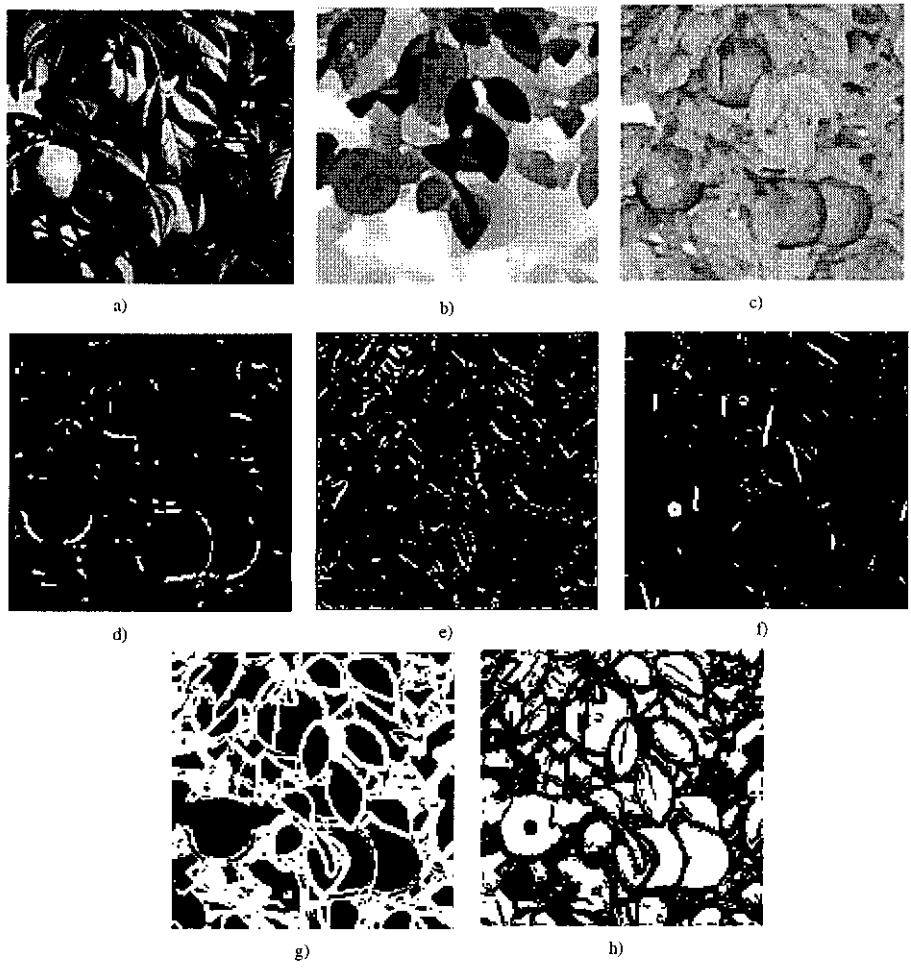


Figura 4.24: Clasificación de puntos mediante el método propuesto correspondiente a una imagen de un árbol frutal artificial: (a) fotografía de la escena en B/N, (b) y (c) imágenes en distancia y reflectancia, (d) bordes esféricos (primitivas puntuales contorno), (e) bordes cresta, (f) bordes marca, (g) otros bordes y (h) puntos suaves.

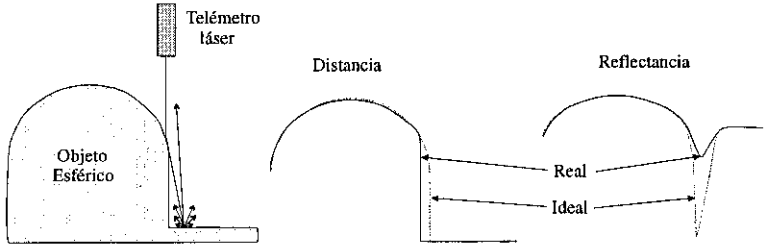


Figura 4.25: Situación anómala que provoca un alejamiento entre el modelo ideal y los datos reales en la imagen de reflectancia y distancia.

Esta corona circular, que podría no ser cerrada si existen problemas de oclusión sobre las esferas, permite generar indicios de esfericidad y por tanto se ha elegido como fundamento para la obtención de las primitivas denominadas *primitivas puntuales corona.*

Para definir las primitivas corona primeramente necesitamos elegir la pendiente correspondiente a la corona con mejor carácter discriminante. La corona más discriminante será aquella con un alto número de puntos y que a la vez contenga un mínimo número de puntos que correspondan a objetos no esféricos. Pero, ¿cómo determinar la pendiente óptima para obtener las coronas más discriminantes?. En principio las siguientes consideraciones se pueden aplicar:

- *No pendientes pequeñas.* Las pendientes pequeñas generan coronas de radios reducidos con lo cual al tener perímetros pequeños se generan pocos indicios de esfericidad. Por tanto son preferibles coronas con un radio lo mayor posible.
- *No pendientes máximas.* Por el contrario, las coronas de máximo radio que son aquellas que se corresponden con altas pendientes y por tanto captan bordes, no son deseables precisamente por captar todos los bordes de forma indiscriminada, tal y como lo hace un filtro de Sobel o Canny. La detección de forma discriminante de los bordes que corresponden a esferas se trató en el apartado anterior al definir las primitivas contorno.
- *Una pendiente intermedia.* Por tanto la solución ideal se encuentra en una posición intermedia entre las dos soluciones anteriores.

Para determinar esta pendiente óptima podemos aplicar una técnica de decisión estadística y elegir aquella pendiente que presente un máximo en la probabilidad de pertenecer a una esfera. Por definición la función de densidad de probabilidad se corresponde con el cociente entre el diferencial de la función de distribución de probabilidad dF y el diferencial de la variable aleatoria considerada. Por tanto podemos plantear la función de densidad de probabilidad condicional $f(\theta|esfera)$ tal y como la ecuación 4.59 indica. Suponiendo el modelo de una esfera y un muestreo espacial uniforme sobre un plano, podemos desarrollar dF a partir de $F(x)$ y $F(x+dx)$, donde las distribuciones de probabilidad $F(x)$ y $F(x+dx)$ se obtienen al dividir el área de un círculo de radio x y $x+dx$, respectivamente, entre el área de uno de radio R (fig. 4.26 y ec. 4.60). De esta forma $f(\theta|esfera)$ queda definido por la ecuación 4.61.

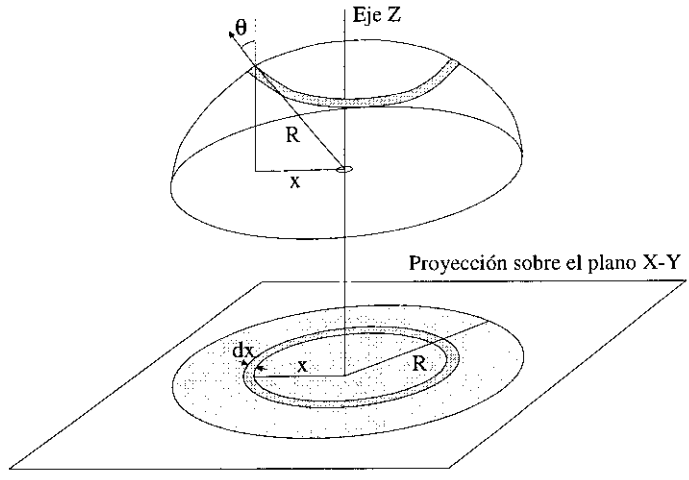


Figura 4.26: Esquema que muestra una *corona* sobre una esfera correspondiente a los puntos con una pendiente tal que su normal forma un ángulo θ con el eje Z.

$$f(\theta|esfera) = \frac{dF}{d\theta} \tag{4.59}$$

$$\begin{aligned} dF &= F(x) - F(x+dx) = \frac{\pi x^2}{\pi R^2} - \frac{\pi(x+dx)^2}{\pi R^2} \simeq \\ &\frac{2\pi x dx}{\pi R^2} = (x = R\sin\theta; dx = R\cos\theta d\theta) \\ &\frac{2\pi R \sin\theta R \cos\theta d\theta}{\pi R^2} = 2\sin\theta\cos\theta d\theta \end{aligned} \tag{4.60}$$

$$f(\theta|esfera) = 2\sin\theta\cos\theta \tag{4.61}$$

Como puede verse en la figura 4.27a, la función densidad de probabilidad $f(\theta|esfera)$ presenta un máximo en 0.78 radianes y se hace cero en 0 y $\pi/2$ radianes. Esto concuerda con lo esperado ya que la probabilidad debe aumentar de una forma monótona al hacerlo el radio puesto que el perímetro de las coronas aumenta, sin embargo y debido a los efectos del muestreo realizado en ejes paralelos al eje z, la zona visible para valores próximos a $\pi/2$ se reduce según el coseno de θ con lo cual la probabilidad de captar puntos con pendiente $\pi/2$ se hace cero.

Supongamos que los objetos esféricos de interés se encuentran sobre un fondo de objetos planos orientados aleatoriamente, si esta orientación es totalmente aleatoria se puede demostrar que las normales a los planos presentan una distribución de probabilidad idéntica al caso de una esfera (fig. 4.27b). Aplicando el teorema de Bayes (ec. 4.62)

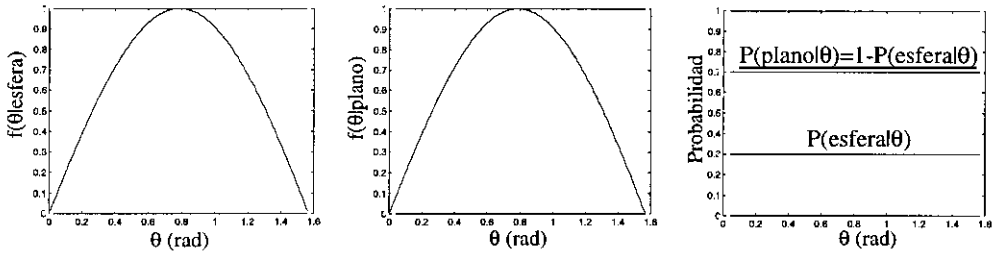


Figura 4.27: Función densidad de probabilidad condicional para: (a) una esfera y (b) un plano. (c) Probabilidad de que un punto con normal θ pertenezca a un plano o una esfera.

podemos ver que la probabilidad de pertenecer a una esfera, de un punto cuya normal forma un ángulo θ con el eje Z, es constante para cualquier valor de θ (fig. 4.27c). Es por ello que no existe ningún valor concreto de θ que maximice dicha probabilidad y por tanto no hay una pendiente óptima en el mapa de distancias que nos permita seleccionar la corona más discriminante.

$$\begin{aligned} P(esfera|\theta) &= \frac{f(\theta|esfera)P(esfera)}{f(\theta)} = \\ &= \frac{f(\theta|esfera)P(esfera)}{f(\theta|esfera)P(esfera) + f(\theta|plano)P(plano)} = \\ &= \frac{P(esfera)}{P(esfera) + P(plano)} \end{aligned} \tag{4.62}$$

$$\begin{aligned} P(plano|\theta) &= \frac{f(\theta|plano)P(esfera)}{f(\theta|esfera)P(esfera) + f(\theta|plano)P(plano)} = \\ &= 1 - P(esfera|\theta) \end{aligned} \tag{4.63}$$

Sin embargo, en un caso real, el modelo de distribución de objetos con caras planas de forma totalmente aleatoria siguiendo la función de densidad $f(\theta|plano) = 2\sin(\theta)\cos(\theta)$, no se llega a cumplir. Esto se debe a que en un caso no ideal existen ligaduras que restringen la orientación aleatoria y uniforme de las normales sobre un casquete esférico, y como consecuencia la distribución se limita a un subconjunto de las posibles direcciones. Un ejemplo claro puede ser el de una caja cúbica colocada sobre el suelo. La normal de la cara frontal de la caja, al estar siempre colocada sobre una superficie horizontal, puede variar aleatoriamente pero limitada siempre a estar situada en un plano horizontal. El resultado de este fenómeno hace que la probabilidad de encontrarnos con planos con θ próximas a cero aumente considerablemente y se produzca un corrimiento del máximo ideal centrado en 0.78 radianes en mayor o menor grado hacia 0 radianes.

Para ilustrar este fenómeno y para derivar un valor óptimo de θ para el caso concreto de objetos planos orientados de forma aleatoria que forman parte del fondo de una

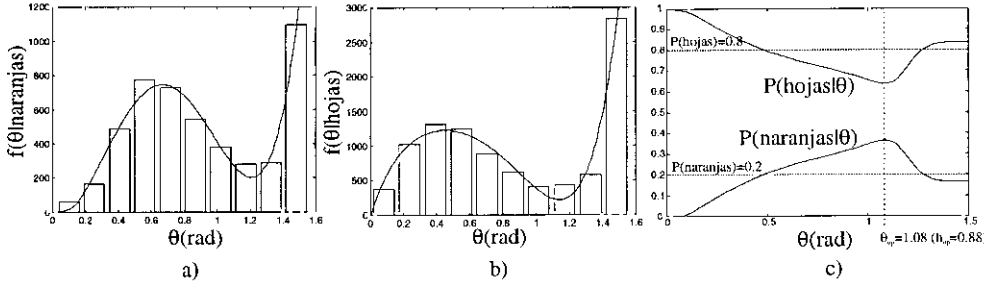


Figura 4.28: Función densidad de probabilidad condicional para: (a) naranjas y (b) hojas. (c) Probabilidad de que un punto con normal θ pertenezca a una naranja o al fondo de hojas.

escena de árboles frutales, vamos a derivar de forma experimental las funciones de densidad de probabilidad para el caso de hojas (planos) y naranjas (esferas). Hemos adquirido diversas imágenes conteniendo de forma aislada una muestra de solo naranjas y de solo hojas; se ha calculado el ángulo de sus normales con el eje de medición y se han representado los correspondientes histogramas (4.28a y b). A continuación hemos ajustado los valores experimentales por un polinomio, pudiéndose observar los resultados en las líneas continuas de la misma figura. Puede observarse que la distribución de probabilidad para el caso de las naranjas se corresponde de forma fiel a la distribución ideal de una esfera. Igualmente en el caso de las hojas, se observa un corrimiento del modo central hacia la zona de menor θ. El pico en los histogramas para valores de θ próximos a $\pi/2$ se corresponde al cálculo de las discontinuidades existentes entre los objetos presentados y el fondo existente en las imágenes, y por tanto al no ser propios del objeto correspondiente no supone ninguna desviación respecto a lo esperado.

Aplicando de nuevo el teorema de Bayes obtenemos las probabilidades condicionales de la figura 4.28c. Podemos ver que en este caso existe un valor de $\theta = 1.08$ rad, para el cual se maximiza la probabilidad de que un punto pertenezca a una naranja y por tanto se minimiza la probabilidad de que corresponda al fondo. Las probabilidades a priori utilizadas han sido de $P(naranja) = 1/5$ y de $P(hoja) = 4/5$, lo cual se deduce del estudio de distribución espacial de naranjas en un árbol mediante el cual se encontraba que en una ventana cuadrada de 500×500 mm el promedio de naranjas presentes era de 10. Debido a que las probabilidades a priori son menores para el caso de una naranja que para las hojas, con este valor óptimo de $\theta_{op} = 1.08$ vamos a ser capaces de marcar una serie de puntos en la imagen de los cuales una minoría de ellos pertenecerán a naranjas $P(naranja|\theta_{op})$ y la mayoría corresponderán a hojas $P(hojas|\theta_{op})$. Sin embargo, el cociente entre ambas será máximo en ese punto óptimo, con lo cual la relación señal ruido también lo será.

A partir del valor óptimo de θ ($\theta_{op} = 1.08$) podemos deducir la fracción h del radio de una esfera en la que la superficie tiene una normal con una inclinación de θ_{op} radianes, esta fracción resulta ser $h_{op} = 0.88$. Si aceptamos que una corona debe tener una anchura de al menos un punto para detectar una esfera, entonces aplicando un margen doble de seguridad para que los factores de discretización se atenúen, las primitivas corona se

obtienen marcando aquellos puntos que estén en el intervalo $[h_{op} - \frac{s}{R}, h_{op} + \frac{s}{R}]$; donde s es la resolución espacial y R el radio de la esfera.

Expresando el rango de interés en función de la pendiente mD en el mapa de distancias, que se relaciona con h mediante la relación $mD = \frac{h}{\sqrt{1-h^2}}$, obtenemos la siguiente condición a satisfacer por las primitivas corona:

$$\frac{h_{op} - \frac{s}{R}}{\sqrt{1-(h_{op} - \frac{s}{R})^2}} \leq mD \leq \frac{h_{op} + \frac{s}{R}}{\sqrt{1-(h_{op} + \frac{s}{R})^2}} \tag{4.64}$$

Para mejorar la relación señal ruido, es necesario reducir el número de puntos marcados como primitivas *corona* que realmente no pertenecen a superficies esféricas. Para ello utilizaremos, a parte de las pendientes en los mapas de distancia mD, la información de reflectancia disponible. Asumiendo superficies con reflectividad uniforme tenemos que los objetos esféricos al tener una curvatura convexa, presentan una pendiente característica en el mapa de reflectancia que viene definida por la expresión $m\Re|_{esfera} = \frac{-80.9}{ln(10)} \frac{h}{R(1-h^2)}$. La integración de la condición en el mapa de distancias (ec. 4.64) con la siguiente condición para las pendientes en el mapa de reflectancia

$$\frac{h_{op} - \frac{s}{R}}{1-(h_{op} - \frac{s}{R})^2} \frac{80.9}{R\ln(10)} \leq m\Re \leq \frac{h_{op} + \frac{s}{R}}{1-(h_{op} + \frac{s}{R})^2} \frac{80.9}{R\ln(10)} \tag{4.65}$$

permite eliminar la mayoría de puntos que corresponden a planos que casualmente presentaban una orientación igual a θ_{op}, ya que un plano con reflectancia uniforme presenta siempre un $m\Re$ igual a cero, independientemente de cual sea θ. Además, para mejorar aun más la relación señal ruido exigimos que haya una igualdad de orientaciones en las pendientes de ambos mapas, lo cual se expresa con la siguiente condición:

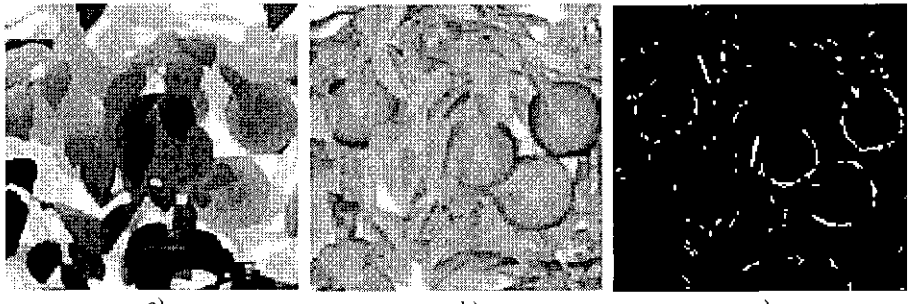
$$\left| \arctan(\frac{\nabla D_y}{\nabla D_x}) - \arctan(\frac{\nabla \Re_y}{\nabla \Re_x}) \right| \leq \pi/6 \tag{4.66}$$

donde el rango angular permisible de $\pi/6$ no es crítico y fué seleccionado por ajuste empírico.

Finalmente, en la figura 4.29 mostramos un ejemplo de una pareja de imágenes distancia/reflectancia (a y b) de las cuales se han extraído las primitivas *corona* (c) aplicando las tres condiciones que acabamos de exponer.

4.3.3 Primitivas regionales convexas

Las primitivas *puntuales* vistas con anterioridad (contorno y corona) centran su atención sobre la periferia de los objetos esféricos. Al ser puntuales soportan altos grados de oclusión del objeto, sin embargo cuando la periferia queda ocluida las regiones internas visibles no generan ningún indicio referente a la presencia del objeto. El objetivo de las primitivas regionales es la captura de indicios a partir de las regiones internas de la superficie de una esfera. Una de las características que nos da indicios de esfericidad es la existencia de una superficie con curvatura negativa lo cual da evidencia de su convexidad. Este principio de convexidad es el que va a ser utilizado para definir las primitivas denominadas *primitivas regionales convexas.*



a) b) c)

Figura 4.29: Ejemplo de extracción de primitivas *corona*: (a) y (b) imágenes de distancia y reflectancia, y (c) primitivas puntuales corona; apréciese los semiarcos que dan una clara impresión visual de la ubicación de los cuatro frutos.

En primer lugar es necesario obtener las regiones sobre las cuales se ha de realizar la caracterización. Los puntos suaves obtenidos en la etapa de clasificación de puntos propuesta en la sección 4.3.1, serán la base para obtener cada uno de los segmentos regionales. Estas son las etapas de procesamiento necesarias para obtener dichos segmentos a partir de los puntos clasificados como suaves:

- *Erosión.* Se realiza una erosión en la imagen de *puntos suaves* con el objeto de separar puntos conectados que pertenecen a regiones de diferentes objetos.

- *Dilatación.* Posteriormente se realiza una dilatación con preservación del número de Euler para recuperar el área inicial pero sin volver a unir las regiones previamente separadas.

- *Conexión de componentes.* La conexión de componentes es un proceso de etiquetado de cada punto de la imagen donde se asigna una misma etiqueta a aquellos puntos que pertenecen a la misma región. Por cada región aislada adicional que se va encontrando se utiliza una nueva etiqueta de tal forma que al final existen tantas etiquetas como regiones encontradas.

- *Eliminación de regiones pequeñas.* Aquellas regiones con un área inferior a un cierto umbral (típicamente el 15% del área de una esfera de radio R), son rechazadas por no presentar un tamaño suficiente como para generar a partir de ellas estimaciones paramétricas estables.

Una vez que tenemos las regiones perfectamente identificadas estimamos la curvatura media $\hat{K}$ y la desviación estándar de cada una de ellas $\hat{\sigma_K}$. En el cómputo de la curvatura se aplica el método definido en las ecuaciones 4.41 a 4.43 para cada punto de la región, y a partir de estos, se estima el valor de curvatura medio y la varianza en la región. El cálculo de los vectores unitarios normales a las superficies en el mapa de distancias (necesarios para derivar la curvatura) lo hacemos a partir de las siguientes expresiones:

$$\vec{n} = \vec{x}_u \times \vec{x}_v = \frac{1}{\sqrt{1 + \frac{\partial f}{\partial x}^2 + \frac{\partial f}{\partial y}^2}}(-\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1) \tag{4.67}$$

$$\vec{n} = \frac{1}{\sqrt{1 + a^2 + b^2}}(-a, -b, 1) \tag{4.68}$$

donde a y b son los parámetros de un plano ajustado a una región de tamaño 3×3 que se define mediante la ecuación $f(x, y) = ax + by + c$.

Aquellas regiones que tengan una convexidad próxima a la esperada y dicha convexidad provenga de la integración de curvaturas puntuales con una dispersión acotada, pasarán a la categoría de primitivas regionales convexas. La curvatura esperada K será igual al inverso del radio medio esperado de las esferas a detectar, es decir, $K = 1/R$. Para evitar considerar regiones casi planas o con alta convexidad, solamente aceptaremos regiones con convexidad acotada entre $-2K$ y $-0.5K$, con lo cual estamos permitiendo una variabilidad del tamaño de las esferas de un 100% en ambos sentidos. La restricción de baja dispersión se consigue al exigir que debe haber una probabilidad casi nula de que haya puntos en la región con características de concavidad, lo cual se consigue exigiendo que $\hat{\sigma_K} \leq |0.5\hat{K}/3|$. Por tanto estas son las condiciones a satisfacer simultáneamente:

$$(-2K \leq \hat{K} \leq -0.5K)\ Y\ \left(\hat{\sigma_K} \leq |0.5\hat{K}/3|\right) \tag{4.69}$$

4.3.4 Primitivas regionales reflectividad

Estas primitivas tratan de aportar indicios sin basarse en las formas, como se hacía en los casos anteriores, sino basándose en propiedades ópticas superficiales como la reflectividad. Si los objetos que queremos reconocer tienen unas propiedades ópticas que son separables frente a la de los objetos del fondo, un análisis como el de reflectividad permitirá realizar la clasificación de una forma muy directa.

Supongamos que conocemos un valor umbral de reflectividad ρ_u que nos permite separar los objetos deseables del resto del fondo. Bastaría calcular la reflectividad ρ de cada una de las regiones y evaluar a qué lado del umbral de reflectividad nos encontramos, con lo cual determinaríamos si una región pertenece al fondo o a una esfera. Para determinar la reflectividad de una región realizamos el promedio de las reflectividades puntuales correspondientes a cada punto de la región. Para ello aplicamos el modelo del láser deducido en el capítulo 2, de donde despejando la reflectividad tenemos:

$$\rho = \frac{10^{\Re/80.9}\ D^2}{\cos\theta\ \alpha(D)\ \frac{A\ F_i}{\pi}} \tag{4.70}$$

donde $\Re$ es la reflectancia captada por el telémetro, D es la distancia, F_i es el flujo radiante emitido (que es de 8 mW en nuestro caso), A es el área de la óptica de recepción, $\alpha(D)$ es un término experimental que depende de la distancia y θ es el ángulo formado entre el eje de emisión del haz láser y la normal de la superficie.

Para calcular θ utilizamos una aproximación válida para incrementos angulares pequeños, que permite obtener θ trabajando directamente sobre el mapa de distancias en

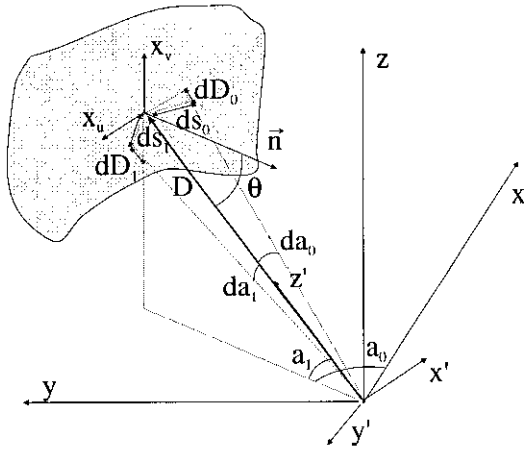


Figura 4.30: Esquema explicativo de la aproximación utilizada para el cálculo de θ.

coordenadas esféricas sin tener que hacer una transformada de coordenadas al sistema cartesiano XYZ. Para ello utilizamos un sistema de coordenadas X'Y'Z' solidario con el haz láser (fig. 4.30) que nos permite definir los vectores directores $\vec{x}_u$ y $\vec{x}_v$ de la superficie en términos de incrementos en distancia dD_0, dD_1 y diferenciales de desplazamiento ds_0, ds_1 en las direcciones X' y Y'. A continuación presentamos la deducción de θ planteada como el producto escalar de la normal a la superficie $\vec{n}$ y el vector $\vec{D}$ que define el punto de medida respecto al sistema de coordenadas X',Y',Z'.

$$\vec{D} = (0, 0, D) \tag{4.71}$$

$$|\vec{D}| = D \tag{4.72}$$

$$\vec{n} = \vec{x}_u \times \vec{x}_v = (1, 0, \frac{\partial D}{\partial s_0}) \times (0, 1, \frac{\partial D}{\partial s_1}) = (-\frac{\partial D}{\partial s_0}, -\frac{\partial D}{\partial s_1}, 1) \tag{4.73}$$

$$|\vec{n}| = \sqrt{\left(\frac{\partial D}{\partial s_0}\right)^2 + \left(\frac{\partial D}{\partial s_1}\right)^2 + 1} \tag{4.74}$$

$$\cos\theta = \frac{\vec{n} \cdot \vec{D}}{|\vec{n}||\vec{D}|} = \frac{D}{|\vec{n}|D} = \frac{1}{|\vec{n}|} = \frac{1}{\sqrt{(\frac{\partial D}{\partial s_0})^2 + (\frac{\partial D}{\partial s_1})^2 + 1}} \simeq \tag{4.75}$$

$$\simeq \frac{1}{\sqrt{(mD_x)^2 + (mD_y)^2 + 1}} \tag{4.76}$$

Por tanto θ se calcula, a partir de las pendientes mD_x y mD_y en las direcciones x e y del mapa de distancias, de esta forma:

$$\theta \simeq \arccos\left(\frac{1}{\sqrt{(mD_x)^2 + (mD_y)^2 + 1}}\right) \tag{4.77}$$

Para el cálculo del valor umbral de reflectancia, ρ_u, que nos permite separar el fondo de los objetos, utilizamos un proceso de aprendizaje automático que consiste en captar dos imágenes diferentes, una conteniendo una muestra de los objetos de interés y otra presentando una configuración típica de fondo. Por cada una de estas imágenes se calcula la reflectividad de cada punto y se parametriza una campana de Gauss mediante su valor medio y su desviación estándar, obteniendo por tanto dos gaussianas cada una de ellas representando la función de densidad de probabilidad para los objetos de interés y el fondo. Dichas campanas se solaparán, y para algún valor de ρ se cumplirá la condición de igual probabilidad de pertenecer a ambas clases. El cálculo del valor ρ_u que marca la frontera de decisión de este clasificador bayesiano se obtiene igualando las funciones de probabilidad:

$$pdf_1(\rho) = \frac{P(1)}{\sqrt{2\pi}\sigma_1} \cdot e^{-1/2(\frac{\rho-\mu_1}{\sigma_1})^2} \tag{4.78}$$

$$pdf_2(\rho) = \frac{P(2)}{\sqrt{2\pi}\sigma_2} \cdot e^{-1/2(\frac{\rho-\mu_2}{\sigma_2})^2} \tag{4.79}$$

$$pdf_1(\rho_u) = pdf_2(\rho_u) \tag{4.80}$$

tomando logaritmos neperianos tenemos,

$$ln(\frac{P(1)}{\sqrt{2\pi}\sigma_1}) - \frac{1}{2}(\frac{\rho_u - \mu_1}{\sigma_1})^2 = ln(\frac{P(2)}{\sqrt{2\pi}\sigma_2}) - \frac{1}{2}(\frac{\rho_u - \mu_2}{\sigma_2})^2 \tag{4.81}$$

$$\underbrace{(\sigma_1^2 - \sigma_2^2)}_{A}\rho_u^2 + \underbrace{2(\mu_1\sigma_2^2 - \mu_2\sigma_1^2)}_{B}\rho_u + \underbrace{\mu_2^2\sigma_1^2 - \mu_1^2\sigma_2^2 + \sigma_2^2\sigma_1^2 2ln(\frac{P(1)\sigma_2}{P(2)\sigma_1})}_{C} = 0 \tag{4.82}$$

con lo cual el umbral de reflectancia se obtiene así:

$$\rho_u = \begin{cases} \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} & \text{si } \sigma_1 \neq \sigma_2 \\ \frac{\mu_1 + \mu_2}{2} + \frac{\sigma^2 \, ln(P(1)/P(2))}{\mu_1 - \mu_2} & \text{si } \sigma_1 = \sigma_2 \end{cases} \tag{4.83}$$

El aprendizaje o cálculo de ρ_u se hace con anterioridad a que el proceso de reconocimiento comience (off-line). Este método permite realizar una clasificación de cada región segmentada de una forma eficaz pues solamente es necesario hacer una comparación entre la reflectividad de cada región ρ_{region} y ρ_u, y no es necesario utilizar funciones de decisión que operen continuamente con las funciones de densidad de probabilidad para realizar la clasificación.

Como ejemplo en la figura 4.31 mostramos el resultado de realizar un aprendizaje del valor umbral ρ_u, el cual se ha hecho captando una imagen con sólo naranjas (objetos deseables) y otra con un fondo compuesto de hojas. En ambos casos los objetos utilizados

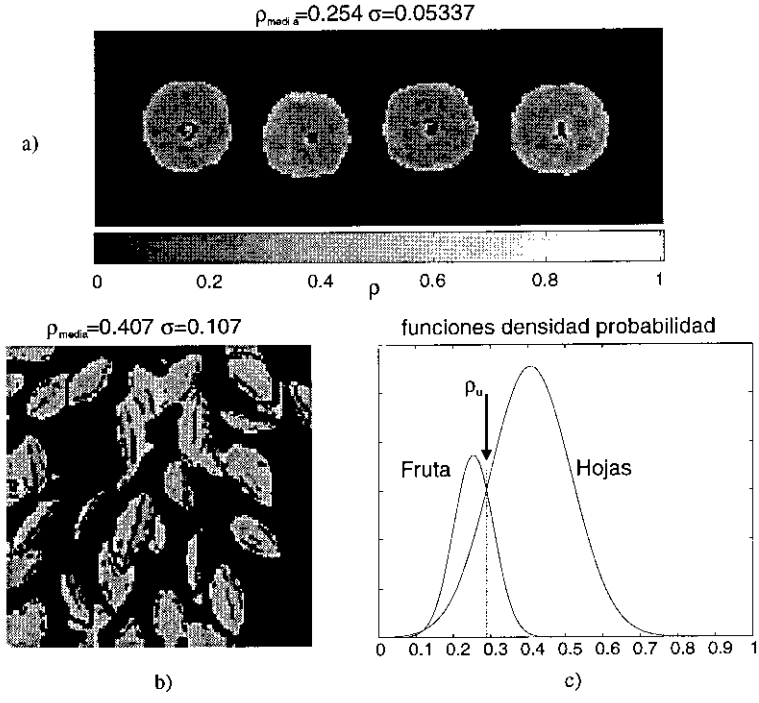


Figura 4.31: Selección del umbral de reflectividad ρ_u para objetos artificiales: (a) cálculo de la reflectividad media y varianza de una muestra de naranjas, (b) el mismo cálculo para una muestra de hojas, (c) funciones de densidad de probabilidad y el valor óptimo de reflectividad que separa ambas clases.

no son naturales y se corresponden a los frutos y hojas artificiales disponibles en el laboratorio. En este caso el valor de reflectividad medio de las naranjas es de 0.254, mientras que las hojas tienen una reflectividad media de 0.407, obteniéndose un valor umbral $\rho_u = 0.28$ para una probabilidad a priori de las naranjas y hojas de P(1)=1/5 y P(2)=4/5, respectivamente.

4.4 Estimación de parámetros e indicios

Una vez obtenidas las diversas primitivas que captan indicios de esfericidad (en los tres primeros tipos de primitivas) e indicios basados en propiedades ópticas (en las primitivas reflectividad), el siguiente paso consiste en el cálculo de los parámetros que definen cada esfera, es decir, el radio, su posición espacial y la reflectividad de su superficie. Este último parámetro podría ser interesante como información adicional para posteriores etapas de clasificación atendiendo a propiedades ópticas, o incluso, para realizar recolecciones selectivas de los frutos. El proceso de extracción de parámetros se hará de una forma individual por cada tipo de primitivas puesto que cada una de ellas

requiere un tratamiento especial.

Una vez que se tenga la lista de los parámetros de las presuntas esferas, será también necesario dar un grado de confianza a cada una de estas estimaciones indicando la certeza que tenemos sobre esa estimación. Es decir, no se tendrá la misma confianza sobre los parámetros estimados a partir de un número reducido de primitivas puntuales, que si estas son muy numerosas y por tanto generan un alto número de indicios; en este último caso los parámetros estimados serán mucho más fiables y el hecho de que esos datos provengan de objetos realmente esféricos también será mayor. Por tanto, junto a los *parámetros* también generaremos los *indicios* acumulados, los cuales serán fundamentales en futuras etapas de integración de las hipótesis parciales generadas a partir de cada tipo de primitiva.

Debido al carácter diferenciado de las primitivas puntuales (contorno y corona) frente a las regionales (convexas y reflectividad), en la estimación de parámetros vamos a utilizar dos procesos claramente diferenciados. En el primer caso se aplicará un proceso que utiliza la transformada circular de Hough, y en el segundo, utilizaremos una técnica basada en un ajuste por mínimos cuadrados. A continuación describimos ambos procesos.

4.4.1 Estimación con primitivas puntuales

Las primitivas puntuales presentadas captan arcos esféricos que delimitan el contorno de una esfera. De una forma visual hemos visto en la sección anterior que efectivamente cuando no hay alta oclusión este tipo de primitivas generan a un observador humano una clara información para determinar la presencia de una esfera. Sin embargo, el volumen de información susceptible de ser analizado ha disminuido drásticamente, ya que las primitivas puntuales se corresponden simplemente con los "unos" de una imagen binaria y el fondo, que es mayoritario y viene representado por "ceros", no debe ser analizado. Por tanto, mediante la utilización de primitivas puntuales el proceso se ha simplificado, pasando de una detección de esferas tridimensionales en entornos altamente no estructurados, a la detección de arcos esféricos en imágenes binarias con una relación señal ruido altamente mejorada.

El método más evidente para detectar los arcos semiesféricos que generan las imágenes de primitivas, es el del emparejamiento de plantillas (template matching) [172]. En este caso bastaría realizar una correlación de un modelo de arco con diferentes radios e ir explorando toda la imagen moviendo el modelo punto a punto por ella; los puntos de alta correlación se corresponderían con presuntas esferas y sus parámetros serían los del modelo utilizado. Sin embargo esta técnica no es muy eficiente ya que exige explorar todos los puntos de la imagen. Plá [164], presenta un método de detección de contornos circulares ocluidos utilizando una técnica de agrupamiento de puntos contiguos para formar segmentos de arco que tienen una curvatura uniforme. Posteriormente agrupa estos segmentos por proximidad y estima los parámetros mediante un ajuste por mínimos cuadrados. La debilidad de esta técnica radica en que los puntos del contorno requeridos a la entrada del algoritmo deben ser contiguos o estar formando contornos cerrados. Sin embargo, esta condición no es siempre posible que se cumpla como sucede con nuestras primitivas puntuales. Análogamente, Jacobs [94] también presenta resultados similares detectando grupos de segmentos lineales convexos.

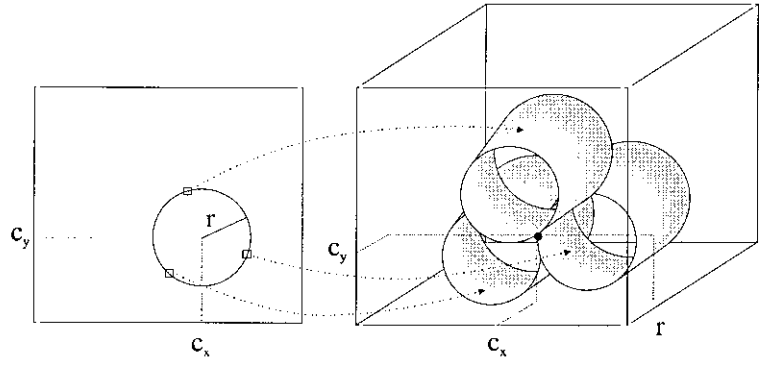


Figura 4.32: Proceso de votación mediante la transformada circular de Hough.

La técnica conocida como transformada de Hough [47, 120, 219, 91, 90, 92, 3] es un método más eficiente de "template matching" ya que no recorre toda la imagen sino solo los puntos activos de ella; en nuestro caso solo utilizaría las primitivas puntuales y el fondo no sería explorado. Además es robusta en el sentido de que soporta ruido impulsional y no requiere que los puntos que definen un contorno estén agregados y formen arcos continuos. Por ello es una técnica muy adecuada a nuestro propósito.

En la primera sección de este capítulo explicamos en qué consiste la transformada de Hough para el caso de querer detectar líneas rectas, y también para arcos esféricos. En el segundo caso, los parámetros a detectar son: el centro de la circunferencia (c_x,c_y) y su radio r. Por tanto el espacio de parámetros a donde vamos a realizar la transformación es un espacio tridimensional con coordenadas c_x, c_y y r. En la figura 4.32 se puede ver el proceso de transformación o votación, de una imagen conteniendo una circunferencia, en su espacio transformado. En este proceso por cada punto de la circunferencia se genera un tronco de cono de votos que se acumula en el espacio de parámetros. Vemos que existe un punto común de intersección donde se produce una alta votación que se corresponde con los parámetros de la circunferencia que queremos detectar. Por tanto, mediante la transformada circular de Hough, la detección de circunferencias se ha simplificado en la detección de picos o máximos locales en el espacio transformado.

El principal problema de la transformada de Hough es el tamaño del espacio transformado, el cual si es muy grande provoca problemas de reducción de eficiencia ya que el proceso se basa en la votación y detección de picos en este espacio, y por tanto al crecer dicho espacio, la complejidad de los algoritmos aumenta con el mismo orden de magnitud [91]. Para reducir la complejidad de los algoritmos hemos utilizado las siguientes técnicas de optimización:

- *Discretización del espacio transformado.* Para llevar a cabo el proceso de votación sobre un computador digital, se recurre a la utilización de una matriz de acumuladores que es una versión discreta del espacio transformado. Por simplicidad, el espacio transformado lo hemos discretizado, en sus coordenadas c_x y c_y, de igual forma que está discretizada la imagen sobre la cual queremos

detectar los arcos esféricos. Una discretización más grosera sería también adecuada ya que el número de acumuladores se reduciría más, aumentándose la eficiencia, aunque en contrapartida la precisión en la determinación de las coordenadas se vería reducida. La discretización en r se hace centrada en torno al radio R esperado de las esferas, con intervalos regulares correspondientes a la resolución espacial de paso entre puntos de la imagen inicial.

- *Incorporación del parámetro distancia en la votación.* Los tres parámetros c_x, c_y y r que definen la circunferencia formada por las primitivas puntuales, también son parte de las coordenadas tridimensionales del objeto esférico que estamos buscando, sin embargo, para terminar de definir su posición tridimensional necesitamos conocer la distancia d a la que se encuentra el centro de dicho objeto. Para obtener este cuarto parámetro aprovechamos el proceso de votación de la transformada de Hough y por cada acumulador, a parte de almacenar el número de votos, añadimos en promedio la distancia $D_{i,j}$ de cada punto votante. Debido a que todas las primitivas puntuales que contribuyen en la votación de cada esfera deben encontrarse a la misma distancia, el promedio final obtenido es una buena estimación de la distancia a la que se encuentra el objeto esférico. Además, la existencia de puntos votantes situados a una distancia alejada del valor final, permite determinar aquellos puntos que no se deben considerar para realizar votaciones, por provenir de otras superficies.

- *Votación reducida usando dirección del gradiente [120].* La transformada circular de Hough realiza por cada punto imagen y para cada uno de los radios considerados, una votación sobre una circunferencia en el plano $c_x - c_y$ (fig. 4.32). Este proceso de transformación de un punto a múltiples se puede optimizar conociendo la dirección del gradiente de cada punto imagen, ya que el centro de una circunferencia siempre está en la dirección opuesta a la del gradiente de un punto perteneciente al contorno de una esfera. De esta forma es posible reducir la votación por cada punto, ya que en vez de hacerlo sobre una circunferencia completa ahora solo se hace sobre un arco de circunferencia. El rango angular de la votación depende de la incertidumbre en la determinación de la dirección del gradiente, que idealmente debería pasar por el centro de la esfera que generó las primitivas puntuales. En nuestro caso hemos trabajado con un rango angular de 30^0 con lo cual se consigue reducir el proceso de votación en más de un orden de magnitud.

- *Retrotransformación [60].* La retrotransformación (Backtransform) se fundamenta en repetir el proceso de votación de nuevo, pero a diferencia del primer paso, en este segundo proceso se utiliza la información de la votación anterior para depositar solamente un voto por cada punto imagen. La casilla del acumulador elegida será aquella que presente el mayor número de votos de las casillas dentro del ámbito de votación del punto imagen. El resultado que se obtiene es una acentuación de los picos en el volumen de votación, por tanto se mejora la relación señal ruido y se facilita la detección de picos, la cual se puede hacer aplicando un umbral. Para eliminar la mayor parte del ruido y dejar los candidatos más votados aplicamos un umbral que se corresponde a una votación de un arco circular de 30^0.

Una vez efectuada la umbralización, realizamos un *agrupamiento* (clustering) de las

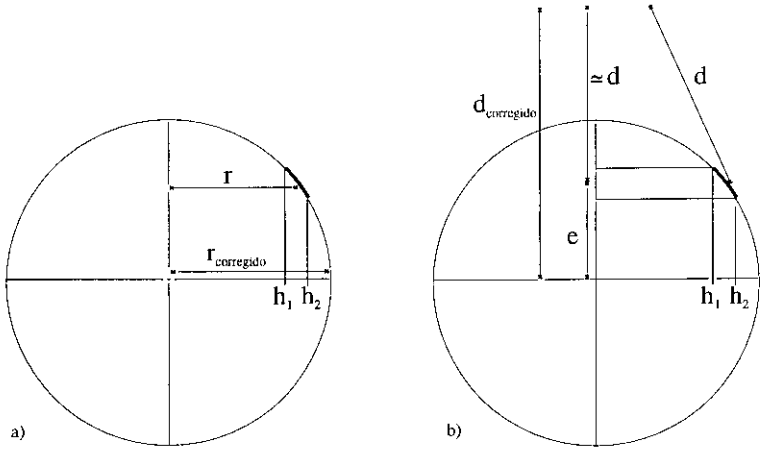


Figura 4.33: Correcciones en radio y distancia a efectuar después de aplicar la transformada circular de Hough sobre las primitivas puntuales.

casillas que han superado el umbral de votación, de esta forma conseguimos obtener las agrupaciones naturales existentes [54]. El algoritmo de agrupamiento se basa en la distancia euclídea en el espacio transformado realizándose agrupaciones de las agregaciones más compactas. Los grupos finalmente detectados determinan los parámetros de las esferas, que se calculan promediando los parámetros de cada uno de los componentes agregados y sumando el número de votos.

Debido a la posición de las primitivas sobre la esfera, existen correcciones que hacer a dos de los cuatro parámetros obtenidos mediante la transformada circular de Hough:

- *El radio de la esfera.* Tanto las primitivas contorno como corona forman circunferencias de radio r menor que el de la esfera que las genera (fig. 4.33a), por tanto, la corrección a aplicar es esta:

$$r_{corregido} = \frac{r}{\frac{h_1+h_2}{2}} \tag{4.84}$$

- *La distancia a la esfera.* La distancia d obtenida en la votación es inferior a la distancia real al centro de la esfera, por ello, el error cometido e debe ser compensado (fig. 4.33b):

$$d_{corregido} = d + e = d + r_{corregido}\sqrt{1 - (\frac{h_1 + h_2}{2})^2} \tag{4.85}$$

Los indicios de esfericidad están directamente relacionados con el número total de votaciones realizadas y con la proximidad entre el tamaño de la esfera esperada R y la estimación de radio corregido $r_{corregido}$. Por tanto los indicios obtenidos a partir de las primitivas puntuales los calculamos así:

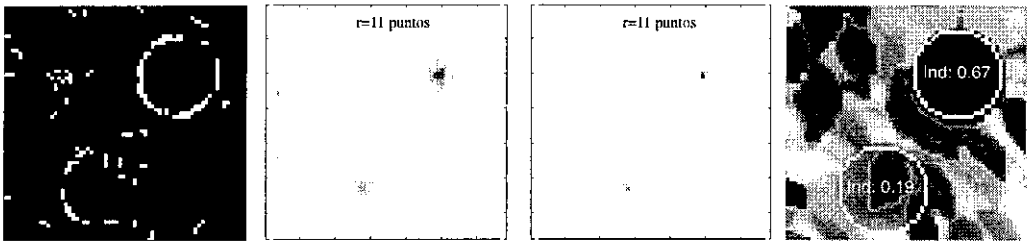


Figura 4.34: Estimación de parámetros e indicios a partir de primitivas puntuales utilizando la transformada circular de Hough: (a) primitivas puntuales, (b) CHT, (c) CHT después de aplicar backtransform, (d) parámetros e indicios estimados.

$$\text{Indicios} = \frac{\text{votos_grupo}}{\pi(\frac{R}{s})^2(h_2 - h_1)^2} \cdot \exp^{-|r_{corregido} - R|/R} \tag{4.86}$$

donde h_2 y h_1 son las fracciones de radio utilizadas para calcular las primitivas corona o contorno, y s es la resolución espacial.

Como ejemplo ilustrativo del proceso de transformación y estimación de los parámetros a partir de primitivas puntuales, en la figura 4.34 se muestran los resultados intermedios de las diferentes etapas de procesamiento aplicadas. Las imágenes b) y c) de esta figura muestran dos secciones del espacio transformado de Hough correspondiente a un radio r de 11 puntos. La primera sección (4.34b) se corresponde a la primera votación optimizada utilizando información de la dirección de los gradientes. Se puede apreciar ya un claro agrupamiento de las votaciones. En la segunda sección (fig. 4.34c) se muestra la votación después de aplicar la retrotransformación (backtransform). Se aprecia que los picos anteriores no tienen tanta dispersión y quedan mejor localizados. Las restantes secciones del espacio de Hough no han sido mostradas ya que no generan una acumulación de votos significativa, y por tanto no suponen una contribución apreciable en la determinación de los parámetros de la esfera.

4.4.2 Estimación con primitivas regionales

Una primitiva regional es un conjunto de puntos que pertenecen a una misma superficie y que se formaron por generar indicios de esfericidad por su curvatura o indicios de corresponder a la superficie del objeto buscado por tener una reflectividad típica. Los parámetros de la esfera se obtienen a partir de las primitivas regionales y la imagen de distancias, ajustando la ecuación de una esfera sobre la región en el mapa de distancias definida por cada una de las primitivas regionales. A continuación desarrollamos el proceso de ajuste aplicado, en el cual x_i, y_i y z_i son los datos o coordenadas de cada punto en la región, y x_c, y_c, z_c, r son los parámetros a determinar.

$$(x_i - x_c)^2 + (y_i - y_c)^2 + (z_i - z_c)^2 = r^2 \tag{4.87}$$

$$x_i^2 + y_i^2 + z_i^2 - 2x_i x_c - 2y_i y_c - 2z_i z_c \underbrace{+ x_c^2 + y_c^2 + z_c^2 - r^2}_{w} = 0 \tag{4.88}$$

$$2x_i x_c + 2y_i y_c + 2z_i z_c - w = x_i^2 + y_i^2 + z_i^2 \tag{4.89}$$

Si la región contiene n puntos podemos plantear un sistema de ecuaciones que se puede expresar de forma matricial de la siguiente forma:

$$\underbrace{\begin{pmatrix} 2x_1 & 2y_1 & 2z_1 & -1 \\ 2x_2 & 2y_2 & 2z_2 & -1 \\ & \vdots & & \\ 2x_i & 2y_i & 2z_i & -1 \\ & \vdots & & \\ 2x_n & 2y_n & 2z_n & -1 \end{pmatrix}}_{A} \cdot \underbrace{\begin{pmatrix} x_c \\ y_c \\ z_c \\ w \end{pmatrix}}_{p} = \underbrace{\begin{pmatrix} x_1^2 + y_1^2 + z_1^2 \\ x_2^2 + y_2^2 + z_2^2 \\ \vdots \\ x_i^2 + y_i^2 + z_i^2 \\ \vdots \\ x_n^2 + y_n^2 + z_n^2 \end{pmatrix}}_{b} \tag{4.90}$$

$$A \cdot p = b \tag{4.91}$$

Para realizar un ajuste por mínimos cuadrados la función de error $e(p)$ a minimizar, que depende del vector de parámetros a estimar p, es

$$e(p) = (Ap - b)^T \cdot (Ap - b) \tag{4.92}$$

con lo cual realizando sus derivadas parciales e igualando a cero tenemos

$$2A^T(Ap - b) = 0 \tag{4.93}$$

cuya solución es:

$$p = (A^T A)^{-1} A^T b \tag{4.94}$$

A este método de ajuste se le conoce como técnica de la seudoinversa [221], y permite obtener mediante un cómputo directo los parámetros x_c, y_c, z_c y r; este último después de utilizar la relación $r = \sqrt{x_c^2 + y_c^2 + z_c^2 - w}$.

En cuanto a los indicios de que los parámetros obtenidos correspondan realmente al objeto buscado, podemos decir en general que estos dependerán del tamaño de la primitiva regional. Específicamente para el caso de las primitivas regionales convexas, la similitud entre el radio esperado y el obtenido, así como el error obtenido en el ajuste a la esfera serán otros factores relacionados con la confianza en la estimación. De esta forma hemos utilizado la siguiente expresión para obtener la evidencia generada a partir de primitivas convexas:

$$\text{Indicios} = \frac{n}{\pi(\frac{R}{s} - 1)^2} \exp^{-\sum_{i=1}^{n}(z_{modelo}(x_i,y_i) - z_i)^2/n} \exp^{-|r-R|/R} \tag{4.95}$$

donde n es el número de puntos en la región, y $z_{modelo}(x_i, y_i) = z_c - \sqrt{r^2 - (x_i - x_c)^2 - (y_i - y_c)^2}$, representando el valor de distancia que corresponde al punto (x_i, y_i) según el modelo obtenido mediante el ajuste presentado.

En el caso de las primitivas regionales reflectividad, la proximidad entre el valor de referencia aprendido para los objetos deseables y la reflectividad superficial estimada para cada región, marcará igualmente el grado de confianza que podemos obtener a partir de esta propiedad discriminante. Por tanto para las primitivas reflectividad utilizamos esta expresión

$$\text{Indicios} = \frac{n}{\pi(\frac{R}{s}-1)^2} \exp^{-0.5(\frac{|\rho-\rho_{esfera}|}{\sigma_{\rho_{esfera}}})^2} \tag{4.96}$$

donde ρ es la reflectividad estimada en la región, ρ_{esfera} es la reflectividad aprendida inicialmente, y $\sigma_{\rho_{esfera}}$ es la desviación estándar obtenida en el proceso de aprendizaje.

4.5 Generación de hipótesis finales y verificación

En la etapa anterior por cada tipo de primitiva éramos capaces de obtener unas hipótesis parciales. Estas hipótesis consistían en el conjunto de parámetros definiendo la posición, radio y reflectividad del objeto buscado. Cada hipótesis parcial se generaban basándose solamente en el análisis de cada clase de primitivas. En esta última etapa pretendemos integrar las cuatro fuentes de indicios en una única hipótesis global que sea el resultado de la contribución de todas las fuentes de indicios. De esta forma aquellas hipótesis parciales demasiado débiles, o con pocos indicios para ser consideradas hipótesis globales consistentes, mediante la contribución de otros indicios provenientes de otras primitivas, pueden convertirse en candidatos válidos. De esta forma, estamos realizando una integración o fusión en la cual los cuatro métodos de análisis contribuyen en la detección de los objetos, aportando cada uno la información necesaria para poder derivar una decisión final.

La integración de las hipótesis parciales en las definitivas hipótesis globales, se realiza agrupando aquellas que se encuentran suficientemente próximas, en términos de la distancia euclídea, en el espacio que define su posición tridimensional. En concreto integramos aquellas hipótesis locales cuya distancia entre centros sea inferior a 3 veces la desviación estándar en la estimación de las coordenadas tridimensionales, que como se verá en el siguiente capítulo es inferior a 3 mm. En este caso, es decir, cuando hay integración entre dos o más hipótesis locales, los parámetros globales se obtienen haciendo un promedio ponderado, donde los pesos son los cocientes entre los indicios parciales y la suma total de indicios. Los indicios que acompañaran a la hipótesis global generada, se obtiene sencillamente sumando las indicios correspondientes a las hipótesis parciales. Finalmente, debe ser aplicado un umbral para eliminar aquellas hipótesis globales con indicios insuficientes y por tanto con alta probabilidad de que puedan ser debidos a ruido y ser causantes de que aparezcan detecciones falsas.

El umbral aplicado deberá ser seleccionado en función del nivel de ruido presente. Normalmente interesará conseguir el máximo número de detecciones correctas y evitar todas la detecciones falsas. En ese caso, y considerando imágenes de árboles frutales como se verá en el siguiente capítulo, un umbral de indicios en torno a 0.15 es el más adecuado. Nótese que en el caso de que las cuatro fuentes contribuyan de una forma total, al captar todas las primitivas de una forma plena, los indicios máximos alcanzables son 4.0, por tanto el umbral de 0.15 se corresponde a decir que con captar un 3.75%

de los máximos indicios posibles se tiene evidencia o certeza sobre la existencia de una esfera.

Debido a que estamos especialmente interesados en eliminar al máximo las detecciones indeseables, además de la umbralización basada en indicios también aplicamos una etapa de verificación que elimina aquellas poco probables hipótesis finales que con suficientes indicios no deben aceptarse, o bien por que no nos interesa su detección (p.ej. por ser esferas con parámetros alejados de los deseados) o bien por que admitir su existencia conduce a un resultado absurdo. Para eliminar esta circunstancia aplicamos una etapa basada en reglas que aplica y comprueba que se cumplan las siguientes condiciones:

- *Distancia absoluta.* Rechazamos cualquier hipótesis cuya posición espacial esté fuera del alcance del robot manipulador y del sistema de captación. Es decir

$$1 \text{ m} < \text{Distancia} < 4 \text{ m} \tag{4.97}$$

- *Tamaño esperado.* Eliminamos aquellas posibles hipótesis que presenten radios de la esfera r alejados del valor esperado R. En particular, debido a que la aplicación de detección de frutos presenta una variabilidad acotada, el rango válido lo establecemos entre el doble y la mitad del radio esperado R.

$$R/2 < r < 2R \tag{4.98}$$

- *Opacidad.* Rechazamos hipótesis que presenten puntos en su superficie que estén situados a mayor distancia que la que les correspondería atendiendo a su posición espacial y su radio. Es decir, estamos eliminando presuntas esferas que en su volumen presenten agujeros a través de los cuales se pueden ver superficies en planos posteriores a la esfera.

$$\{i, j | D(x+i, y+j) > \text{distancia} - \sqrt{r^2 - i^2 - j^2}, i^2 + j^2 < r^2\} = \emptyset \tag{4.99}$$

- *No Intersección.* Debido a que dos esferas no pueden ocupar el mismo volumen al mismo tiempo, dada una pareja de hipótesis rechazamos la de menor evidencia si los volúmenes ocupados por ambas esferas interseccionan.

$$\sqrt{(x_{c_1} - x_{c_2})^2 + (y_{c_1} - y_{c_2})^2 + (d_1 - d_2)^2} > r_1 + r_2 \tag{4.100}$$

4.6 Conclusiones

A lo largo de este capítulo los aspectos más destacables son los que a continuación se citan:

- Hemos presentado una breve pero completa *revisión de las técnicas* más habituales en la *interpretación de imágenes*, que tienen una relación directa con el reconocimiento de patrones y que se fundamentan en la extracción de características de cada patrón y su posterior clasificación. En el caso del procesamiento de imágenes vemos que existe una etapa previa conocida como segmentación que trata

de aislar cada objeto del resto de la imagen. De esta forma estamos obteniendo el segmento correspondiente a la entidad a clasificar, y por tanto a partir de aquí el procesamiento restante se centra en la extracción de características de cada segmento y en su posterior clasificación.

- Hemos propuesto una *estrategia de reconocimiento de objetos esféricos* basada en la acumulación de indicios generados mediante cuatro tipos de primitivas. Estos indicios aparecen al extraer primitivas, tanto de carácter puntual como regional, que han sido diseñadas especialmente para detectar propiedades discriminantes de los objetos buscados.

- Se han presentado nuevas *primitivas puntuales*, que integrando información de distancia y reflectancia, permiten clasificar los puntos de la imagen en dos clases: 1) puntos que corresponden a bordes de transición esférica (primitivas contorno) y 2) puntos con pendientes de máxima probabilidad de pertenecer a la superficie de una esfera (primitivas corona).

- Planteamos la *estimación de parámetros e indicios* a partir de primitivas puntuales y regionales. En el primer caso, se aplica una transformada circular de Hough modificada en varios aspectos para optimizar su eficiencia, y en el segundo se aplica un ajuste sobre las regiones definidas por las primitivas.

- Finalmente, se presenta la *integración de las diversas hipótesis parciales* generando hipótesis definitivas mediante el promedio ponderado de parámetros y acumulando los indicios provenientes de diferentes fuentes. Como criterio adicional para evitar posibles detecciones indeseables, se aplica un test basado en reglas que comprueba varias condiciones que conducen a determinar que la esfera no es de interés o a un absurdo sobre su existencia.

Capítulo 5

Evaluación conjunta del sistema y los algoritmos propuestos en condiciones simuladas y reales

Resumen. *En este capítulo presentamos la evaluación conjunta de los algoritmos de restauración y análisis propuestos en los capítulos 3 y 4 de la tesis, así como el comportamiento del sistema de exploración láser al operar en entornos naturales.*

Los algoritmos desarrollados admiten como entrada una pareja de imágenes distancia-reflectancia y generan una lista de hipótesis de objetos esféricos incluyendo la posición tridimensional del objeto, su radio y reflectividad. Ahora se presenta su evaluación utilizando datos simulados y empíricos. En la evaluación simulada utilizamos como criterios para medir la calidad de los resultados la precisión y repetitividad de los parámetros estimados, y los indicios de esfericidad generados. Dichos criterios son analizados al variar diversos parámetros (nivel de ruido, oclusión, resolución, etc.) afectando tanto a la degradación de las imágenes de entrada como a las referencias ajustables del algoritmo, observándose una aceptable robustez y unos errores de estimación subpuntuales. Así mismo, analizamos la influencia que tiene el uso de diferentes etapas de restauración, encontrándose que el uso del filtrado 3σ-MPF *presentado en el capítulo 3, genera estimaciones más precisas e incrementa el número de indicios de esfericidad. En la evaluación empírica utilizamos imágenes de árboles frutales tanto artificiales como naturales. Analizando los porcentajes de detecciones correctas y falsas, encontramos una solución de compromiso que con imágenes de 3 mm de resolución espacial y un umbral de indicios de 0.15, es posible detectar un 80% de los frutos presentes, no encontrándose detecciones falsas. Las pruebas en campo ponen de manifiesto que los frutos son distinguibles del fondo por reflectividad y que existen dos aspectos degradantes adicionales que afectan a la calidad de las imágenes captadas; nos referimos al viento y a la iluminación solar.*

En el capítulo 3 presentamos una técnica nueva de restauración de imágenes, denominada filtrado *3σ-MPF*, con la cual podíamos mejorar la calidad de las imágenes sin introducir distorsiones. En el capítulo anterior, hemos presentado una estrategia de reconocimiento, localización y caracterización de objetos esféricos; que trabajando a partir de una pareja de imágenes distancia-reflectancia, permite generar hipótesis de objetos esféricos con diferentes grados de confianza. Los algoritmos desarrollados en ambos capítulos, tras ser encadenados forman un sistema completo de procesamiento y análisis de imágenes que hemos aplicado a la detección de frutos con forma cuasi-esférica situados en entornos poco estructurados como los agrícolas. Sin embargo, teniendo en cuenta que estamos interesados en que estos algoritmos se apliquen en dicho caso práctico, es imprescindible realizar una evaluación del conjunto de los algoritmos, para analizar de forma pormenorizada sus posibilidades y sus limitaciones.

5.1 Introducción a la evaluación de algoritmos

5.1.1 Consideraciones iniciales en la evaluación

Hace más de 10 años, se mencionaba que los trabajos desarrollados en el campo de la visión artificial sufrían de una carencia en el desarrollo de trabajos teóricos [72]. Sin embargo en los últimos años se ha reconocido que los desarrollos teóricos y algorítmicos, no son útiles si no vienen acompañados de una evaluación utilizando datos empíricos [99]. La importancia de una correcta evaluación es doble. Por un lado se facilita la implantación de los algoritmos desarrollados en aplicaciones reales en las que existen ciertos problemas, ya que gracias a una buena evaluación el posible usuario es capaz de comprobar si los algoritmos presentados por el investigador van a operar correctamente para el tipo de datos y restricciones prácticas que van a caracterizar el caso real. En segundo lugar, la evaluación va a permitir comparar claramente diferentes algoritmos que persiguen resultados similares, permitiendo determinar cual es mejor que otro y bajo qué circunstancias. Así mismo, se evita la necesidad de codificar los algoritmos de otros investigadores para compararlos bajo los mismos criterios, lo cual no suele ser posible debido a que en las publicaciones no siempre se muestran todos los detalles de los desarrollos realizados.

En general podemos decir que no existen criterios comúnmente aceptados para evaluar, ni una clara metodología de prueba, y la experiencia en pruebas de otras áreas de la ingeniería no ha llegado suficientemente a la investigación en visión artificial. Además existen otras objeciones que disuaden a los investigadores de realizar y presentar evaluaciones en sus trabajos. Estos son algunos de los argumentos esgrimidos para justificar la no evaluación de algoritmos [53]:

- *"La evaluación depende de la tarea a realizar".* Sin embargo, si caracterizamos la evaluación en función de diversos parámetros afectando a los datos de entrada y al ajuste de los algoritmos, es posible juzgar la validez de los algoritmos para un amplio rango de aplicaciones distintas sin tener que ejecutar los algoritmos en cada caso.

- *"Las medidas de calidad no son comparables".* Efectivamente, si cada

investigador utiliza su propia métrica es difícil comparar unos algoritmos con otros, especialmente en etapas de bajo nivel, pero utilizando medidas objetivas como las aportadas por la estadística (varianzas y probabilidades), es posible comparar fácilmente los resultados y comprender el resultado de la evaluación.

- *"Hay muchos parámetros que ajustar en los algoritmos".* Es cierto, por tanto la evaluación crece exponencialmente con el número de parámetros. Este problema se puede reducir eligiendo solo aquellos parámetros que tienen un significado muy definido y un interés claro.

- *"La evaluación no está reconocida".* Las pruebas requieren mucho tiempo, y de forma aproximada se puede decir que la relación de tiempos teoría:codificación:prueba tiene la distribución 1:10:100. Este hecho hace que la evaluación quede marginada y se realicen publicaciones de teorías no probadas o que funcionan para uno o dos ejemplos. Sin embargo, estamos viendo que la evaluación es muy importante, es por ello que los trabajos en este campo deben ser reconocidos por editoriales y organismos de ayuda a la investigación.

5.1.2 Conceptos fundamentales en la evaluación de algoritmos

La idea básica en la evaluación de un algoritmo consiste en mostrar la dependencia de la calidad de los resultados del algoritmo con el tipo de entrada y de sus parámetros de control [33]. Es decir evaluar consiste en obtener una relación del siguiente tipo:

$$\text{Calidad_resultados} = f(\text{datos_entrada}, \text{parámetros_control_algoritmo}) \qquad (5.1)$$

donde la relación, representada por f, puede venir dada mediante una tabla, una ecuación o una gráfica.

La *caracterización de los datos* de entrada se puede hacer simplemente mostrando las imágenes procesadas, especificando el proceso de generación de estas imágenes, ó en el caso más habitual y útil, parametrizando las características propias de la imagen. En este último caso, se puede dar la magnitud del ruido que afecta a los datos, la distribución de este ruido, la frecuencia de muestreo, etc.

Los *parámetros de control de los algoritmos*, evidentemente también afectan a los resultados obtenidos. Estos parámetros, que normalmente tienen forma de umbrales ajustables, deben representar una función claramente comprensible, de tal forma, que sea fácil la interpretación de los resultados de la evaluación.

En cuanto a las *medidas para caracterizar los resultados* del algoritmo, estas deben ser también simples e intuitivas. Cuando la salida del algoritmo es discreta (p.ej en un proceso de clasificación) las medidas típicas pueden ser la probabilidad de detección correcta, de no detección ó de detección falsa. En el caso de salidas continuas (p.ej propiedades de un objeto como posición o tamaño), lo más conveniente es utilizar medidas estadísticas como la media, varianza, errores medios, etc. Estas métricas tienen la propiedad de ser fácilmente interpretables, pueden aplicarse en todo tipo de algoritmos, y por tanto, permiten una comparación entre aquellos que hayan sido evaluados de esta misma manera. Otro tipo de métricas de calidad más particulares, como la medida

de fragmentación de segmentos u otras métricas tan específicas como esta, no son recomendables por ser demasiado particulares y limitar el ámbito de la evaluación [181].

Bajo este concepto de evaluación, existen fundamentalmente tres tipos de estudios, que se diferencian en la forma de generar la relación de dependencia entre la calidad de los resultados, los datos de entrada y los parámetros de control del algoritmo:

- *Evaluación mediante análisis matemático.* En este caso a partir de la expresión matemática de los algoritmos, es posible propagar las características de los datos de entrada y obtener las métricas que caracterizan la calidad de los resultados. Normalmente los datos de entrada se representan mediante funciones de distribución de probabilidad o mediante matrices de covarianza, y mediante la propagación de esta información a lo largo del algoritmo, es posible obtener las funciones de distribución o covarianzas referidas a los resultados del algoritmo. La expresión obtenida constituye una evaluación matemática del algoritmo, ya que estamos relacionando características de los resultados con características de los datos de entrada, apareciendo igualmente dependencias con los parámetros ajustables del algoritmo. Este tipo de soluciones son válidas para algoritmos muy sencillos y linealizables, encontrándose situaciones de difícil tratamiento al crecer ligeramente la complejidad de los algoritmos [38, 211].

- *Evaluación mediante simulación con datos sintéticos.* La evaluación utilizando datos sintéticos con diferentes niveles de ruidos y degradaciones, es el método tradicional de evaluación para algoritmos complejos. Como característica positiva destacable está la disponibilidad de los valores de referencia reales, con lo cual es posible estudiar las desviaciones de los resultados del algoritmo frente a los valores ideales. Las distribuciones de probabilidad o las covarianzas en los resultados se obtienen a base de repetir la ejecución de los algoritmos con diferentes patrones de ruido y promediando, lo cual hace que se necesiten largos tiempos de evaluación al ser un proceso iterativo [211].

- *Evaluación empírica utilizando datos reales.* Este método de evaluación se suele utilizar para determinar la utilidad práctica en situaciones reales con datos y entornos sin controlar. En este caso no se puede apreciar la desviación de los resultados frente a las soluciones ideales, pues en muchos casos es desconocida, pero se puede evaluar la capacidad de trabajo del algoritmo ante degradaciones, que siendo influyentes en los resultados, no se contemplaron en la fase de evaluación simulada, usualmente por desconocimiento de su existencia o por un incompleto conocimiento de su comportamiento.

En nuestro caso, los algoritmos presentados son suficientemente complejos y ricos en no linealidades, como para no considerar el primer tipo de evaluación como una alternativa abordable de forma inmediata. Por tanto nos vamos a restringir a realizar la evaluación mediante datos simulados y empíricos. En la siguiente sección presentamos la evaluación simulada, y en la sección 5.3 mostraremos la evaluación empírica con imágenes captadas con el sensor láser presentado en el capítulo 2.

5.2 Evaluación de los algoritmos propuestos mediante simulación

En esta sección vamos a presentar tres tipos de evaluación utilizando imágenes simuladas. Por un lado deduciremos la complejidad de los algoritmos, o dicho de otro modo el orden de variación del tiempo de procesamiento en función del volumen de datos de entrada. En segundo lugar, analizaremos la sensibilidad de los algoritmos a degradaciones en los datos de entrada que podrían corresponder a perturbaciones reales, afectando de esta forma a la calidad de las imágenes, y por tanto, a la capacidad de reconocimiento y localización de los algoritmos. Finalmente, estudiaremos la influencia que tiene la fase de restauración de imágenes presentada en el capítulo 3, sobre los resultados generados por el conjunto de los algoritmos. Estos tres tipos de análisis se presentan en las siguientes subsecciones.

5.2.1 Complejidad algorítmica

El método estándar para analizar la calidad de un algoritmo, referida a los tiempos de cálculo, se hace utilizando el concepto de *órdenes de complejidad.* Los tiempos absolutos son importantes, pero no son tan significativos puesto que depende de otros factores, como el tipo de sistema de cálculo utilizado, la herramienta de programación, etc. Al medir la complejidad de los algoritmos lo que se analiza es cómo crece el tiempo de ejecución al aumentar el número de datos en la entrada. De esta forma si N representa el volumen de los datos de entrada es frecuente encontrarse con algoritmos cuyo tiempo de ejecución depende de N mediante una función logarítmica, lineal, lineal-logarítmica, cuadrática, cúbica, etc. En estos casos la representación utilizada es $O(\log N)$, $O(N)$, $O(N \log N)$, $O(N^2)$, $O(N^3)$, respectivamente. Evidentemente cuanto menor sea su orden de complejidad mejor resultará el algoritmo, especialmente para grandes volúmenes de datos, si bien es cierto que también pueden existir algoritmos de complejidad alta, que para un tamaño de datos limitado, tengan tiempos absolutos menores que otros de menor complejidad.

En nuestro caso, los algoritmos elaborados básicamente realizan un número limitado de visitas a cada punto de las imágenes, y por cada uno de estos puntos, en casi todos los casos, analizamos un vecindario de tamaño 3×3 ó 5×5. Esto significa que si denominamos N al número de puntos que contiene una imagen, la complejidad sería $O(N \times 3 \times 3)$, ó $O(N \times 5 \times 5)$ considerando el peor caso. Sin embargo, como N suele ser muy grande comparado con 5 ($N = 10.000$ para una imagen de 100×100 puntos) y además el tamaño del vecindario es constante y no depende de la entrada, la complejidad final debe ser lineal con los datos, es decir, $O(N)$. Este hecho lo podemos comprobar en la figura 5.1 donde se vé una clara relación lineal entre el tiempo de ejecución experimental y el número de puntos en la imagen.

Hay que destacar que los tiempos absolutos obtenidos, se refieren a la ejecución de los algoritmos sobre un computador Pentium 150 MHz, sin utilizar ningún tipo especial de placa para procesamiento de imágenes y por tanto utilizando solamente el propio microprocesador de la placa madre. Tiempos de 60 segundos para imágenes de 20.000 puntos no son excesivos teniendo en cuenta que los algoritmos no están optimizados

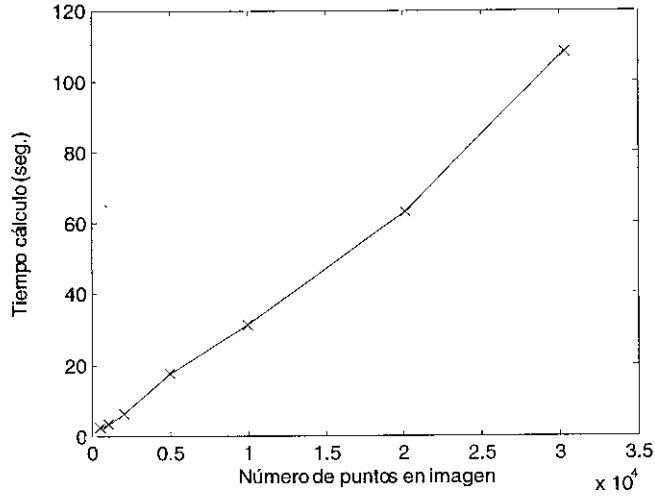


Figura 5.1: Dependencia del tiempo de ejecución con el tamaño de las imágenes.

para velocidad, y como se ha comentado, no se utiliza ningún tipo de soporte físico de procesamiento especial. Cabe esperar mejoras entre un orden y dos órdenes de magnitud en los tiempos absolutos de ejecución, cuando se realice una labor de ingeniería con el objeto de alcanzar ciclos de trabajo en tiempo real.

5.2.2 Sensibilidad ante degradaciones

Los algoritmos de percepción propuestos, partiendo de una pareja de imágenes distancia-reflectancia, generan hipótesis de detección de esferas, incluyendo la posición tridimensional del centro de cada una de ellas, su radio y la reflectividad media correspondiente a su superficie. Así mismo, cada hipótesis generada viene acompañada de los indicios totales acumulados mediante los cuatro métodos presentados de generación de hipótesis parciales. Estos indicios totales están acotados entre los valores 0 y 4, y por tanto cuanto más cerca estén de 4, mayor certeza sobre la existencia de la esfera existirá y mayor fiabilidad tendrán los parámetros estimados.

La ventaja de la evaluación simulada consiste en que se conoce de una forma fiel los parámetros reales que el algoritmo va a estimar. Por tanto, con este tipo de evaluación estamos en condiciones de obtener los errores medios y las repetitividades en la estimación de cada uno de los parámetros: posición 3-D, radio y reflectividad. Estos criterios de evaluación junto con los indicios de esfericidad serán los siete criterios de evaluación que utilizaremos para analizar la calidad del resultado de los algoritmos.

En cuanto al tipo de entrada utilizada, vamos a usar una imagen sintética de una esfera de radio 40 mm y con una reflectividad de 0.2, situada en el centro de una imagen con un fondo uniforme. La figura 5.2 muestra la pareja de imágenes distancia-reflectancia generadas sintéticamente y que serán la base a partir de la cual se generarán otras versiones degradadas de éstas, que permitirán estudiar los resultados ante diferentes

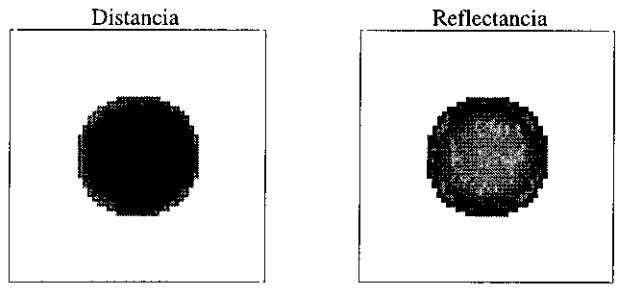


Figura 5.2: Pareja de imágenes sintéticas distancia-reflectancia generadas como base para la evaluación simulada.

configuraciones en la entrada del algoritmo. Los tipos de degradación introducidos a los datos son los siguientes:

- *Ruido gausiano.* Sobre la imagen de distancias añadiremos ruido con una distribución gausiana con desviaciones estándar entre 0 y 6 mm, con lo cual analizaremos el comportamiento del algoritmo no solo para los valores típicos de nuestra aplicación, 1-2 mm, sino para otras circunstancias más desfavorables.

- *Ruido impulsivo.* Un ruido impulsivo de ±300 mm de amplitud y diferentes porcentajes de actuación se adicionarán a la imagen de distancias. El rango aplicado variará entre un 0% y un 50% con lo cual podremos comprobar como varían los resultados cuando pasamos de no degradar ningún punto de la imagen hasta que la mitad de los puntos están alterados de forma impulsiva. Para la configuración láser presentada en el capítulo 2, los valores típicos de presencia de ruido impulsivo son prácticamente nulos, existiendo una cierta probabilidad ($< 10\%$) cuando las frecuencias de muestreo se aproximan a 50 KHz.

- *Oclusión.* La pareja de imágenes se irá ocluyendo paulatinamente de tal forma que el área visible de la esfera se reduzca. El porcentaje de oclusión introducido se define como la fracción entre el área no visible de la esfera y el área total del circulo resultante de proyectar la esfera sin oclusión sobre el plano. En un plantación típica de naranjos es visible para un ser humano entre un 40 y un 50% de la fruta existente, y dentro de este grupo los frutos presentan una distribución uniforme en el rango de oclusiones parciales [112].

- *Excentricidad.* La excentricidad de las esferas sintéticas, que para el caso de una esfera perfecta es 1 puesto que los valores de los ejes mayor y menor coinciden, se irá incrementando hasta conseguir elipsoides representando un modelo degradado de una esfera. De esta forma se podrá estudiar la sensibilidad del algoritmo ante discrepancias entre un objeto elipsoidal y el modelo perfecto de esfera que asumen los algoritmos. En un caso típico de fruta como la naranja, el grado de excentricidad de la fruta no suele ser superior a 1.1, lo cual indica que la longitud del eje mayor no suele superar en más de 10-14 mm la longitud del eje menor.

- *Resolución espacial.* La resolución espacial de muestreo de la imagen, dada por el intervalo en milímetros entre puntos, se variará para analizar su influencia en los resultados finales. Este análisis es muy importante puesto que nos interesa trabajar con la menor resolución posible para reducir tiempos de captura y procesamiento; aunque habrá que buscar un compromiso de tal forma que las estimaciones de los parámetros y la capacidad de captar indicios de esfericidad no se vea seriamente afectada.

En cuanto a los parámetros ajustables del algoritmo, variaremos la diferencia relativa entre el valor real de la esfera presente y el valor de referencia del radio esperado que se introduce como parámetro al algoritmo. El algoritmo está diseñado para buscar esferas de un determinado tamaño, lo cual se consigue al introducir el radio de referencia como parámetro, si las esferas presentes son de otro radio los indicios deberán disminuir y los parámetros estimados podrán verse afectados. Por tanto, junto a las cinco características de los datos de entrada, la *diferencia relativa de radios* definida como $|\frac{R_{real}-R_{refe}}{R_{refe}}|$, constituye la sexta característica con la cual vamos a realizar la evaluación sintética.

En la presente evaluación, para poder obtener valores de repetitividad, y a la vez hacer que los resultados no dependan de configuraciones particulares del ruido, cada una de las pruebas realizadas se repetirá 100 veces con lo cual debido a que hay seis características de degradación y vamos a emplear cinco valores discretos por cada una de ellas, tenemos un total de 3000 ejecuciones del algoritmo de reconocimiento sobre otras tantas parejas de imágenes que serán necesarias para poder obtener los resultados que a continuación mostraremos.

Las figuras 5.3a y 5.3b muestran respectivamente la influencia del *ruido gausiano*, en imágenes con 3 mm de resolución espacial, sobre los indicios medios de esfericidad captados y la precisión en la estimación de los parámetros de la esfera. Podemos ver que un aumento en el nivel de ruido provoca una reducción paulatina en los indicios, y por tanto en la evidencia de esfericidad, y un aumento de los errores de las estimaciones. Los errores de posición 3-D son siempre inferiores a 3 mm con una repetitividad menor de 1 mm, por tanto se alcanzan precisiones subpuntuales. En cuanto al radio, la precisiones están acotadas por 2 mm y la repetitividad es inferior al milímetro. El error detectado en el valor de reflectividad es menor siempre del 2.5%, ya que el peor caso se da con errores de 0.005 para un valor real de 0.2.

Cabe destacar que el valor máximo de indicios teóricamente alcanzable, 4, no lo estamos obteniendo utilizando la esfera libre de degradaciones, ya que la suma de indicios solamente alcanza 3.3. Este fenómeno se debe fundamentalmente al muestro espacial limitado que en este caso es de 3 mm y que como veremos posteriormente en el análisis de resolución, va mejorando paulatinamente al incrementar la resolución.

Las figuras 5.4a y 5.4b muestran respectivamente la influencia del *ruido impulsivo*, en imágenes con 3 mm de resolución espacial, sobre los indicios de esfericidad captados y sobre la precisión en la estimación de los parámetros de la esfera. En este caso podemos ver que existe un nivel de ruido a partir del cual los indicios de esfericidad se hacen nulos con lo cual no es posible detectar la esfera, y por tanto no se muestran los valores de error para un ruido del 50% puesto que en este caso no hay hipótesis generadas. El comportamiento altamente robusto de los algoritmos ante este tipo de ruido, proviene del filtrado *3σ-MPF* que como vimos en el capítulo 3, trabajaba correctamente ante

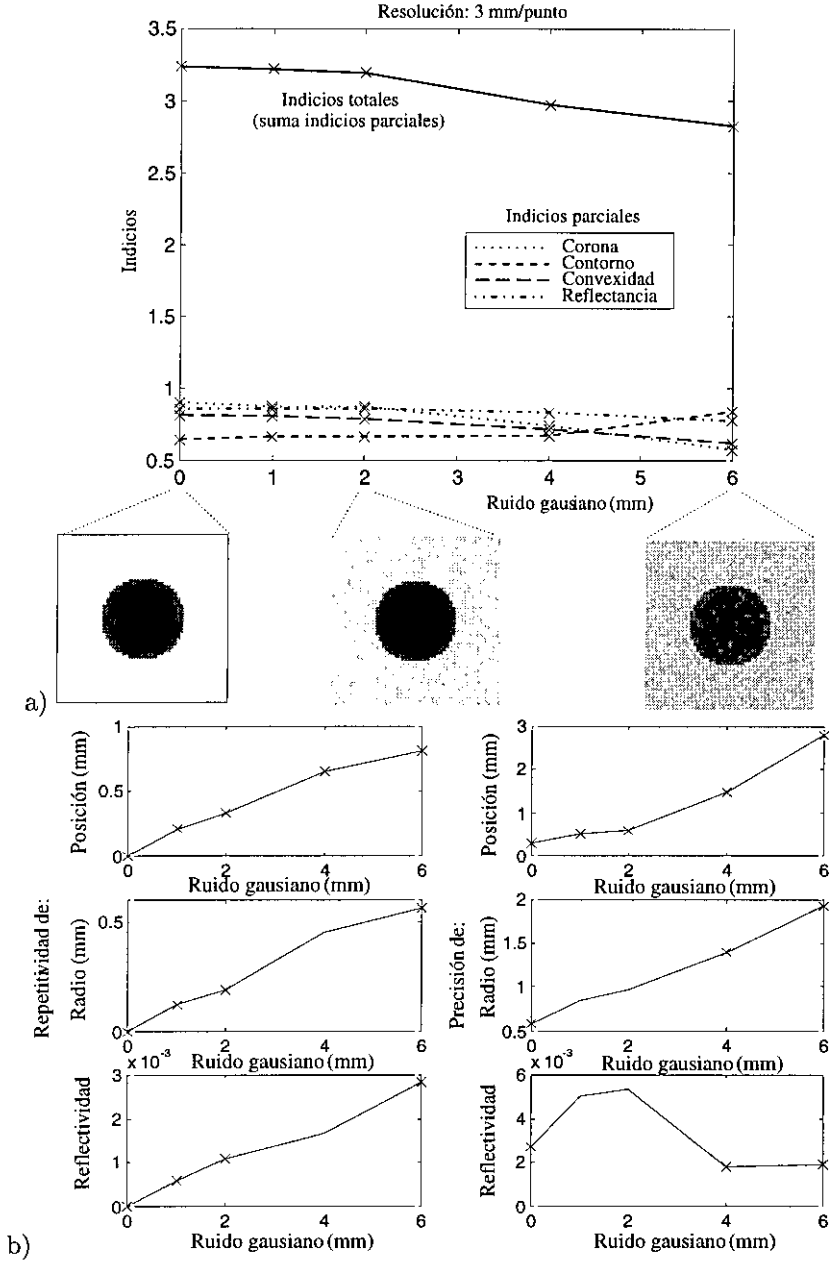


Figura 5.3: Influencia del nivel de ruido gausiano, en una imagen de distancias, sobre: a) el conjunto de indicios de esfericidad captados por los algoritmos propuestos; b) la repetitividad y precisión en la estimación de la posición tridimensional, el radio y la reflectividad media de la esfera detectada.

niveles significativos de ruido impulsional. La tendencia general al aumentar el ruido es una disminución progresiva del grado de evidencia y un aumento del mismo orden en los errores de estimación de parámetros. Podría parecer que existe una significativa sensibilidad a este tipo de degradación, sin embargo el análisis realizado ha sido muy exigente y en la mayoría de los casos este tipo de ruido no suele superar un 5% o 10%, con lo cual la influencia es casi despreciable.

En las figuras 5.5a y 5.5b, podemos ver los resultados de la evaluación cuando el *porcentaje de oclusión* varía, utilizando una imagen contaminada con ruido gausiano de desviación estándar 1 mm y resolución espacial de 3 mm. Como era de esperar, a medida que el área visible de la esfera disminuye, los indicios también lo hacen, alcanzándose un punto crítico correspondiente a una oclusión del 75% (25% visible) en el cual los indicios de esfericidad se hacen casi nulos y por tanto también su capacidad para detectar esferas. En cuanto a los errores, la tendencia general es la de aumentar al incrementar la oclusión. Los errores de posición están acotados en un rango de 4 mm y las discrepancias en la estimación del radio son menores que 1 mm, aunque para oclusiones próximas al 75% se pueden apreciar repetitividades próximas a 3 mm. En el caso de la estimación de la reflectividad, los errores son menores del 2.5% salvo en el caso de la oclusión límite, donde no se generan evidencias a partir de primitivas regionales y por tanto no se genera ningún valor de reflectividad para la esfera detectada, con lo cual el error es 0.2 (reflectividad de la esfera sintética creada).

Como puede apreciarse en las figuras 5.6a y 5.6b, la influencia que tiene la variación del *grado de esfericidad* respecto a un modelo ideal es significativa, aunque bastante limitada en un rango de 1 a 1.2. Los errores absolutos de posición quedan acotados por 4 mm, con repetitividades menores de 1 mm. Estos errores absolutos de posición que suponen desviación relativamente alta al compararla con sus repetitividades, se deben a errores acumulados en la coordenada de distancia de la posición tridimensional en coordenadas esféricas. Los algoritmos al asumir esferas ideales y encontrarse con elipsoides con diferentes radios de curvatura sobre su superficie, detectan el centro como una integración conjunta de evidencias sobre todo el elipsoide, cuando en realidad solo deberían utilizar los puntos del elipsoide con curvatura igual a la definida por el eje menor del elipsoide orientado según el eje de medida del sensor. En cuanto a los errores absolutos en el radio y la reflectividad, vemos que están limitados por 3 mm y un 4% del valor real de reflectividad, con repetitividades menores de 1 mm y del 10%, respectivamente.

La influencia que la *resolución* o el *grado de muestreo* de la imagen tiene sobre los resultados del algoritmo, puede apreciarse en las figuras 5.7a y 5.7b. Podemos ver, como indicamos antes, que un aumento en la resolución (muestreo mayor) hace que la esfera tienda a ser realmente perfecta y por tanto los indicios totales de esfericidad tienden a 4, que es el máximo número de indicios alcanzable con los algoritmos propuestos. Las primitivas de tipo regional no se ven afectadas por una menor resolución alcanzando valores de indicios parciales próximos a 1. Sin embargo y como es lógico las más sensibles son las hipótesis generadas a partir de primitivas puntuales, puesto que en estos casos el carácter local y puntual implícito deja de presentarse al trabajar con puntos de tamaño creciente, y por tanto los cálculos quedan muy influenciados por los puntos del vecindario que ya no se encuentran realmente próximos. Los errores absolutos en posición, radio y reflectividad son siempre menores de 2 mm, 1.5 mm y 1%; con lo cual vemos que la precisión es siempre menor que el tamaño del punto. Igualmente, las repetitividades son

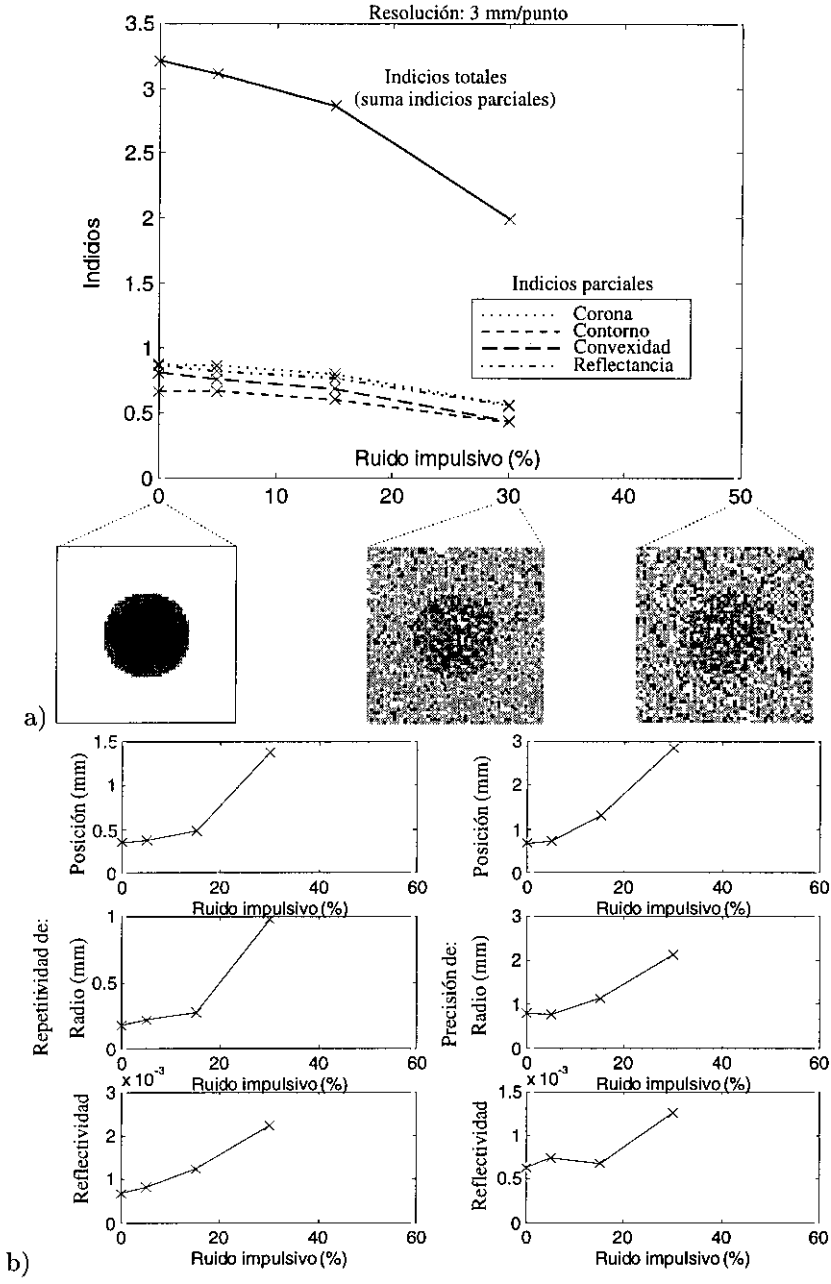


Figura 5.4: Influencia del nivel de ruido impulsional, en una imagen de distancias, sobre: (a) el conjunto de indicios de esfericidad captados por los algoritmos propuestos, (b) la repetitividad y precisión en la estimación de la posición tridimensional, el radio y la reflectividad media de la esfera detectada.

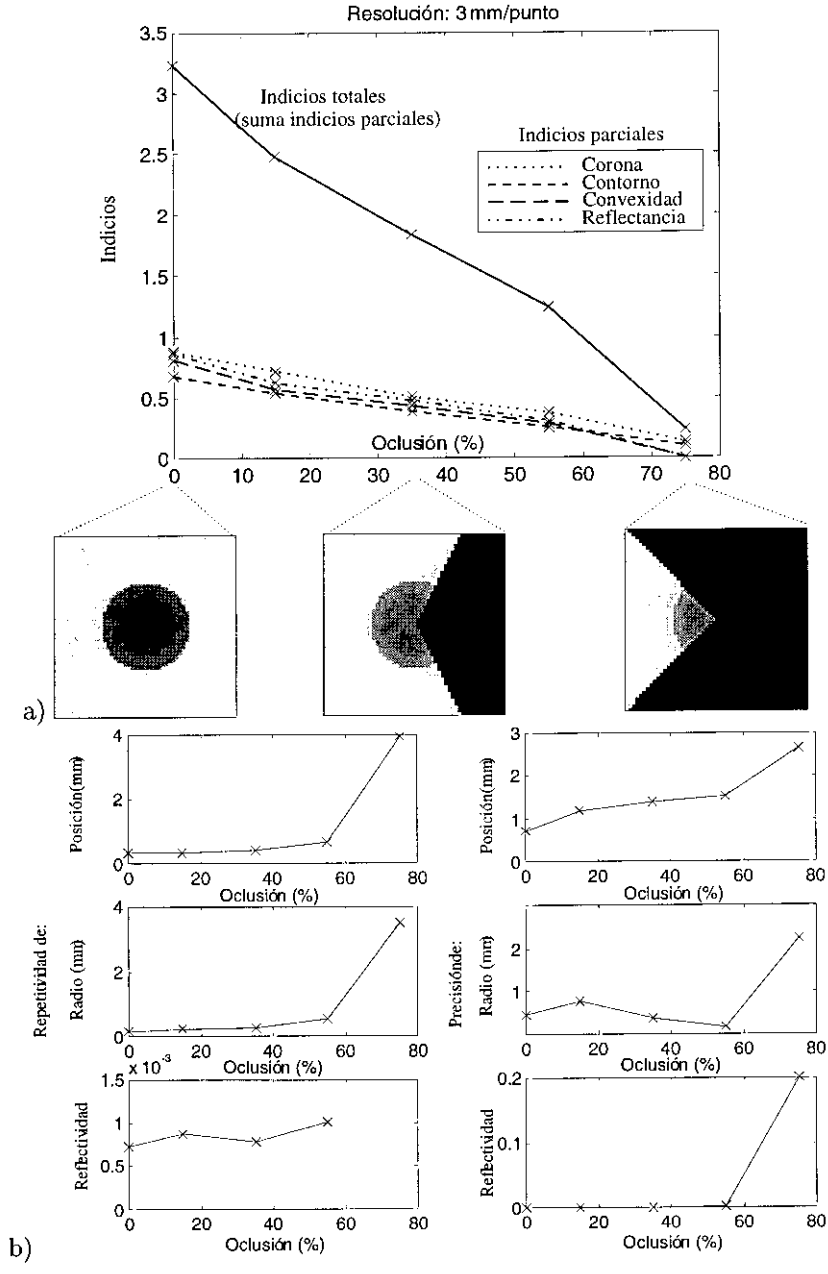


Figura 5.5: Influencia del grado de oclusión de una esfera sobre: (a) el conjunto de indicios de esfericidad captados por los algoritmos propuestos, (b) la repetitividad y precisión en la estimación de la posición tridimensional, el radio y la reflectividad media de la esfera detectada.

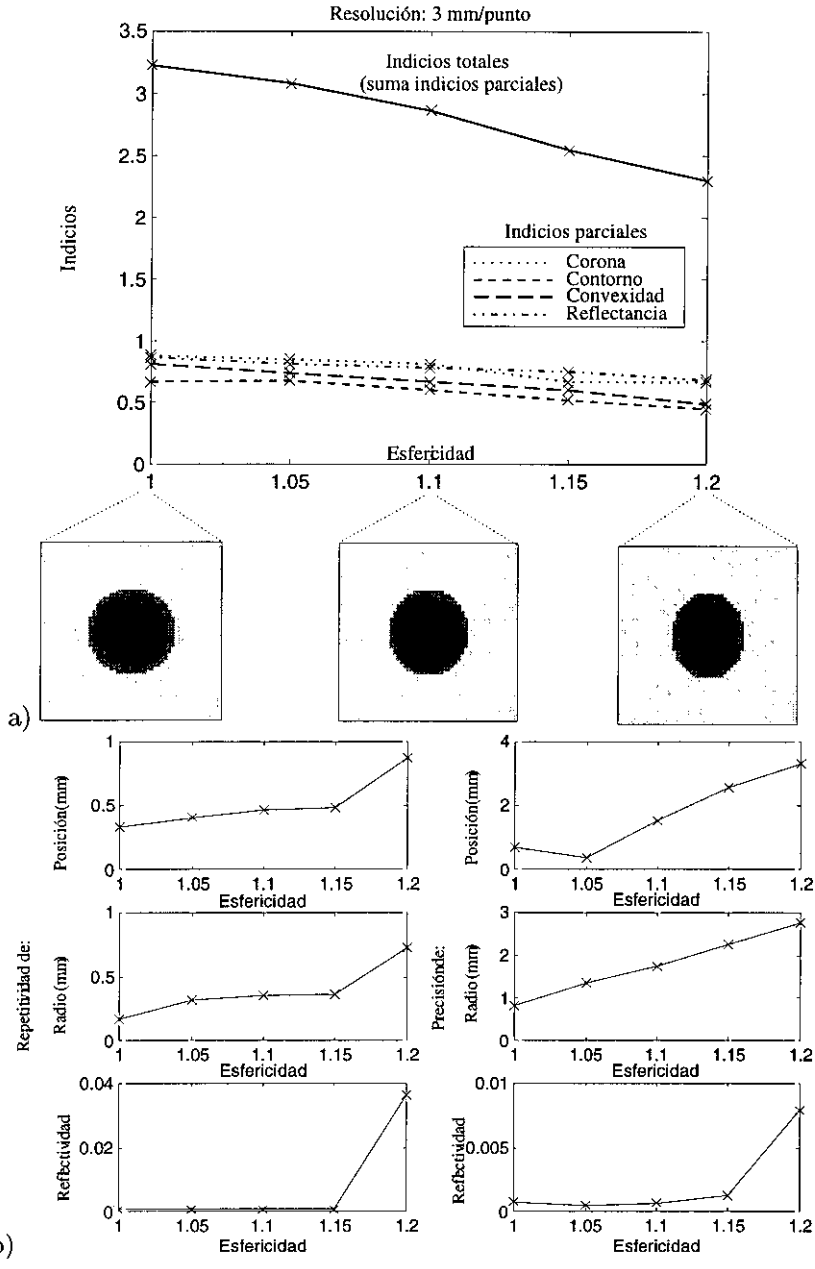


Figura 5.6: Influencia de la excentricidad de una esfera sobre: (a) el conjunto de indicios de esfericidad captados por los algoritmos propuestos, (b) la repetitividad y precisión en la estimación de la posición tridimensional, el radio y la reflectividad media de la esfera detectada.

menores que 1 mm, 0.5 mm y un 1%, para posiciones, radios y reflectividades.

Finalmente, la sensibilidad del algoritmo a la búsqueda de esferas de *radio diferente* al dado como referencia, se aprecia en las figuras 5.8a y 5.8b, las cuales contienen imágenes degradadas con ruido gausiano de $\sigma = 1$ mm y una resolución de 3 mm por punto. Como se vio en el capítulo anterior, los algoritmos fueron diseñados para penalizar aquellas esferas detectadas con radios diferentes que la referencia buscada. Este hecho, que es adecuado para restringir el ámbito de objetos a buscar, y por tanto para aumentar la eficiencia de los algoritmos y reducir la probabilidad de encontrar detecciones falsas, hace que los indicios de esfericidad captados se vean afectados, con lo cual disminuye paulatinamente la capacidad de detectar esferas a medida que estas aumentan o disminuyen de tamaño respecto al valor referencia. Vemos que la reducción de evidencia es menor del 50% para una diferencia relativa de radios de 0.3, lo cual posibilita la detección de esferas con una variabilidad en radio mayor del $\pm 30\%$, siendo esto suficientemente flexible para un amplio número de aplicaciones, entre las cuales también se encuentra la detección de frutos cuasi-esféricos cuya variabilidad en radio para una misma especie no suele ser mayor que la indicada [112].

En cuanto a los errores en la estimación de parámetros, estos son mínimos y no varían apreciablemente al diferir los radios de referencia y real. Es decir, el grado de evidencia de esfericidad disminuye pero no lo hace la fiabilidad de los parámetros estimados lo cual es muy interesante. Los errores absolutos son menores de 2 mm en posición, 2 mm en radio y 0.5% en reflectividad, con unas repetitividades menores de 1 mm, 0.5 mm y 0.5%.

La evaluación simulada que acabamos de hacer, nos da una idea clara de los límites que presenta el algoritmo al modificarse las características de las imágenes de entrada y al variar el radio de referencia pasado como parámetro al algoritmo. Hemos podido ver que los errores en la estimación de los parámetros de la esfera están casi siempre limitados por el muestreo de la imagen, posibilitando la utilización de estas estimaciones en aplicaciones donde no se requieran altas precisiones y por el contrario sea más importante mantener la capacidad de reconocimiento de esferas utilizando imágenes de reducido tamaño, favoreciendo así la aceleración de su captura y procesamiento.

También hemos visto que los indicios totales disminuyen al aumentar cualquiera de las degradaciones, sin embargo esta disminución del grado de evidencia no está linealmente relacionada con una reducción en la capacidad de detección de esferas. Los indicios totales son simplemente una medida del grado de certeza que tenemos en la formulación de una hipótesis sobre la existencia de un objeto esférico. Sobre estos indicios se aplican umbrales para finalmente generar las hipótesis globales válidas. Este umbral se deberá elegir para que las detecciones falsas se minimicen y las detecciones correctas sean máximas. Un valor típico del umbral es 0.1, lo cual indica que todas aquellas hipótesis con indicios totales entre 0.1 y 4 son aceptadas como esferas. Esto hace que la sensibilidad mostrada por los algoritmos ante las degradaciones introducidas, no se refleje directamente en la capacidad de detección de esferas, aunque sí en el grado de confianza con la que dicha afirmación se hace.

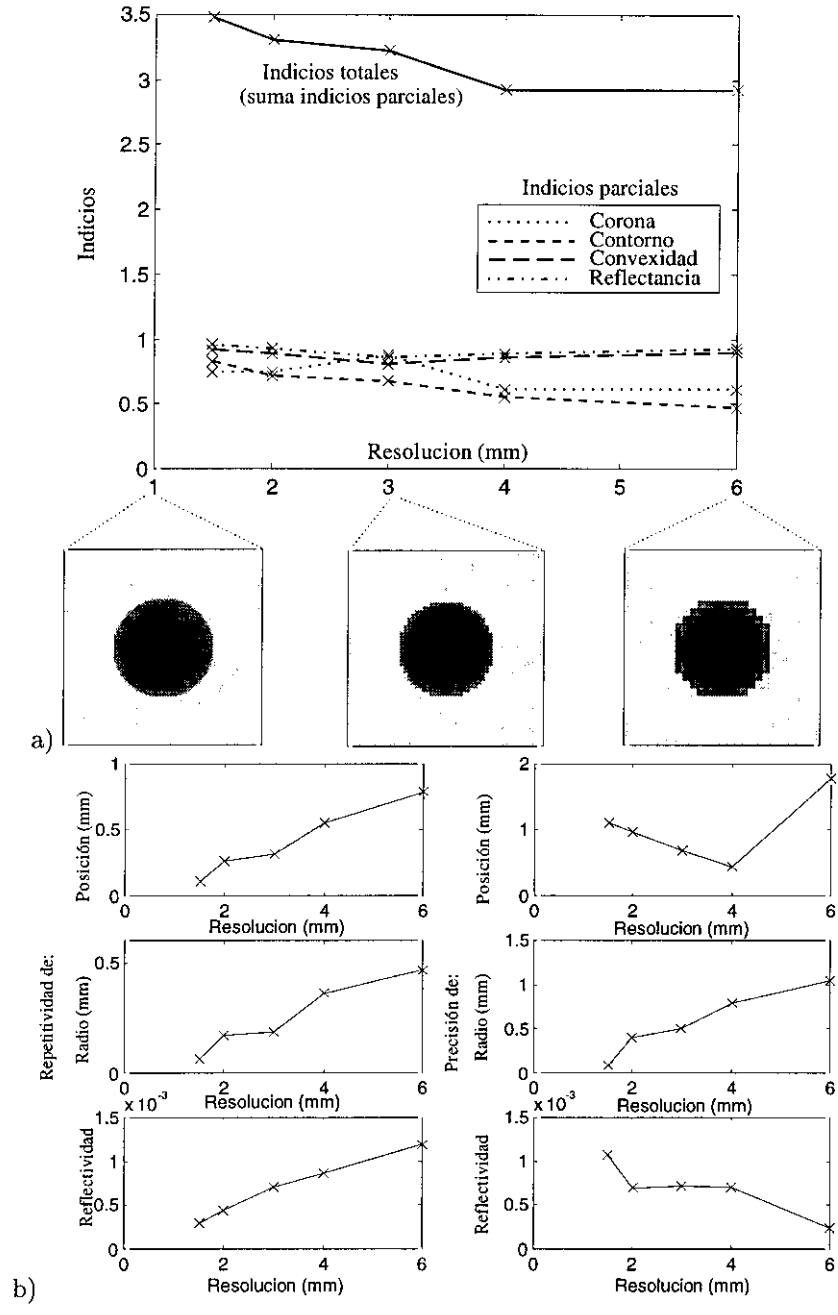


Figura 5.7: Influencia de la resolución en el muestreo sobre: (a) el conjunto de indicios de esfericidad captados por los algoritmos propuestos, (b) la repetitividad y precisión en la estimación de la posición tridimensional, el radio y la reflectividad media de la esfera detectada.

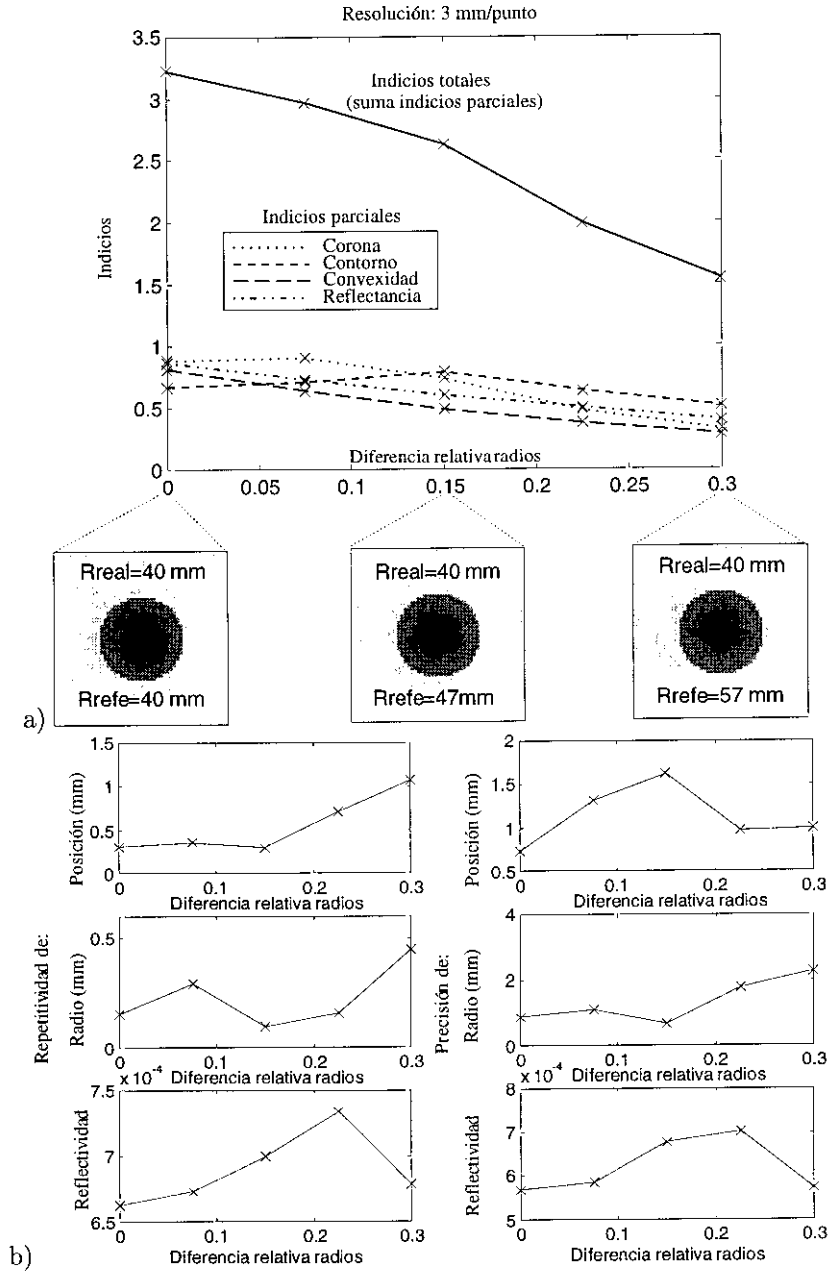


Figura 5.8: Influencia de la diferencia relativa entre el radio real de una esfera y el radio referencia introducido al algoritmo, sobre: (a) el conjunto de indicios de esfericidad captados por los algoritmos propuestos (b) la repetitividad y precisión en la estimación de la posición tridimensional, el radio y la reflectividad media de la esfera detectada.

5.2.3 Influencia de la restauración sobre los resultados

En el capítulo 3 presentamos el nuevo método de restauración que denominamos *3σ-MPF* y su evaluación utilizando la métrica *GRI*. Esta evaluación comparaba los resultados generados por el filtro con la referencia ideal libre de ruido. La medida de calidad de restauración captada por la métrica *GRI* valoraba positivamente aquellos filtros que proporcionaban reconstrucciones fieles, minimizando el error absoluto, y a la vez generaban superficies continuas con bajos niveles de rugosidad. Ante esta métrica, el filtro *3σ-MPF* resultó ser mejor que el resto de filtros con los que se realizó la comparación.

La cuestión no presentada hasta el momento, y que quedaba por clarificar, como quedó indicado en el capítulo 3, consiste en ver si la utilización de este filtro supone alguna mejora, desde un punto de vista práctico, a la hora de facilitar la labor a las siguientes etapas de reconocimiento y localización; o por el contrario, la utilización de cualquier otro filtro también permite obtener resultados similares. En este último caso la utilización del filtro *3σ-MPF* no tendría ninguna utilidad de tipo práctico, al menos para la configuración algorítmica que hemos presentado. Lo que cabe esperar, teniendo en cuenta que los algoritmos se basan en modelos de esferas ideales, sin rugosidad superficial y sin distorsiones, y que los algoritmos calculan magnitudes tales como gradientes, pendientes y curvaturas, las cuales son muy sensibles a ruidos y deformaciones, es que la precisión en los resultados y la capacidad de detección de objetos se incremente utilizando un filtro como el *3σ-MPF*.

Para comprobar este hecho, vamos a analizar de nuevo los indicios generados y los errores absolutos y repetitividades cuando variamos el nivel de ruido gausiano sobre la imagen sintética utilizada anteriormente. A diferencia de lo mostrado en la subsección anterior, en este caso cambiaremos el tipo de filtrado utilizado. Los filtros que vamos a probar son tres: nuestro filtro *3σ-MPF*, el filtro óptimo de Wiener o MMSE y el filtro de orden de la mediana. Además veremos lo que sucede cuando eliminamos por completo la etapa de filtrado. En la figura 5.9 podemos ver cómo los máximos indicios se alcanzan utilizando nuestro filtro, seguido muy de cerca por el filtro de Wiener y por el filtro de la mediana, este último a más de 0.3 indicios de distancia para un ruido de 3 mm. En el caso de no utilizar ningún filtro, se aprecia que a partir de pequeños niveles de ruido la degradación es tal que la capacidad de detección de esferas se degrada fuertemente, con lo cual se justifica claramente la necesidad de utilizar una etapa de restauración.

Debe hacerse notar, que los resultados obtenidos con los filtros *MAS* y *DW-MTM*, que presentaban altos índices *GRI* en el estudio del capítulo 3, no se han presentado por obtenerse resultados sensiblemente peores que en el caso del filtro de Wiener, con lo cual este último fue elegido como representante de la familia de los mejores métodos de filtrado, incluyéndose el filtro de la mediana como representante de los filtros con resultados intermedios.

El hecho de que un filtro óptimo como el de Wiener ó MMSE, se vea superado por un filtro no óptimo como es el nuestro, no debe resultar extraño. El filtro de Wiener es considerado óptimo al filtrar una señal contaminada con ruido gausiano, ya que minimiza el error cuadrático medio, pero solamente es óptimo en este sentido. Sin embargo, alcanzar este mínimo no significa que la señal reconstruida vaya a ser más adecuada que otra para obtener un determinado fin. En mucho casos, como en el nuestro, una señal con error cuadrático mínimo no es la única condición recomendable para obtener buenos

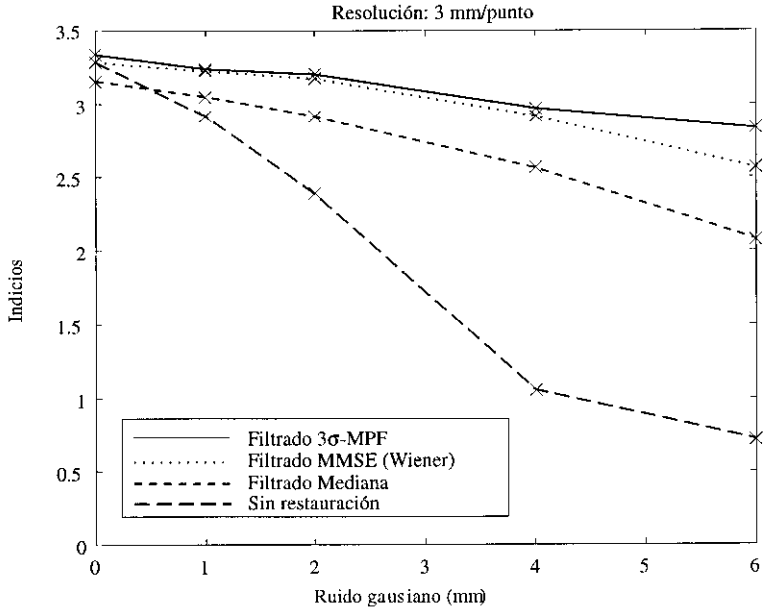


Figura 5.9: Indicios generados al variar el ruido gausiano y utilizando diferentes tipos de restauración.

resultados. Cualquier algoritmo que necesite realizar derivadas primeras y segundas, va a apreciar también la ausencia de rugosidades sobre las superficies que en el caso ideal corresponderían a áreas continuas y suavizadas. El filtro *3σ-MPF* sin ser óptimo, ni minimizando errores cuadráticos ni consiguiendo la máxima suavidad, sí obtiene soluciones muy próximas a las óptimas en ambos aspectos, y esto es lo que provoca que se obtengan mejores soluciones que en el caso óptimo del filtro de Wiener. Además la capacidad multiresolución del filtro *3σ-MPF* favorece que en muchos casos se obtengan errores cuadráticos incluso menores que en el caso de Wiener ya que este trabaja con ventanas fijas y el anterior utiliza ventanas mayores en regiones continuas con lo cual puede llegar a generar errores absolutos menores que en el diseño óptimo, como se pudo comprobar analizando la componente *FI* del índice *GRI* (tabla 3.2).

Analizando los resultados obtenidos referentes a los errores cometidos en la estimación de los parámetros: posición tridimensional, radio y reflectividad media (fig. 5.10), podemos observar que también se obtiene una mejoría clara en los errores absolutos frente a las estimaciones obtenidas utilizando los otros métodos de restauración. Los errores máximos de posición son de 2 mm utilizando nuestro filtro, mientras que para el filtro Wiener y mediana los errores máximos son próximos a 5 mm. En la estimación del radio, con el filtro *3σ-MPF* estamos siempre por debajo de los 2 mm de error mientras que en los otros casos los errores máximos se aproximan a 3 mm. En la estimación de la reflectividad, las diferencias son aún mas acusadas con errores del 2% con nuestro filtro y errores máximos superiores al 10% en los otros casos. En el apartado de repetitividades no hay apenas diferencias entre los tres tipos de filtrado y la única desviación negativa se debe al caso

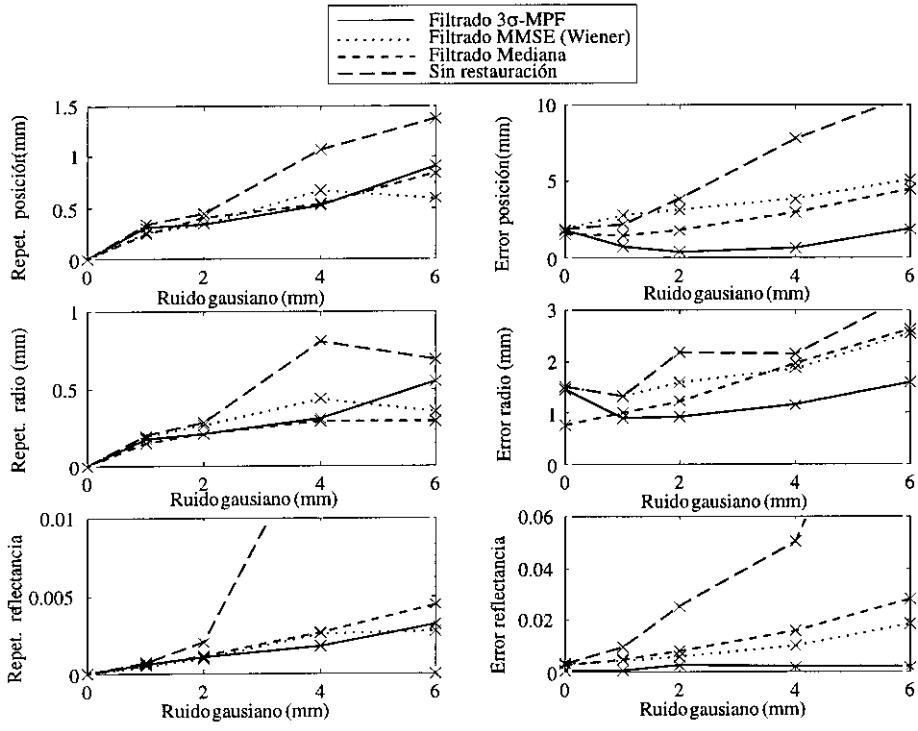


Figura 5.10: Errores en la estimación de parámetros, al variar el ruido gausiano y utilizando diferentes tipos de restauración.

en el que no se aplica filtrado.

5.3 Evaluación de los algoritmos propuestos utilizando datos empíricos

Hasta ahora hemos analizado el comportamiento de los algoritmos diseñados utilizando imágenes generadas sintéticamente e introduciendo diferentes factores degradantes con la intención de ver de forma aislada la sensibilidad de los algoritmos a cada tipo de degradación. En las escenas reales todos los factores degradantes, tanto los presentados en la sección anterior como otros susceptibles de manifestarse y no modelados hasta ahora, influyen, en mayor o menor grado, en la calidad de las imágenes. Para ver este efecto degradante de forma conjunta vamos a realizar dos tipos de análisis con datos empíricos. El primer análisis se basa en ensayos realizados en el laboratorio con escenas artificiales dispuestas para que se aproximen en el mayor grado posible a una escena natural de un árbol frutal conteniendo naranjas. El segundo análisis se realizó en una plantación natural tratando de captar otros factores degradantes que no se pudieron reproducir en la configuración de laboratorio, y que nos permitirá deducir cuales son

las principales limitaciones a la hora de aplicar la estrategia presentada en esta tesis en situaciones prácticas. Los siguientes apartados presentan ambos análisis.

5.3.1 Evaluación de la capacidad de detección con imágenes de laboratorio

En una evaluación empírica, y a diferencia de la evaluación simulada, las imágenes o datos experimentales no están apenas controlados, es decir, los parámetros de degradación que afectan a dichas imágenes son los propios de la escena y los introducidos por el sensor. La posición de los objetos y sus radios o reflectividades, en principio son desconocidos, o al menos, conocidos sin gran precisión. Por tanto en este caso, los *criterios de evaluación* elegidos no van a ser errores de estimación de parámetros, sino los porcentajes de detecciones correctas y falsas, lo cual sí puede ser determinado con absoluta certeza a partir de unas imágenes reales. En concreto por *detecciones correctas* entendemos la relación entre el número de esferas detectadas correctamente y el número de esferas visibles presentes en la escena. Igualmente, por *detecciones falsas* se entiende la relación entre el número de esferas detectadas, que realmente no lo son, y el número total de esferas detectadas, incluyendo tanto las correctas como las falsas.

En cuanto a los *datos de entrada*, solamente vamos a considerar la *resolución espacial* de las imágenes, ya que sobre este parámetro se puede actuar sin más que cambiar el muestreo en la captación de los puntos que forman la imagen. En esta evaluación se utilizó un conjunto de 45 imágenes de laboratorio captadas a diferentes resoluciones conteniendo escenas de un árbol artificial con un total de 117 frutos de plástico, con colores tanto rojizos como verdes, que se dispusieron de forma aleatoria y con diferentes grados de oclusión sobre la periferia e interior del árbol. Una imagen del entorno de trabajo y del árbol artificial utilizado se vé en la figura 5.11. Un subconjunto de las imágenes totales utilizadas se muestra en el apéndice C y un ejemplo particular de una escena se adelanta en la figura 5.12. Las hipótesis finales generadas por los algoritmos, se indican marcando los objetos esféricos detectados mediante un anillo circular centrado sobre la ubicación estimada de los frutos. Los indicios generados correspondientes a las primitivas corona, contorno, convexa y reflectividad, se incluyen en este orden en cada detección presentada quedando etiquetadas como I1, I2, I3 e I4 respectivamente.

El parámetro configurable del algoritmo será el *umbral de indicios*. Sobre su cuantificación no se ha dado ningún criterio de selección automática, y por tanto este estudio permitirá analizar el valor óptimo. El umbral apropiado será aquel con el que se consigan las máximas tasas de detección correctas y al mismo tiempo las detecciones falsas se mantengan mínimas o nulas.

Los resultados obtenidos en esta evaluación experimental, utilizando resoluciones de 2, 3 y 4 milímetros por punto, con umbrales de evidencia entre 0.05 y 0.2, pueden verse en las gráficas 5.13, 5.14 y 5.15. La primera conclusión deducible es que la probabilidad de obtener detecciones falsas crece al disminuir el umbral de evidencia, como es lógico esperar, y también aumenta al utilizar resoluciones menores. En este último caso las regiones superficiales captadas son muy pequeñas y por tanto la estabilidad en las medidas decrece. Por tanto, si lo que nos interesa es trabajar con cierta seguridad de no encontrar detecciones falsas un umbral de 0.15 es el más recomendable, como se puede inferir de

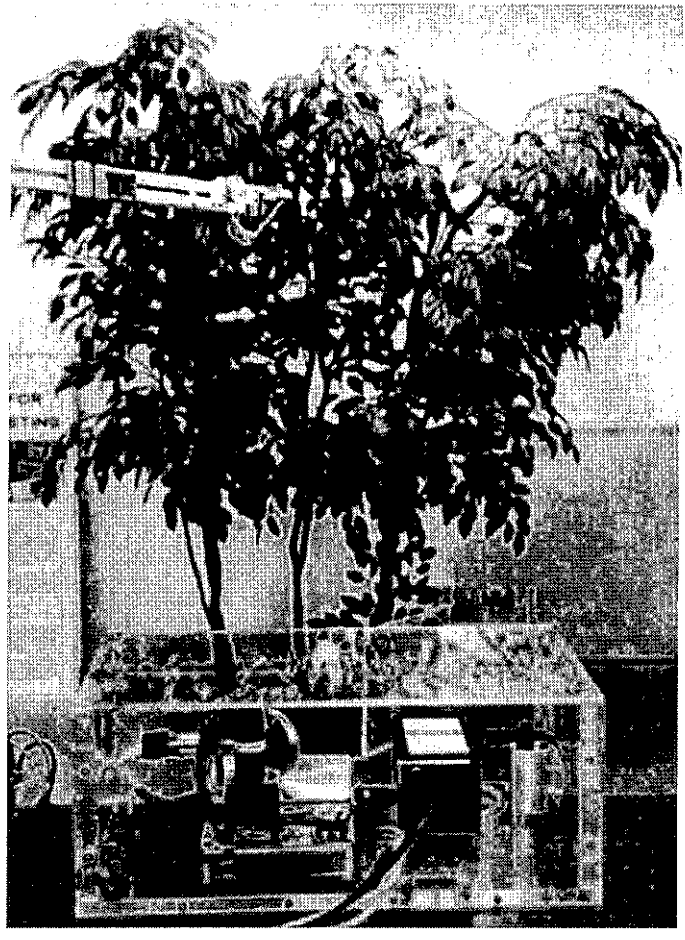


Figura 5.11: Entorno de trabajo donde se muestra el árbol artificial utilizado y el sistema de deflexión láser situado a unos 2-3 metros de este. Las imágenes obtenidas se utilizaron para realizar la evaluación empírica de laboratorio.

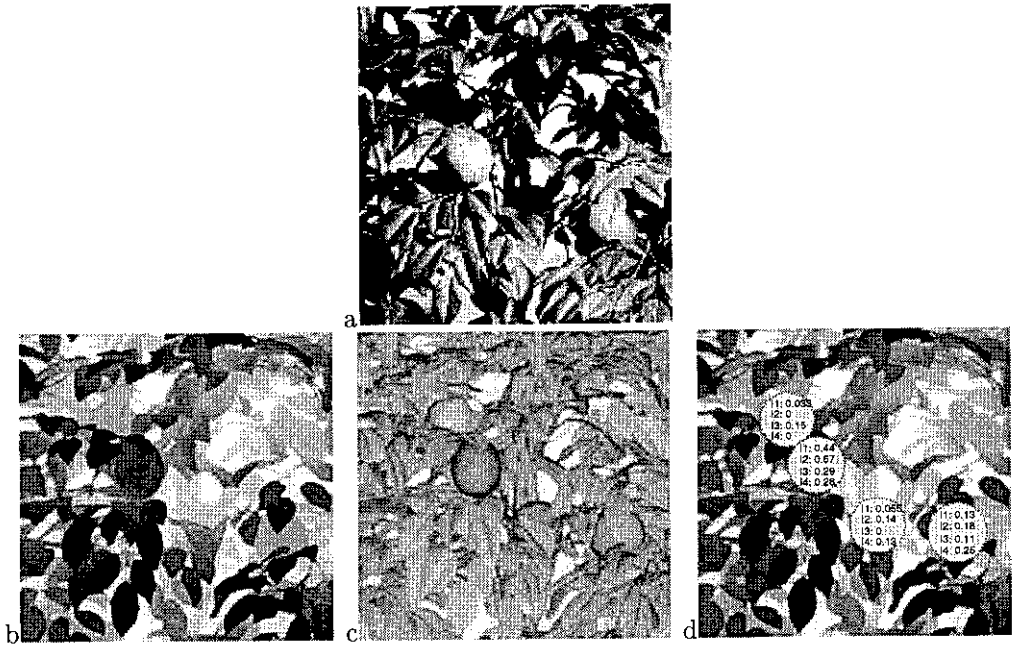


Figura 5.12: Ejemplo de una de las imágenes capturadas en el laboratorio y de las detecciones encontradas. (a) fotografía de la escena, (b) imagen de distancia, (c) imagen de reflectancia y (d) detecciones encontradas superpuestas a la imagen de distancias e indicios que permitieron la generación de estas hipótesis.

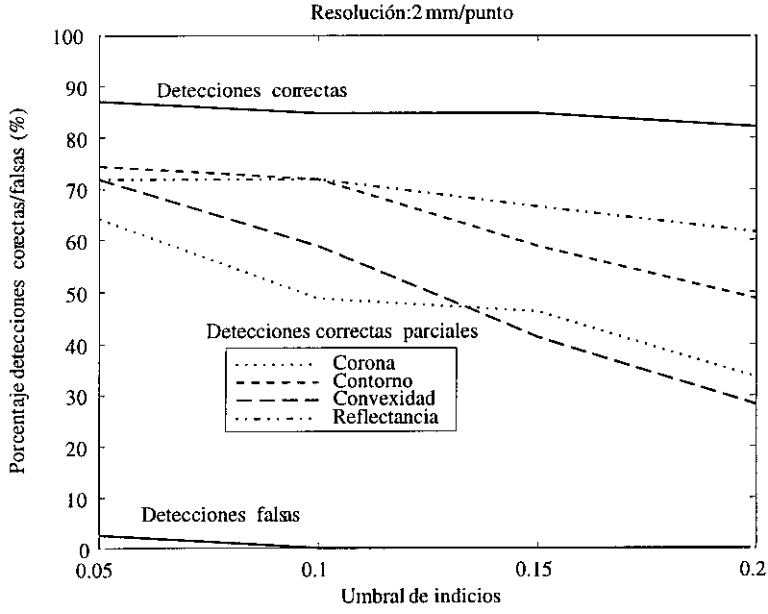


Figura 5.13: Tasa de detecciones correctas y falsas en los ensayos de laboratorio para imágenes con una resolución de 2 mm/punto.

esta evaluación. Con este umbral (0.15) las detecciones falsas son improbables y la tasa de detecciones correctas ronda el 85%, 80% y 72.5% para el conjunto de imágenes con resolución de 2, 3 y 4 milímetros por pixel, respectivamente. Evidentemente, la tasa de detecciones correctas también depende del umbral elegido y aumenta al disminuir este, pudiéndose alcanzar detecciones próximas al 90% utilizando un umbral de 0.05. Sin embargo, esto se consigue a consta de correr el riesgo de aumentar el número de detecciones falsas.

Igualmente en las gráficas 5.13, 5.14 y 5.15 podemos apreciar, a través de las líneas punteadas, el porcentaje de detecciones correctas que se obtendrían en el caso de utilizar solamente uno de los métodos de reconocimiento, es decir, el basado en primitivas contorno, corona, convexas o reflectividad. Podemos ver que no existe ningún método que predomine sobre los demás, con lo que podría dejar de tener sentido aplicar el resto de los métodos. Vemos que los cuatro métodos por aislado contribuyen de una forma moderada a la generación final de hipótesis correctas, rondando entre un 70% y 40%. Sin embargo la integración de las hipótesis generadas por todos, permite obtener una clara mejoría en las detecciones de fruta. Por tanto, se deduce que existe una clara complementariedad y cooperación entre los cuatro métodos aplicados.

Igualmente nos podríamos preguntar si alguno de los métodos es prescindible. Trabajando con tres primitivas a la vez y desactivando la otra, los resultados de detecciones correctas son ligeramente inferiores al caso en que los cuatro métodos están activos. Esta disminución indica que el cuarto método puede tener en determinadas circunstancias influencia positiva. Como ejemplo de que no hay ningún método

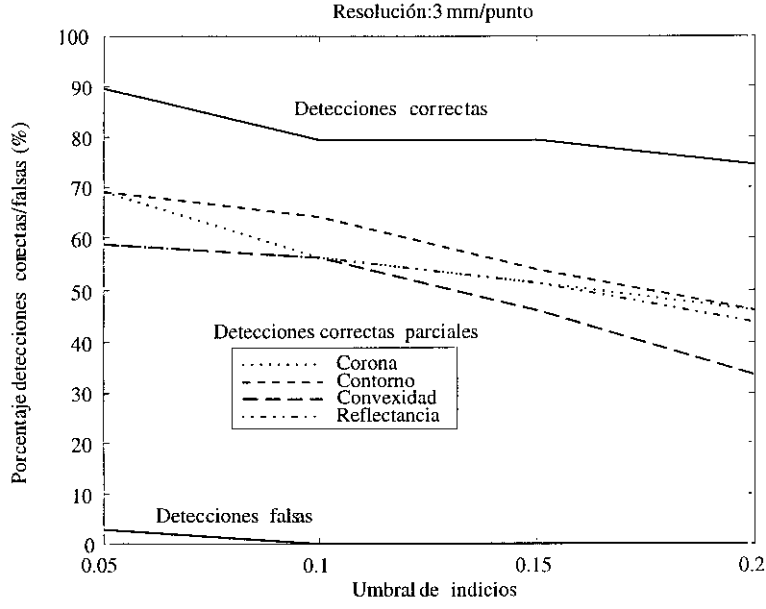


Figura 5.14: Tasa de detecciones correctas y falsas en los ensayos de laboratorio para imágenes con una resolución de 3 mm/punto.

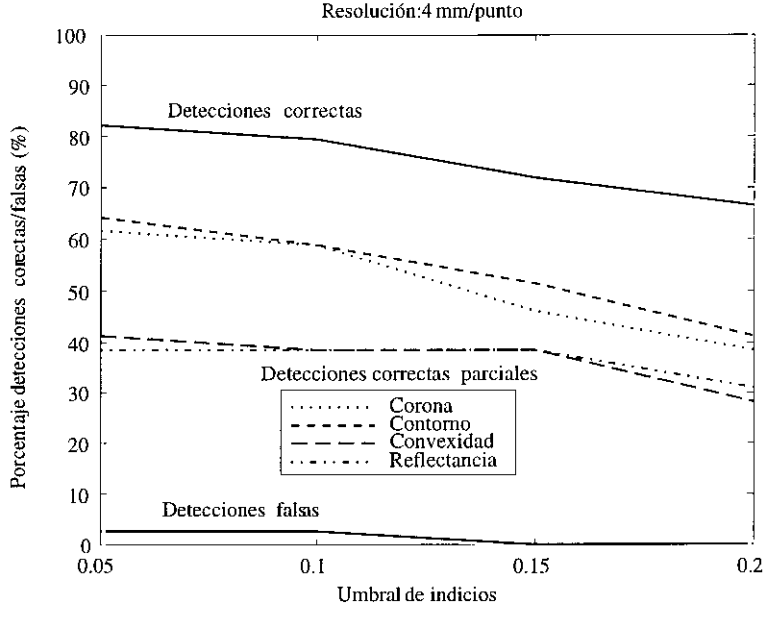


Figura 5.15: Tasa de detecciones correctas y falsas en los ensayos de laboratorio para imágenes con una resolución de 4 mm/punto.

prescindible y que todos aportan indicios en mayor o menor grado para llegar a la detección correcta, en la figura 5.16 se muestran casos en los que existen detecciones correctas obtenidas gracias a la activación de solamente uno de los cuatro métodos, con lo cual trabajar solo con las otras tres etapas provocaría una detección correcta menos.

El hecho de que no haya ningún método introducido gratuitamente era de esperar teniendo en cuenta que las primitivas puntuales y regionales fueron diseñadas para captar propiedades diferentes de una esfera. Las primitivas corona permiten detectar esferas cuando no son visibles los bordes ni el centro de la esfera pero sí una fracción de su corona. Las primitivas contorno se centran en los bordes esféricos con lo cual una oclusión de la mayor parte del área de la esfera, apareciendo visible solo parte del contorno esférico, permite la detección. Las primitivas regionales tienen sentido cuando tanto los contornos como las coronas están prácticamente ocultas y solo es visible la superficie central de la esfera. En el caso de esferas detectables por reflectividad (esfera y fondo con diferentes reflectividades) y si es visible una región significativa de su superficie, el método que opera utilizando primitivas reflectividad generará hipótesis correctas. En el caso de que no haya separabilidad por reflectividad, serán las primitivas convexas las que darán los indicios adecuados.

Considerando el modo de percepción mixto propuesto en el capítulo 2, los resultados mostrados ponen de manifiesto que seleccionando un modo de trabajo con captura de imágenes de 3 mm de resolución y un umbral de indicios de 0.15, el porcentaje de detecciones en el modo automático sería del 80% de la fruta visible, con lo cual solo un 20% de los frutos visibles quedarían en el árbol, pudiendo ser recogidos mediante el modo de señalización asistida, lo que permitiría disminuir la actividad del operador en cuatro quintas partes.

En cuanto a la generalidad de las soluciones propuestas, queremos subrayar que la metodología presentada en esta tesis, aunque está enfocada a solucionar un problema particular como la detección y localización de fruta, ha sido planteada de una forma genérica con lo cual puede aplicarse en la detección de objetos cuasi-esféricos en cualquier tipo de entornos. Como ejemplo ilustrativo mostramos en las figuras 5.17 y 5.18 dos imágenes típicas utilizadas en visión artificial compuestas de cubos, esferas, cilindros y fondos planos. Como puede verse, los objetos esféricos son detectados perfectamente a pesar de que existen oclusiones importantes.

5.3.2 Pruebas de detección en una plantación natural

Este segundo estudio con datos empíricos, a diferencia del anterior, utiliza imágenes captadas en entornos naturales. Se pretende detectar posibles factores degradantes importantes no modelados correctamente en el entorno artificial de laboratorio que se creó. No se ha hecho un estudio exhaustivo de la capacidad de detección de los algoritmos en entornos reales, cosa que se deberá hacer en el futuro y que no debería diferir mucho de los resultados presentados en la subsección anterior, una vez que los factores degradantes adicionales detectados en condiciones naturales que afectan a la calidad de las imágenes sean solucionados.

El estudio se realizó en Valencia en una plantación de naranjos de la variedad "navelate" en pleno estado de madurez. En la figura 5.19a se muestran imágenes del entorno en el

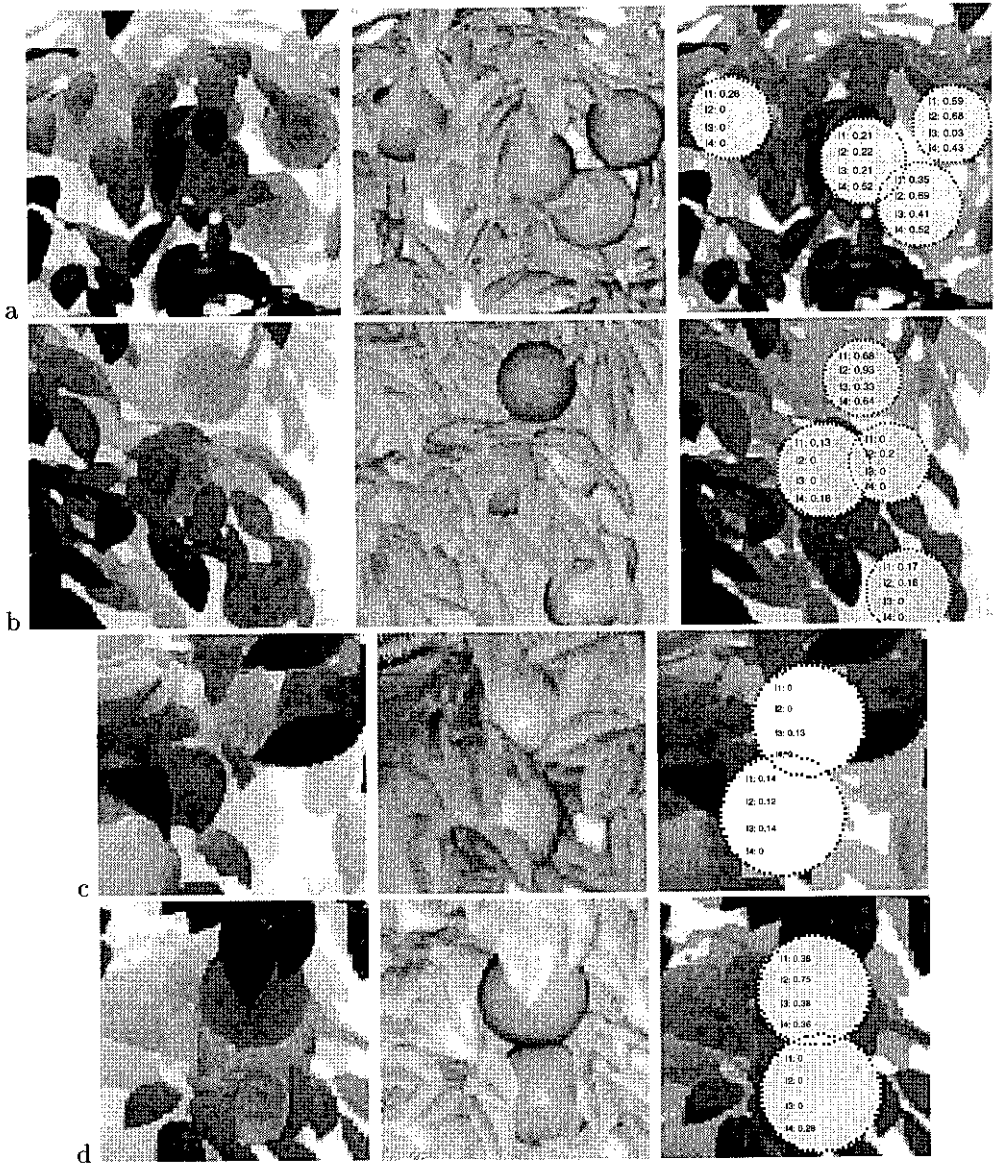


Figura 5.16: Ejemplos donde se ve que ninguno de los cuatro métodos de detección es prescindible.(a) la naranja de la izquierda es detectada por indicios corona, (b) la naranja del centro a la derecha solo se detecta por indicios contorno, (c) la hipótesis superior se genera a partir de indicios de convexidad, (d) la detección inferior se consigue por indicios de reflectividad.

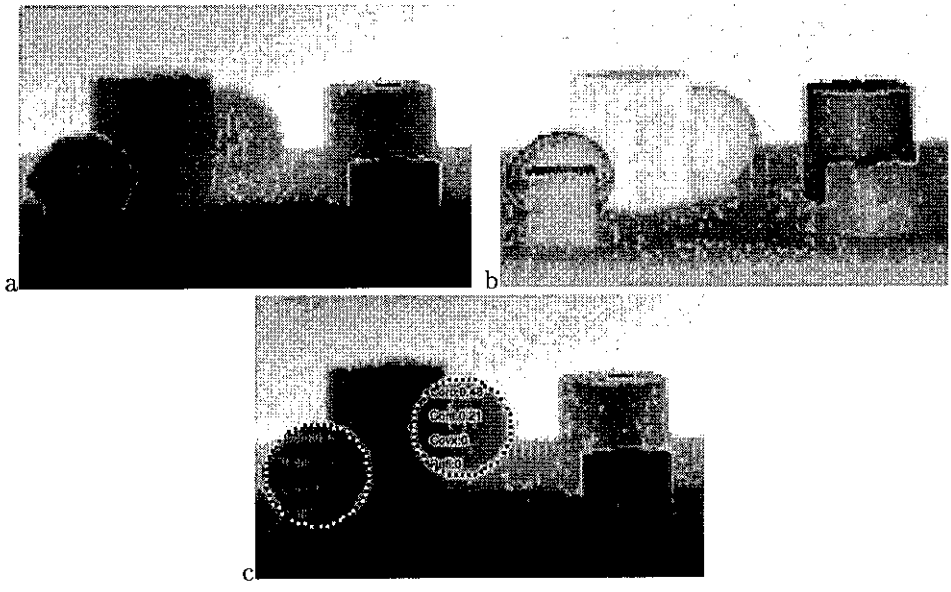


Figura 5.17: Ejemplo de detección de esferas en una escena conteniendo objetos estándar. (a) y (b) imagen de distancia y reflectancia respectivamente, (c) objetos esféricos encontrados.

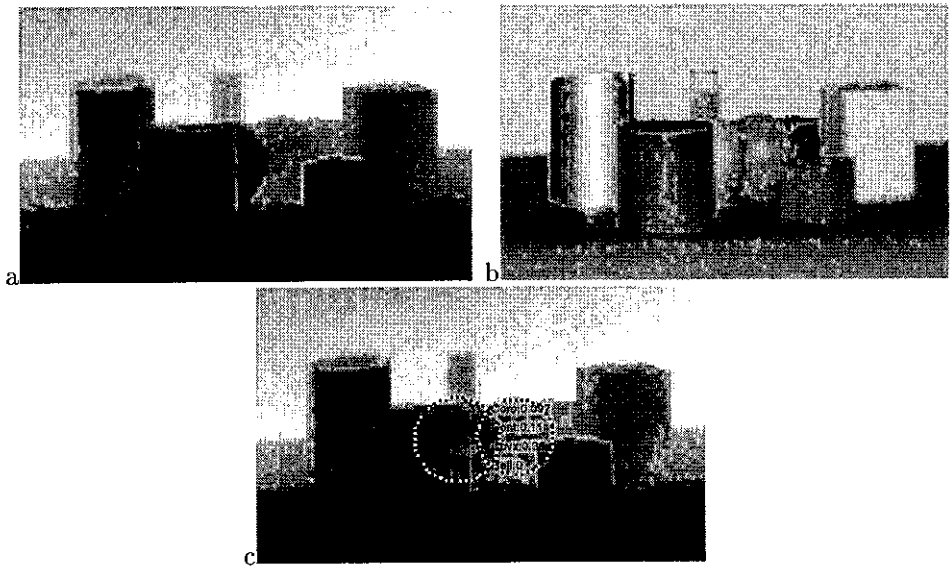


Figura 5.18: Ejemplo de detección de esferas en una escena conteniendo objetos estándar. (a) y (b) imagen de distancia y reflectancia respectivamente, (c) objetos esféricos encontrados.

Figura 5.19: Montaje y condiciones de trabajo en las pruebas de campo realizadas.

que se realizaron las pruebas, el equipo del robot Agribot utilizado (todos los módulos excepto el propio brazo recolector) y un detalle del sistema de deflexión láser captando una escena natural (fig. 5.19b). El resultado de uno de los procesos de barrido se muestra en la figura 5.20 donde presentamos una fotografía en detalle de la zona explorada, la pareja de imágenes distancia/reflectancia obtenida y los resultados de ese proceso de detección.

Los factores degradantes detectados en condiciones naturales, que se suman a los ya vistos en los estudios simulados y artificiales, son fundamentalmente dos: 1) la iluminación intensa del Sol, y 2) el viento que mueve los frutos. Estos factores perturbadores, que no se han tratado en ningún estudio hasta la fecha, los vamos a analizar a continuación, así como también analizaremos la capacidad de discriminación entre las frutas y el fondo utilizando información de reflectividad centrada a una longitud de onda de 780 nm.

5.3.2.1 Discriminación por reflectividad

Tanto en el capítulo 4, cuando presentábamos las primitivas regionales reflectividad, como en la subsección anterior veíamos que en muchos casos la fruta se podía discriminar del fondo utilizando información de reflectividad. Este estudio se realizó en ambos casos con el montaje artificial y por tanto utilizando fruta de plástico y hojas de tela. Veíamos que la reflectividad de las hojas era mayor que la de la fruta (0.407 frente a 0.254), y que esta última tenía una cierta componente difusa de tipo especular que se podían apreciar como picos de reflectancia en el centro de cada fruto.

La cuestión que nos planteamos ahora es ver si la capacidad de discriminación por reflectividad se sigue conservando en los entornos naturales. En la figura 5.21c presentamos una imagen de reflectividad que se calculó a partir de las imágenes de

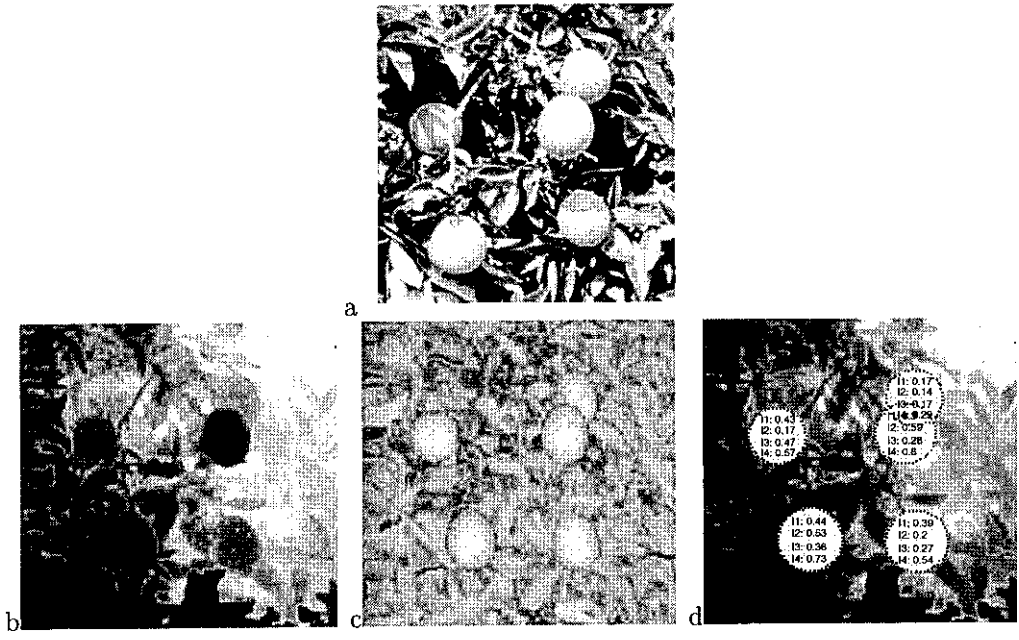


Figura 5.20: Imagen natural captada y procesada correspondiente a la escena visible en la figura 5.19b.

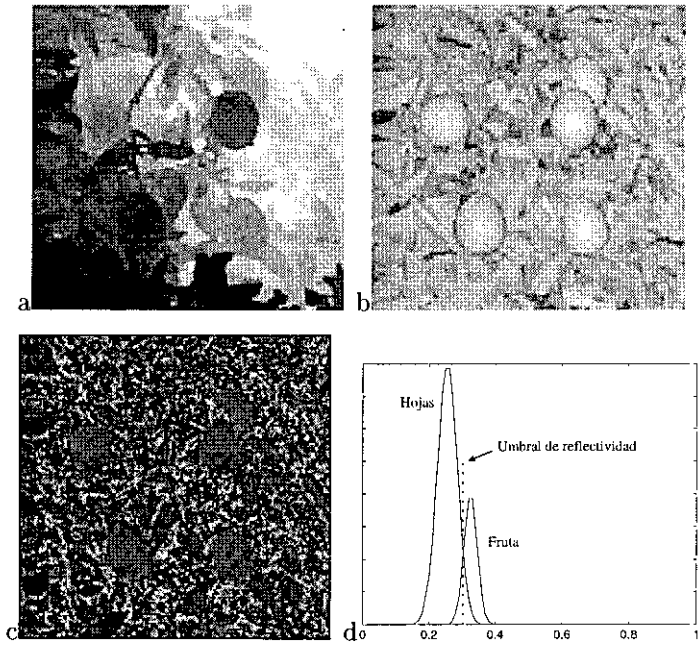


Figura 5.21: (a) y (b) Imágenes de distancia y reflectancia de una escena natural para el cálculo de reflectividades de fondo y fruta. (c) Imagen de reflectividad ($\lambda = 780$) calculada a partir de las dos anteriores. (d) Funciones de densidad de probabilidad de la fruta y las hojas.

distancia y reflectancia (figs.5.21a y 5.21b). Debido a que una mayor reflectividad se representa con un nivel de gris más claro, podemos apreciar visualmente en esta figura que las regiones correspondientes a las cinco naranjas tienen una reflectividad ligeramente mayor que la de las regiones donde hay hojas. Cuantitativamente el valor medio de reflectividad de la fruta es de 0.325 y el de las hojas es de 0.255, existiendo unas desviaciones estándar reducidas de 0.02 y 0.03 respectivamente.

En comparación con el estudio realizado con la fruta y el árbol artificial, podemos concluir que en el caso natural la fruta sigue siendo distinguible del fondo basándose en reflectividad. Es cierto que el margen de separación entre los valores medios de reflectividad es más reducido y por tanto podría resultar más difícil discriminar entre ambas clases, sin embargo la varianza en el caso natural también ha disminuido con lo cual las funciones gausianas, que definen la densidad de probabilidad, no se solapan fuertemente. El valor utilizado como umbral discriminante entre ambas clases, como se ve en la figura 5.21d, es 0.303.

Como también se comprobó en el estudio realizado en el CEMAGREF [170] (figs. 1.5 y 1.6), las manzanas son distinguibles de las hojas de los manzanos, trabajando a longitudes de onda entre 700 y 900 nm. Sin embargo las frutas pertenecientes a distintas variedades de manzana no son distinguibles entre sí trabajando en esa banda espectral infrarroja,

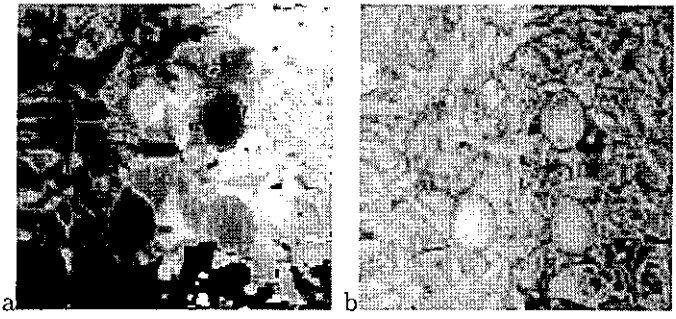


Figura 5.22: Influencia de la iluminación Solar sobre la distancia y reflectancia captada por el sensor. (a) y (b) imágenes de distancia y reflectancia de una escena natural. La parte izquierda de la escena fue captada con el Sol ocluido por una nube (iluminación tenue) y la parte derecha con el Sol sin ocluir (iluminación muy intensa).

incluso aunque las frutas tengan colores diferentes. En nuestro caso, debido a que el sensor infrarrojo es insensible a los colores, es de esperar igual comportamiento. Por tanto, las naranjas de diferentes variedades o en diferentes estados de madurez son difícilmente distinguibles trabajando con 780 nm. Esto conlleva a que si se desea realizar una recolección selectiva basada en la reflectividad de la fruta, sea necesario utilizar una o varias líneas espectrales en el visible. En concreto, para apreciar el cambio en el estado de madurez, que normalmente pasa por un estado inicial con tonos verdes y finaliza en muchos casos con colores amarillos, naranjas o rojos, se recomienda utilizar líneas espectrales entre 600 y 640 nm (rojo).

5.3.2.2 Influencia del Sol

En los estudios realizados en el laboratorio, la iluminación ambiente utilizada es de tubos fluorescentes, con niveles de iluminación tenues que típicamente tienen irradiancias de 1 W/m^2. En un entorno natural la iluminación de tipo solar, que llega a la superficie terrestre a cielo despejado, tiene irradiancias de 1000 W/m^2. Sin embargo, el ser humano no suele ser consciente de esta gran diferencia en tres ordenes de magnitud, ya que la energía luminosa que entra en nuestro ojo está regulada por el iris de forma refleja y la respuesta a la intensidad de luz es logarítmica. La luz solar tiene un amplio rango espectral que va desde el ultravioleta al infrarrojo; por tanto también emite a 780 nm que es la frecuencia del telémetro, aunque afortunadamente en esta banda espectral la atmósfera absorbe el 90% de esta radiación. La irradiancia debida al haz láser infrarrojo que se genera después de incidir sobre una superficie mate a un metro de distancia está en torno a 1 mW/m^2. Esto significa que existe un solapamiento espectral entre ambas fuentes de luz con una relación señal ruido muy baja. A pesar de que la radiación del telémetro está modulada en amplitud y la ambiente no, la interacción aditiva entre ambas señales puede alterar negativamente al proceso de medida, como de hecho se ha observado experimentalmente.

En la figura 5.22 se muestra la influencia del Sol en un caso práctico. Para obtener

esta imagen se comenzó explorando, de izquierda a derecha y verticalmente, la escena con el Sol oculto por una nube (mitad izquierda de la imagen), y en el transcurso de la adquisición el Sol se despejó repentinamente hasta que se acabó de adquirir el resto de la imagen (mitad derecha). Se ve claramente que la reflectancia $\Re$, o nivel de señal válida captada por el sensor, disminuye cuando la iluminación solar es mayor. Este fenómeno puede crear variaciones de hasta un 20% en la señal de reflectancia. Las consecuencias pueden ser diversas:

- *Falseado de la reflectividad calculada.* Al no estar modelada la influencia de los cambios de iluminación sobre el nivel de señal válida que recibe el sensor, la presencia de luz solar directa provoca que las reflectividades calculadas para una determinada región sea inferior a la real. Este hecho puede hacer que las primitivas regionales reflectividad dejen de ser discriminantes. La solución a este efecto podría consistir en utilizar un detector de iluminación ambiental y compensar adecuadamente los valores de reflectividad ρ.

- *Aumento de la repetitividad en distancia.* Como quedó modelado en el capítulo 2, la repetitividad de la distancia depende de la señal de reflectancia $\Re$ captada. Este efecto no es muy importante ya que disminuciones de $\Re$ del 20% provocan que la repetitividad se incremente en menos de 0.2 mm (si $D_{max} = 4000$ y $T = 1 \cdot 10^{-3}$ s).

- *Aumento de la probabilidad de medidas de distancia falsas.* En el capítulo 2 también se explicó que reflectancias absolutas inferiores a 35 (para un rango típico de reflectancia entre 10 y 160) generan medidas de distancia falsas. La iluminación solar por tanto, provoca que medidas que en condiciones de iluminación tenues tenían reflectancias entre 40 y 45, ahora estén muy próximas a 35 y por tanto sus correspondientes datos de rango sean falsos. En la franja vertical derecha de la figura 5.22a se pueden apreciar unos puntos negros que corresponden a medidas falsas en distancia.

5.3.2.3 Influencia del viento

El viento quizás pueda ser considerado como uno de los aspectos más degradantes encontrados en los entornos naturales. Este factor crea oscilaciones tanto verticales como horizontales de la fruta, con unas amplitudes que dependen de la fuerza del viento, de la longitud del pedúnculo, de la posición y del tamaño del fruto. Las amplitudes típicas encontradas en estos desplazamientos oscilantes están en torno a 10 y 30 mm para el viento relativamente fuerte que pudimos experimentar ($\simeq 30\ Km/h$).

Con el sistema de adquisición por deflexión de haz actual, debido a que el tiempo de adquisición es alto ($\simeq$20 s para imágenes de 500×500 mm y 3 mm/punto de resolución), el movimiento de la fruta en el árbol va a provocar una fuerte deformación en las imágenes, con lo cual es de esperar también un pérdida sustancial en la capacidad de detección. En la figura 5.23 se muestran tres tomas de una misma escena con diferentes intensidades del viento, y por tanto diferentes grados de deformación. En esta figura, de arriba a abajo la amplitud de viento se incrementa, y como se puede apreciar, las detecciones encontradas disminuyen de 7 frutos correctamente detectados, para el caso con viento débil (fig.

5.23b), a 5 y 3 frutos detectados cuando la intensidad del viento crece paulatinamente (figs. 5.23c y 5.23d).

Observando las imágenes degradadas hemos podido determinar que la frecuencia de oscilación de la fruta en el plano vertical está en torno a 1 Hz. En la figura 5.24 podemos ver tres claros ejemplos donde se aprecian cuatro periodos de deformación completos, por oscilaciones verticales en las naranjas de la esquina superior-derecha de la figura 5.24a y 5.24b, y la naranja de la esquina inferior-derecha de 5.24b. Como el tiempo necesario para barrer una franja vertical con un ancho igual al de una naranja es de $20 * 80/500 = 3.2$ segundos, tenemos que la frecuencia de oscilación es de 1.25 Hz. Para eliminar este efecto degradador, será necesario explorar esta franja de 80 mm de ancho a una velocidad superior para que se produzca un efecto de "congelación". Si consideramos que explorar esta franja en $\pi/4$ del periodo de la oscilación es suficiente para que la degradación del viento solo se manifieste como un ligero incremento en la excentricidad de la fruta, tenemos que la franja debe se barrida completamente en 0.1 segundos y por tanto la imagen completa en $0.1 * 500/80 = 0.625$ segundos o a casi 2 Hz.

Debido a que el viento afecta a la capacidad de detección al deformar las imágenes con la velocidad de barrido actual ($t_{adquis} \simeq 20$ s; imagen 500×500 mm, 3 $mm/punto$); tanto para permitir una operación en tiempo real ($t_{adquis} < 5$ s) como para que se pueda "congelar" la imagen ($t_{adquis} < 0.6$ s), se debería trabajar en un futuro con un sistema mejorado de adquisición para captar una imagen de 500×500 mm en 0.6 segundos ($\simeq 2$ Hz). Esta pretensión no debe afectar a otros parámetros característicos esenciales del sistema de medida como la potencia de emisión láser, el rango de medida, la precisión y la repetitividad en distancia. Hoy en día esto parece que puede estar fuera del alcance de la tecnología actual ya que no nos consta que exista ningún sistema telemétrico que cumpla con los requisitos planteados en el capítulo 2 y a la vez capte imágenes con la cadencia que acabamos de indicar (2 imágenes por segundo o 45.000 puntos por segundo). Sin embargo, cada vez se está más próximo a estas especificaciones y es de esperar que en breve tiempo sea posible disponer de sistemas telemétricos barriendo regiones grandes (5 metros), con precisiones y repetitividades elevadas (< 0.5 mm), con grandes velocidades de medida (> 200.000 $puntos/s$) y con potencias de emisión contenidas (clase 3a, clase 3b de menos de 10 mW o ligeramente mayores trabajando con longitudes de onda de más de 1500 nm).

5.4 Conclusiones

En este capítulo hemos evaluado los algoritmos presentados en la tesis de una forma conjunta analizando la calidad de los resultados en función de diversos parámetros degradantes. Las conclusiones más destacables son las siguientes:

- Los algoritmos tienen un *orden de complejidad lineal*, $O(N)$, lo cual es muy importante en percepción artificial por el gran volumen de datos tratados. Sin embargo, los tiempos absolutos de procesamiento no corresponden a un sistema de tiempo real ya que su optimización no se ha llegado a contemplar en esta tesis.

- Los algoritmos elaborados se caracterizan por una *alta robustez*. Esta propiedad surge del método redundante utilizado en el cual se trabaja con múltiples primitivas

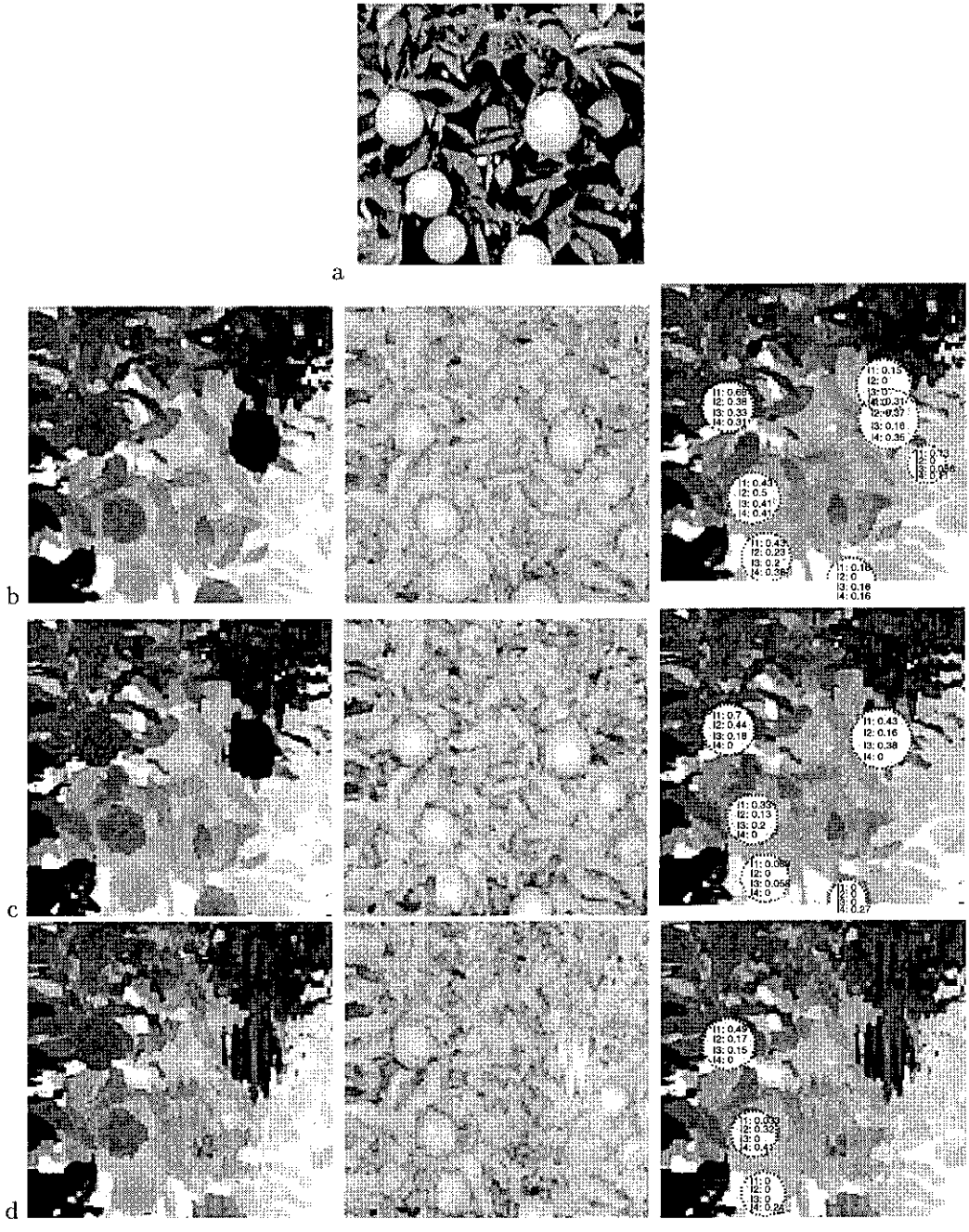


Figura 5.23: a) Fotografía en B/N de una escena de un árbol natural que contiene 9 frutos visibles; En las filas b), c) y d) se muestran de izquierda a derecha, el mapa de distancia, reflectancia y las detecciones encontradas, para diferentes tomas de la escena con intensidad del viento creciente.

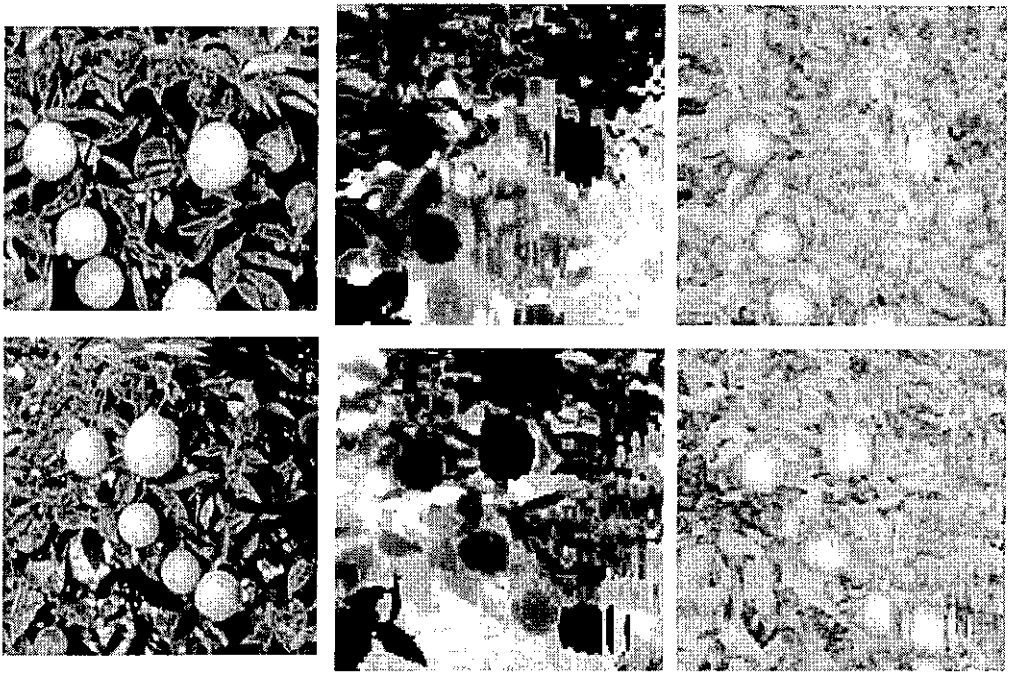


Figura 5.24: a) y b): Dos escenas diferentes captadas para apreciar la frecuencia de oscilación de los frutos con el viento. En cada fila se muestra de izquierda a derecha, una fotografía en B/N y las correspondientes imágenes de distancia y reflectancia.

para conseguir que el fallo de alguna de ellas afecte en la mínima medida al resto del sistema. Es cierto también que los algoritmos son sensibles a las diferentes degradaciones introducidas, pero los indicios de esfericidad disminuyen de una forma controlada permitiendo generar hipótesis válidas en un amplio rango de degradaciones; de ahí su robustez.

- Hemos comprobado que el hecho de utilizar los cuatro métodos de generación de hipótesis parciales, tiene un claro sentido ya que ninguno de los métodos por separado obtiene los mismos resultados que integrando las cuatro fuentes de hipótesis. Es decir, existe una *cooperación y complementación* entre los métodos propuestos, lo que permite mejorar de forma considerable las detecciones y la calidad de las estimaciones.

- La *estimación de parámetros* obtenida, tanto en la posición tridimensional como en el radio de las esferas, presenta una buena precisión, con errores milimétricos, casi siempre inferiores a la resolución utilizada para captar la imagen. Unos errores máximos menores de 3 mm en posición y radio, son resultados perfectamente adecuados para localizar objetos en situaciones complejas como presenta nuestra aplicación.

- Hemos comprobado que el filtro *3σ-MPF* contribuye de una forma positiva y ayuda tanto a las tareas de reconocimiento como a la estimación de los parámetros de las esferas. Al haber presentado resultados sustituyendo nuestro filtro por otras etapas de restauración, se han obtenido siempre resultados inferiores a los generados con el filtro *3σ-MPF*.

- Mediante la evaluación experimental en el laboratorio hemos visto que en un caso práctico de recolección de fruta, podemos conseguir *detecciones correctas* del orden del 80% de la fruta visible, y una probabilidad casi nula de *detecciones falsas*. Estos resultados pueden ser incluso ajustados dependiendo de que interese más, si detectar el máximo número de frutos a costa de aumentar la probabilidad de encontrar detecciones falsas, ó detectar menos frutos pero teniendo certeza absoluta de que son hipótesis correctas.

- Mediante la evaluación experimental en campo, hemos detectado otros *factores degradantes* no percibidos ni modelados hasta entonces. Estos factores perturbadores son el viento y la iluminación solar directa. El último factor es significativo pero no invalida el proceso de detección, sin embargo el factor viento nos obliga a adquirir imágenes a una cadencia de 2 Hz.

- Hemos comprobado que en condiciones naturales la *reflectividad*, al igual que en el montaje de laboratorio, nos permite distinguir las frutas del fondo de hojas, con lo cual el método basado en primitivas reflectividad es efectivo.

Finalmente, y desde un punto de vista global, es necesario comparar los resultados que hemos obtenido con los presentados en trabajos anteriores donde se pretendía automatizar la detección de fruta. En el capítulo 1 se presentó de forma detallada las características, logros y problemas que se obtuvieron en cada una de las estrategias propuestas por otros autores, las cuales se basaban en el uso de cámaras CCD en B/N o

en color. Recordamos que los mayores problemas utilizando cámaras eran las oclusiones, las sombras, la ausencia de información de profundidad y las regiones confusas visibles a través del árbol que generaban detecciones falsas. En la mayoría de los casos los métodos se limitan a distinguir la fruta del fondo por color, con lo cual los algoritmos solo funcionan cuando la fruta está madura. Cuando la fruta es distinguible del fondo por color, circunstancia que no siempre se manifiesta pues en muchos casos la fruta se recoge antes de que madure del todo, los mejores resultados muestran detecciones correctas de hasta el 90% de la fruta visible y detecciones falsas superiores al 5% [112]. En los casos en que el sistema puede detectar tanto fruta madura como verde, la mejor solución obtenida se encuentra en el proyecto CITRUS [165] donde se dan tasas de detección correcta del 75% y un 8% de fallos.

Nosotros, con un sistema telemétrico láser y utilizando una estrategia de preprocesamiento y análisis de formas, reflectividades y distribuciones espaciales, hemos conseguido obtener tasas de detección de fruta, independientemente de cual sea el estado de madurez, del 80% de los frutos visibles; con unas detecciones falsas nulas o muy poco probables. El problema de las sombras, la ausencia de información de profundidad para localizar al fruto y la presencia de regiones visibles a través del árbol, han sido eliminados en nuestra solución. Sin embargo aún existen factores limitantes como la oclusión parcial de la fruta (aparte de la oclusión total de esta que puede llegar a ser del 50-60% [112]), que hace que frutos que son visibles en menos de un 20-25% de su superficie sean difíciles de detectar. Otros factores que actualmente restringen la aplicación de nuestro sistema, aunque son problemas de índole tecnológica, son los tiempos de adquisición y de cómputo.

A pesar de las limitaciones mencionadas del sistema desarrollado, estimamos que las metodologías elaboradas tienen una plena validez como estrategias generales discriminantes para la detección de objetos. Se haría posible la resolución de la aplicación real disponiendo en un futuro próximo de los componentes físicos adecuados para poder trabajar sin limitaciones, captando imágenes a distancias medias (2-5 metros), con alto nivel de detalle ($<$ 1 milímetro) y con altas velocidades ($>$ 2 Hz), y procesándolas en tiempos que no limiten el ciclo de recolección ($<$ 1.5 s/fruto).

Conclusiones, aportaciones y trabajos futuros

Una vez que el trabajo desarrollado ha sido presentado a lo largo de la memoria, ha llegado el momento de recapitular lo que se pretendía hacer, lo que se ha conseguido, las contribuciones novedosas que puedan suponer aportaciones útiles al resto de la comunidad científica y los posibles trabajos que quedan abiertos en esta línea para el futuro.

Conclusiones finales

El objetivo final de la tesis se centraba en estudiar la realización de un sistema automático de reconocimiento y localización de frutos, que basado en técnicas diferentes a las ya presentadas por otros investigadores que utilizan cámaras de visión, permita resolver ciertos problemas encontrados por ellos y que fundamentalmente son las frecuentes detecciones falsas, la limitada capacidad de detectar fruta con tonos de color próximos a los del fondo y las dificultades para determinar la posición tridimensional de cada fruta lo cual dificulta su recolección.

Las técnicas presentadas en esta tesis, han permitido desarrollar un sistema de detección que no genera detecciones falsas, o al menos su probabilidad es muy baja, manteniendo una tasa de detecciones por encima del 80% de la fruta visible. El hecho de que la fruta no sea siempre distinguible del fondo utilizando su color, no influye en la capacidad global de detección del sistema elaborado, ya que el parámetro color no es utilizado en el análisis de imágenes. Junto a la detección del fruto se obtiene la posición tridimensional de su centro con unas precisiones siempre menores a 3 mm, permitiendo que un robot se pueda dirigir a este punto siguiendo la trayectoria que se crea más adecuada y no solamente la del eje de visión. Así mismo, cada una de las esferas detectadas queda parametrizada mediante su radio y reflectividad superficial, con lo cual se podría pensar en realizar recolecciones selectivas basadas en alguno de estos parámetros. Esas características que acabamos de resaltar hacen que el sistema presentado tenga un comportamiento que mejora los resultados obtenidos en otros trabajos previos, sin embargo también presenta ciertas limitaciones.

Una de las principales limitaciones que presenta el sistema planteado, al igual que en las técnicas tradicionales, son los problemas de detección ante la existencia de oclusiones parciales del fruto por parte de las hojas, y que por supuesto utilizando sensores que

trabajen con radiación que no se transmite a través de las hojas no es posible solventar. Sin embargo, el sistema planteado permite detectar correctamente esferas visibles en más de un 20-25%, con lo cual y dependiendo del tipo de árbol explorado, se obtienen casi siempre detecciones superiores al 80% de la fruta visible.

La limitación principal son los tiempos de captura y de procesamiento del prototipo de sistema de detección presentado. En el proceso de adquisición de imágenes los tiempos típicos requeridos para completar la exploración están en torno a los 20 segundos para una imagen de 500 × 500 mm con una resolución de 3 mm/punto. Debido al ciclo de trabajo requerido para la recolección de la fruta, y fundamentalmente al viento que ocasiona deformaciones en las imágenes, se estima que debería realizarse la adquisición de este tipo de imágenes en 0.5 segundos (2 Hz). Conseguir esto es una tarea compleja, ya que simplemente aumentar la velocidad de barrido utilizando espejos con múltiples caras girando a velocidad constante, no es una solución posible si el propio sensor no es capaz de medir con la cadencia requerida para conseguir el muestreo espacial deseado y con la suficiente calidad de medida en la distancia. Este es un aspecto problemático que seguramente el desarrollo tecnológico resolverá en un futuro muy próximo. En cuanto a los tiempos de cálculo que se requieren para procesar y analizar una imagen, que son de 60-80 segundos utilizando un Pentium 150 MHz sin ningún soporte físico adicional para mejorar su rendimiento, tampoco son apropiados para una aplicación en tiempo real. Para garantizar un ciclo de recolección de 1 fruto cada 1.5 segundos, y debido a que en cada imagen hay una media de 10 frutos visibles, los tiempos de cálculo por cada imagen que se estiman adecuados no deben superar 5 segundos. Esto significa que la mejora a realizar debe estar entre uno y dos órdenes de magnitud, lo cual se puede hacer simplificando los algoritmos o añadiendo más unidades de procesamiento.

En cualquier caso, a parte de estas limitaciones de carácter fundamentalmente tecnológicas, queremos resaltar la validez de los planteamientos algorítmicos presentados que permiten dotar al sistema elaborado de una gran capacidad para realizar detecciones correctas de objetos cuasi-esféricos, posibilitando la aplicación de la metodología tanto en el caso agrícola planteado como en otros campos en los que se contemplen labores de automatización.

Aportaciones presentadas

A lo largo de los cinco capítulos de los que consta la tesis, en el apartado final de conclusiones de cada uno de ellos se han citado los aspectos destacables presentados y las aportaciones más relevantes. Fundamentalmente, podemos considerar que las aportaciones principales de esta tesis son las siguientes:

- Se ha realizado una *revisión bibliográfica* detallada de los trabajos de visión orientados a la detección de frutos en el campo, así como un repaso de técnicas de medida de distancia, de restauración y de análisis de imágenes.

- Se ha propuesto una *estrategia de detección mixta* que opera aplicando un modo de detección automático seguido de uno asistido, permitiendo aunar las ventajas de ambas estrategias.

- Hemos propuesto basar el proceso de detección automático en la *forma* de los objetos, en sus *propiedades ópticas* y en la *distribución espacial* de estos. Planteando la necesidad de utilizar un sensor que capte distancia y reflectancia.

- Hemos realizado un *análisis de los sistemas de telemetría láser* actuales, definiendo un índice de eficiencia tecnológico M que nos permite ver las limitaciones actuales y las relaciones de compromiso entre los parámetros que caracterizan un sistema láser de medida de distancias.

- Presentamos un *modelo matemático del telémetro* que relaciona la señal de reflectancia con otros parámetros que intervienen en el proceso de medida, posibilitando la integración de la información de distancia con la de reflectancia.

- Así mismo, hemos *modelado la repetitividad de la distancia* mostrando la dependencia de la reflectancia sobre la desviación estándar del ruido aditivo que afecta a la medida de distancia, posibilitando una estimación del ruido para realizar restauraciones adaptativas.

- Se ha elaborado una *nueva métrica* para la medida de la *calidad global de restauración*, *GRI*, la cual valora tanto las restauraciones fieles como la suavidad en las reconstrucciones.

- Se ha elaborado una *nueva técnica de restauración de imágenes* denominada *3σ-MPF* que, utilizando una estrategia multiresolución y un criterio de fidelidad de ajuste de planos, permite realizar restauraciones con alto grado de suavizado y a la vez preservar discontinuidades sin introducir distorsiones significativas. Tanto por criterios visuales, utilizándolo como una etapa de preprocesamiento en un sistema de reconocimiento completo, como aplicando la métrica *GRI*, este filtro se comporta mejor que los restantes tratados.

- Hemos propuesto un *método de reconocimiento, localización y caracterización de objetos esféricos* que se basa en la extracción de diversas primitivas que, utilizadas como fuentes de indicios y tras una integración final de las hipótesis parciales generadas, permite detectar objetos esféricos generando su posición tridimensional, radio y reflectividad.

- Hemos definido *cuatro tipos de primitivas*, dos de carácter puntual y dos de tipo regional. Tres de ellas captan características discriminantes en la forma de un objeto esférico. La otra capta regiones con propiedades ópticas iguales a las del objeto buscado. La utilización de las cuatro primitivas hace que el método de reconocimiento se base en la complementariedad aportada por cada una de ellas, lo cual le da un carácter robusto ante oclusiones y degradaciones.

- Finalmente, realizamos una *evaluación simulada y empírica de los algoritmos* parametrizando diversos tipos de entrada y diferentes ajustes internos del algoritmo, con lo cual se comprueba la capacidad del sistema propuesto para detectar correctamente frutos y objetos cuasi-esféricos, sin generar detecciones falsas, y con mínimos errores en la estimación de sus parámetros. Así mismo, esta evaluación permite juzgar las posibilidades de utilizar las técnicas mostradas en este trabajo en otros campos diferentes.

Trabajos futuros

Como ya apuntamos anteriormente en el apartado de conclusiones, se considera que la estrategia planteada es adecuada, sin embargo existen limitaciones en los apartados de tiempo de adquisición y análisis de imágenes. Por tanto, se debería proporcionar al sistema del soporte físico que permita su evolución en tiempo real, para lo cual se deberá centrar esfuerzos de desarrollo hacia la mejora de los tiempos de captura, desarrollando sistemas de exploración telemétricos rápidos que generen imágenes de alta calidad, y por otro lado mejorar los tiempos de procesamiento mediante nuevas arquitecturas de procesamiento o simplificando los algoritmos.

Debido a que los algoritmos de procesamiento utilizan varios parámetros ajustables cuya inapropiada elección provoca que el proceso no sea totalmente óptimo, sería interesante disponer de un controlador que observando la evolución de los resultados intermedios y finales del algoritmo, ajustase de forma adaptativa parámetros tales como el número de primitivas utilizadas, la resolución espacial de las imágenes, la velocidad de barrido o el umbral de indicios. De esta forma lo que se conseguiría es modificar la fase de adquisición y procesamiento mediante una sintonía automática con el fin de que los resultados obtenidos sean óptimos.

Así mismo quedan abiertas a la investigación otras estrategias de captación totalmente diferentes a las planteadas, que de alguna forma permitan captar las diferencias en las propiedades volumétricas y de absorción de energía que presentan los frutos y las hojas, con lo cual se estaría en condiciones de obtener unas seudo-imágenes en las cuales la oclusión por parte de las hojas no estuviese presente y por tanto la detección fuese prácticamente inmediata.

También se podría plantear la realización de una fusión entre dos sensores diferentes: telémetros láser y cámaras ópticas. Este posible trabajo tiene sentido si se trata de aprovechar las propiedades más ventajosas de ambos sensores. La ventaja en las cámaras ópticas está en la rapidez de adquisición y en la información más rica de color al generar tres componentes espectrales, lo cual permite realizar detecciones rápidas cuando los frutos son distinguibles por color del fondo y a la vez se posibilita la clasificación de la fruta según su grado de madurez. Los sistemas telemétricos actualmente son más lentos, sin embargo tienen otras propiedades muy interesantes como la generación directa de la distancia absoluta, la posibilidad de realizar un análisis basado en la forma sin problemas de iluminación y la opción de analizar la distribución en el espacio de los objetos detectados. Estos aspectos hacen que las estimaciones sean muy fiables y que las detecciones falsas sean muy poco probables. En este sentido la integración de ambos sensores podría resultar interesante aunque también se incrementaría la complejidad del sistema a desarrollar.

Apéndice A

Hojas técnicas de los componentes del sistema telemétrico de exploración

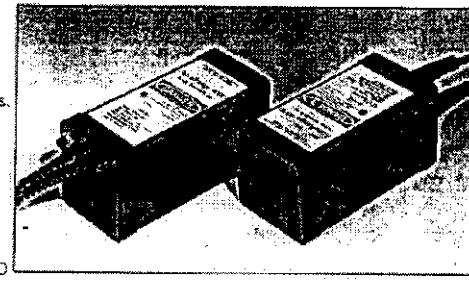
En este apéndice presentamos los datos técnicos de los componentes utilizados en el sistema de deflexión telemétrico que se presentó en el capítulo 2. Estos elementos son los motores, tarjetas de control y etapas de potencia para dotar al sistema de los movimientos angulares en acimut y elevación necesarios para realizar la exploración; los espejos deflexores y monturas de alineamiento; el diodo láser apuntador y el propio telémetro láser junto con su tarjeta de adquisición de alta velocidad. A continuación (figs. A.1-A.10) podemos ver los detalles técnicos que los fabricantes proporcionan.

AccuRange 4000 by Acuity RESEARCH

The AccuRange 4000 is an optical distance measurement sensor with a useful range of zero to 50 feet for most diffuse reflective surfaces. It operates by emitting a collimated laser beam that is reflected from the target surface and collected by the sensor. The sensor is suitable for a wide variety of distance measurement applications that demand high accuracy and fast response times.

Key Features

- Zero to 50 feet operating range for most surfaces
- High precision: 0.02 inch short-term repeatability
- RS-232 serial and pulse width outputs standard. Optional 4-20 mA current loop output.
- Reflected signal amplitude output for grayscale images.
- Fast response time: 50 KHz maximum sample rate
- Lightweight, compact, low power design.
- Tightly collimated output beam for small spot size.
- Two output beam configurations available, visible or infrared.
- Ideally suited to level and position measurement, machine vision, autonomous vehicle navigation, and 3D imaging applications.



AccuRange 4000 Technical Specifications

Standard Configurations

4000-LIR: Emitter: 780 nm
IR laser diode
Optical power:
5 milliwatts max.
Effective Range:
50 feet

4000-LV: Emitter: 670 nm
Visible red laser diode
Optical power:
5 milliwatts max.
Effective Range:
40 feet

Of these configurations, the 4000-LIR has the lower measurement noise, greater sensitivity and maximum range. The 4000-LV features visible light output. See the relevant specifications. Laser power is user adjustable, and custom configurations are also possible.

Mechanical

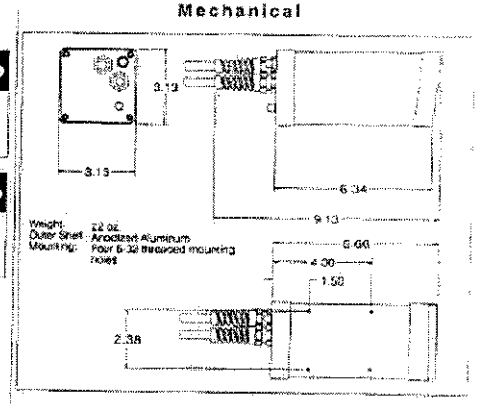


Figura A.1: Telémetro láser AccuRange 4000-LIR (1/2).

Interface

Power

Sensor Power: +5 Volts (+5V min, +6V max) @ 300 mA
Heater Power: +5 Volts (+5V min, +6V max) @ 2 A, temperature dependent. May be used to stabilize sensor temperature in low-temperature environments.

Power and Signal Cable

Color	Function	Direction
Red	Power	In
Black	Ground	
Orange	Heater Power	In
Brown	Heater Return	
Yellow	Temperature	Out
Blue	Pulse or Current	Out
Green	Ambient Light	Out
Purple	Amplitude	Out

Data and Communications

Serial Interface

RS-232 serial output on a standard 9-pin PC-compatible cable. Output may be in an internally calibrated distance format with a resolution of .01", or in an uncalibrated format which includes uncompensated range and signal strength, background light, and sensor temperature.

Data Formats

Ascii Mode:
- Calibrated Output: 3-6 bytes
  - Distance 0 to 655 inches in units of .01"
  - CR,LF terminated
- Low Level Outputs: 9-24 bytes
  - CR,LF terminated

Binary Mode:
- Calibrated Output: 3 bytes
  - Distance 0 to 655 inches in units of .01"
  - FF terminated
- Low Level Outputs: 8 bytes
  - FFFF terminated

Device Configuration via AccuRange Command Set

- Baud Rate: 300-38400 baud*
- Sample Rate
- Ascii or Binary Output
- Maximum Range *
- Laser Power Level
- Laser On/Off*
- Zero Point Set
- Current Loop Zero, Span*
- Current Loop Mode (Off/Calibrated/Uncalibrated)*
- Serial Mode (Off/Calibrated/Uncalibrated)*
- Take Single Sample

* May also be set without serial communication, using pushbutton/LED interface.

Pulse Width and Voltage Outputs

Pulse width indication of uncalibrated range, and analog voltage signals. Standard on the power/signal cable. Voltage outputs are 0-5 volt analog levels indicating target signal strength, background illumination, and sensor temperature. The pulse width and voltage outputs are typically used for external calibration during high-speed sampling (above 1000 samples/second).

Pulse width: uncalibrated range output
18 samples/second to 50,000 samples/second

Optional 4-20 mA Current Loop

Replaces pulse width output on the power/signal cable
Calibrated output to 1 KHz or high-speed uncalibrated output to 10 KHz
Adjustable zero and span points
Adjustable zero current: 4 mA default

Operating Temperature: 0 - 113°F.

Performance

Two forms of output are available from the sensor. Calibrated output is distance in units of .01". Calibration compensates for sensor temperature and target reflectance, and is performed in the sensor. The sample rate for calibrated output is limited to 1 KHz. For higher speed output, the uncalibrated serial outputs or pulse width and voltage signals may be used.

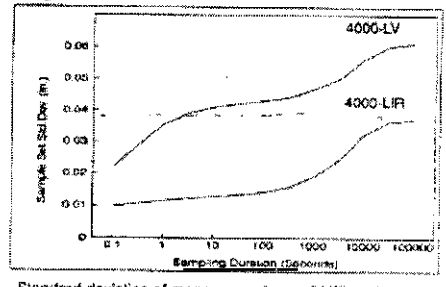
Useable Range

4000- LIR: 0 to 50 feet (calibrated)*
4000-LV : 0 to 40 feet (calibrated)*
0 to 1500 feet (uncalibrated, with reflective target)

Short Term Measurement Noise

.0005 inches / $\sqrt{Hz}$*
(all samples within 0.3 sec)

Long Term Drift



Standard deviation of measurement error (drift) vs. time

Calibrated Absolute Accuracy

Standard Deviation of Range Error:
4000-LIR : .1"
4000- LV: .3"

Range Signal Response Time

< 1 sample for all sample rates

Amplitude, Ambient Light, and Temperature Outputs

Analog signals, 0-5 Volts
Response Times (except temperature): 20 microseconds
Amplitude: Logarithmic signal
Ambient (Background) Light: Linear signal
Temperature: Linear 0-140°F. (31.3 millivolts/degree)

Optical

Laser Beam Diameter: 0.1"
Laser Beam Divergence: 0.5 milliradians
Return Light Collection Aperture Diameter: 2.5"

The AccuRange High Speed Interface is also available, with sample rates to 50,000 samples/second for all AccuRange 4000 outputs.

Note 1: To 85% diffuse reflectance target.
US Patent # 5,309,212

Acuity Research, Inc. 3475 Edison Way # P Menlo Park, CA 94025 (415) 369-6782 Fax (415) 369-6785

Figura A.2: Telémetro láser AccuRange 4000-LIR (2/2).

AccuRange High Speed Interface

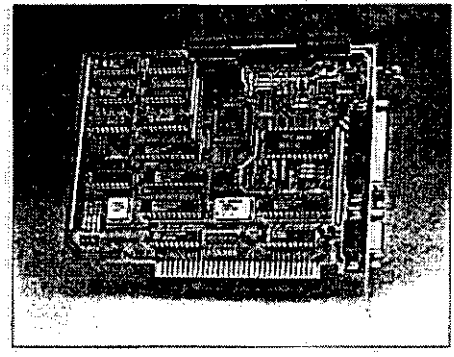
by

Acuity
RESEARCH

The AccuRange™ High Speed Interface can be used to increase the sample rate capability of Acuity Research's line of AccuRange sensors, using any PC-compatible computer. It can sample all the AccuRange outputs at up to 50,000 times per second and buffer the data for reading by the host computer. The High Speed Interface can also be used to control the AccuRange Line Scanner. For a full description of the AccuRange 4000 and AccuRange Line Scanner, see the corresponding data sheets.

Key Features

- Sample rates to 50,000 samples per second.
- Samples AccuRange distance, amplitude, ambient light, and temperature outputs.
- Supplies 8-byte samples over the standard PC ISA bus.
- 2K or optional 16K byte buffer for sample data onboard.
- Buffer empty, half-full, and overflow indicators, and half-full interrupt.
- Start/stop sampling control input.
- 3 general purpose input data lines for event recording.
- Improved resolution and higher sample rate capability than the AccuRange 4000 alone.
- Supplies power to the AccuRange sensor.
- Optional motor power control and shaft encoder reading capability.



AccuRange High Speed Interface Technical Specifications

Mechanical and Power

Form Factor
IBM-PC ISA bus, half-length board, 6.25 by 4.75 inches

Power Requirements: +5V @ 500 ma
Includes power for AccuRange sensor

AccuRange 4000 Interface

The AccuRange High Speed Interface measures the duration of the pulse width output of the AccuRange 4000. Setting sample rate and all other configuration of the 4000 is done through the serial interface on the 4000, over a computer serial port (not included on the High Speed Interface). Signal strength, ambient light, and sensor temperature are also sampled. For maximum accuracy, calibration to obtain actual distance then occurs in software on the host computer.

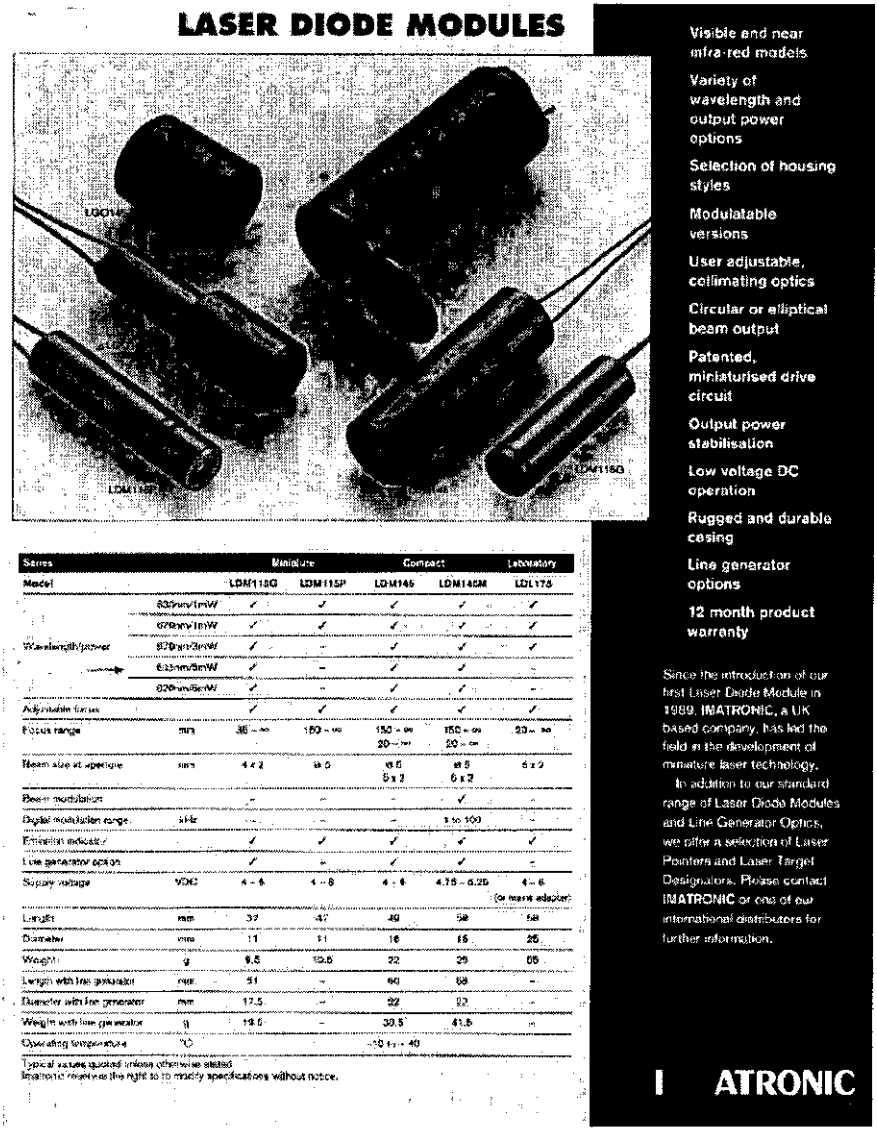
P1: DB-9 power and signal connector

Pin	Function	Direction
1	Power, +5V (5V min, 6V max)	Out
2	Ground	
3	Heater Power, 4.5 to 6.0 V	Out
4	Heater Power Ground	
5	Temperature signal, 0-5V	In
6	Pulse Width Range Signal	In
7	Ambient light signal, 0-5V	In
8	Amplitude signal, 0-5V	In
9	No Connection	
Shield	Ground	

Amplitude, Ambient Light, and Temperature Inputs

Analog signals, 0-5V; 8 bit samples
Response times:
One sample period, as programmed

Figura A.3: Tarjeta de alta velocidad para comunicación con el telémetro láser.



LASER DIODE MODULES

- Visible and near infra-red models
- Variety of wavelength and output power options
- Selection of housing styles
- Modulatable versions
- User adjustable, collimating optics
- Circular or elliptical beam output
- Patented, miniaturised drive circuit
- Output power stabilisation
- Low voltage DC operation
- Rugged and durable casing
- Line generator options
- 12 month product warranty

Series		Miniature		Compact		Laboratory
Model		LDM115G	LDM115P	LDM145	LDM145M	LDL175
Wavelength/power	635nm/1mW	✓	✓	✓	✓	✓
	670nm/10mW	✓	✓	✓	✓	✓
	673nm/3mW	✓	–	✓	✓	✓
	633nm/5mW	✓	–	✓	✓	–
	826nm/5mW	✓	–	✓	✓	–
Adjustable focus		✓	✓	✓	✓	✓
Focus range	mm	35 – ∞	150 – ∞	150 – ∞ 20 – ∞	150 – ∞ 20 – ∞	20 – ∞
Beam size at aperture	mm	4 x 2	ø 5	ø 5 5 x 2	ø 5 6 x 2	5 x 2
Beam modulation		–	–	–	✓	–
Digital modulation range	kHz	–	–	–	1 to 100	–
External adjust		✓	✓	✓	✓	✓
Line generator option		✓	–	✓	✓	–
Supply voltage	VDC	4 – 6	4 – 6	4 – 6	4.75 – 6.25	4 – 6 (or mains adapter)
Length	mm	37	47	49	58	58
Diameter	mm	11	11	16	16	25
Weight	g	9.5	10.5	22	25	65
Length with line generator	mm	51	–	60	68	–
Diameter with line generator	mm	17.5	–	22	22	–
Weight with line generator	g	19.5	–	30.5	41.5	–
Operating temperature	°C			-10 to + 40		

Typical values quoted unless otherwise stated
Imatronic reserves the right to modify specifications without notice.

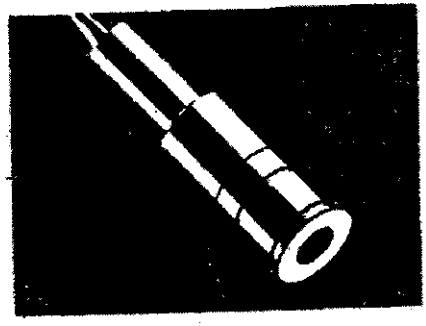
Since the introduction of our first Laser Diode Module in 1989, IMATRONIC, a UK based company, has led the field in the development of miniature laser technology.

In addition to our standard range of Laser Diode Modules and Line Generator Optics, we offer a selection of Laser Pointers and Laser Target Designators. Please contact IMATRONIC or one of our international distributors for further information.

IMATRONIC

Figura A.4: Diodo láser de 633 nm modelo LDM115G.

LASER CROSS GENERATOR LCG115



The Laser Cross Generator (LCG) is a plug-on accessory which, when fitted to the front of a pre-focused miniature laser diode module (LDM115 series), projects a laser line cross. The LCG is retained on the front of the miniature LDM by means of two 'O' rings and a grub screw. Applications include alignment, measurement, position sensing and inspection.

Optical Specification

Typical performance data:

Operational range (metres)	*Cross length (mm)*	*x*	*line thickness (mm)*
0.5	50	x	1
2	160	x	1.5
5	380	x	2

Operational wavelength range: 630nm - 820nm
Perpendicularity (two axis): < 2 degrees

Mechanical Specification

Outside diameter: 17mm Length: 35mm
Overall length when fitted to a miniature laser diode module: 60mm

iMATRONIC

Kingfisher Court, Hambridge Road, Newbury, Berkshire RG14 5SJ UK
Tel: 01635 550477 Fax: 01635 49158
International Tel: +44 1635 550477 Fax: +44 1635 49158

Figura A.5: Generador de cruz láser acoplable a diodos de la serie LDM115.

All Band Laser Diode High Reflector BD.2

Newport's **BD.2** is a high reflectance, broad angle of acceptance dielectric coating designed for laser diode applications. BD.2 mirrors deliver maximum laser energy from 700–950 nm for all polarizations at angles of incidence up to 45°. Reflectivity at 632.8 nm is provided to permit easy alignment with HeNe lasers.

Application: Near-infrared laser diodes

Optical Performance: R_s>99%; R_p>98%; 0–45°, 700–950 nm; R>80%@ 632.8 nm

Durability: MIL-C-675C

Cleaning: Any nonabrasive method. Lens cleaning fluid and tissue advised.

Safe Energy Level: 1000 W/cm² CW; 10 mJ/cm² 10 ns pulse

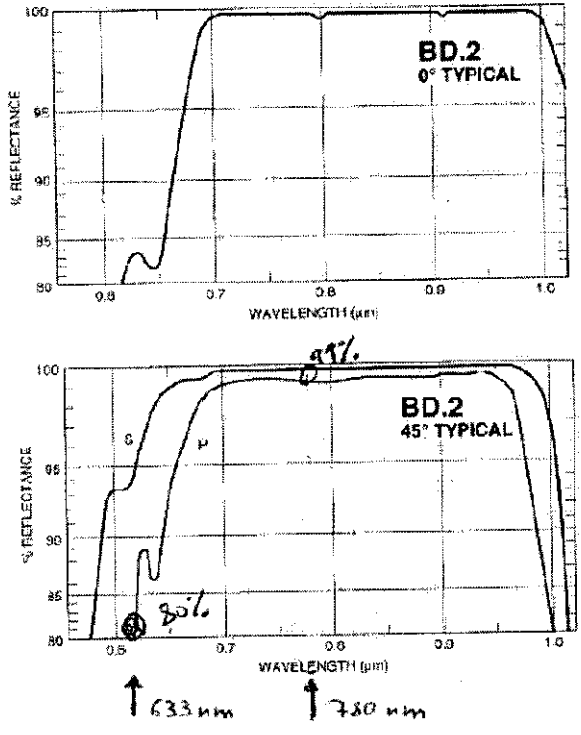


Figura A.6: Características de reflexión para recubrimientos dieléctricos de tipo BD.2 utilizados en el espejo móvil.

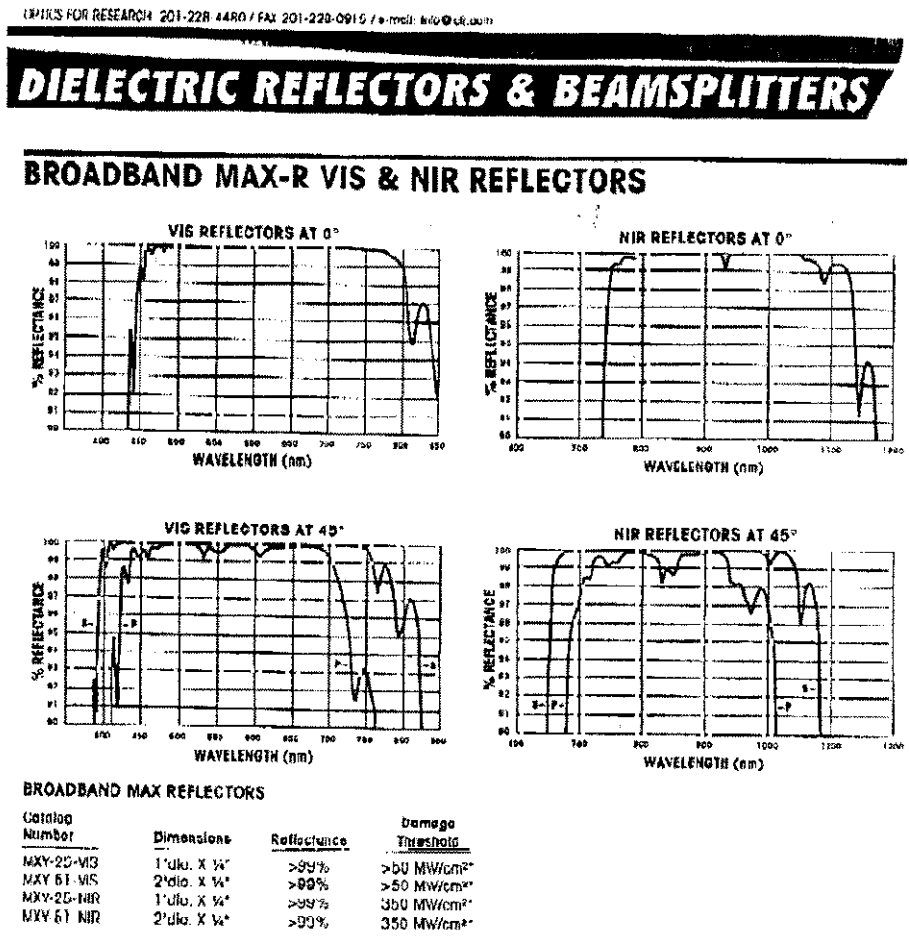
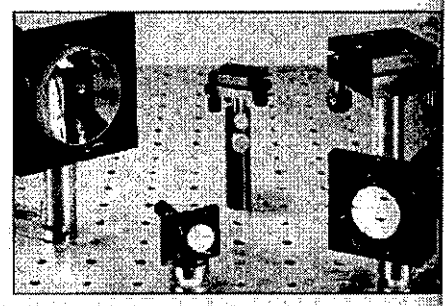
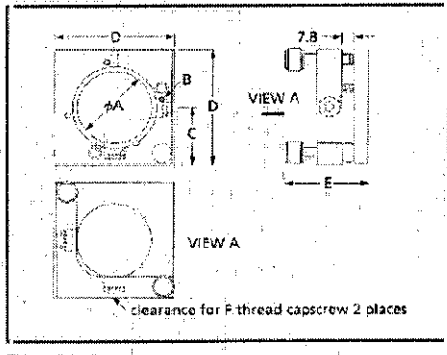


Figura A.7: Características de reflexión del "espejo caliente" utilizado para integrar el haz visible (633 nm) con el invisible (780 nm).

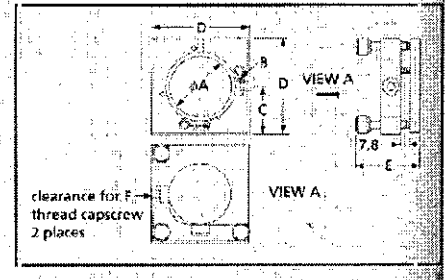
Kinematic mirror/beamsplitter mounts are an ideal low cost solution for two axis tilt. The kinematic design includes hardened surfaces to resist wear and increase stability.

- Precision stainless steel adjustment screws with brass threaded inserts provide exceptionally smooth adjustment.
- Triple-drive option provides translation.
- Counter bored holes provided for easy "top down" mounting to mounting posts (page 404) and pillars (page 415).
- Angular range of 10°.
- M6 and ¼-20 cap screws provided.





07 MHT 02X kinematic mount with two adjusters



07 MHT 03X kinematic mounts with three adjusters

07 MHT Mirror Mounts

	ØA Optic Diameter	Angular* Resolution	Number of Adjusters	B Hole Pattern	C Optical Axis Height (mm)	D (mm)	E (mm)	F (mm)	PRODUCT NUMBER
	16	20 arc sec	2	M2 on 20 mm center	15	27.5	28	M4/8-32	07 MHT 021
→	12.5	20 arc sec	2	M2 on 20 mm center	15	27.5	28	M4/8-32	07 MHT 023
	25	10 arc sec	2	M4 on 38 mm bolt circle	25	50	43	M4/8-32	07 MHT 025**
→	50	6 arc sec	2	M4 on 38 mm bolt circle	37.5	75	53	M4/8-32	07 MHT 027
	16	20 arc sec	3	M2 on 20 mm center	15	27.5	28	M6/1/4-20	07 MHT 031
	12.5	20 arc sec	3	M2 on 20 mm center	15	27.5	28	M6/1/4-20	07 MHT 033
	25	10 arc sec	3	M4 on 38 mm bolt circle	25	50	43	M6/1/4-20	07 MHT 035
	50	6 arc sec	3	M4 on 38 mm bolt circle	37.5	75	53	M6/1/4-20	07 MHT 037

*Angular resolution calculated from 2° rotation of adjuster. **07 MHT 025 replaces 07 MHT 001.

Figura A.8: Monturas de ajuste fino utilizadas para soportar el diodo láser y el "espejo caliente".

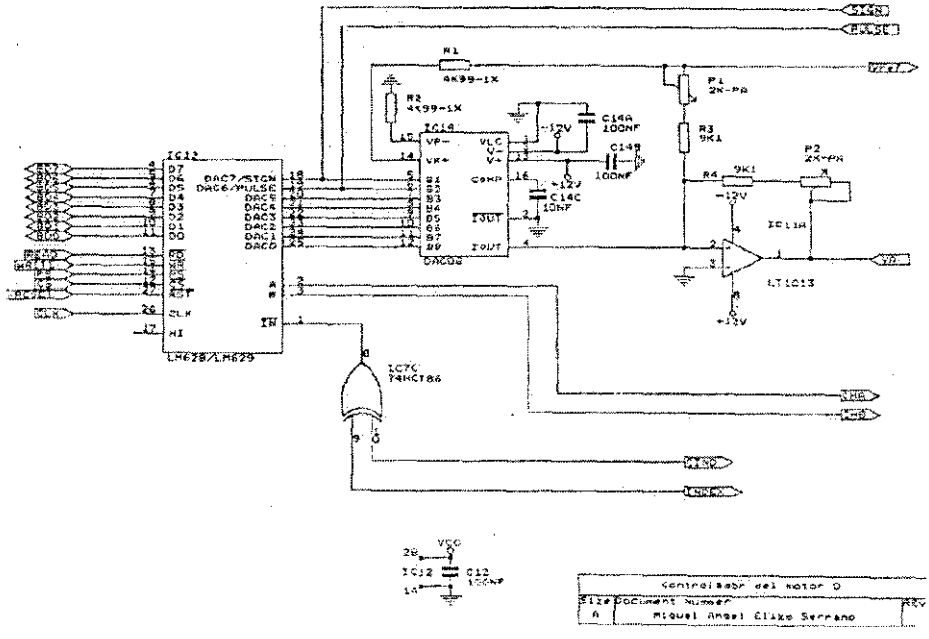


Figura A.9: Esquema de la tarjeta de control de motores utilizada que se basa en el integrado LM629.

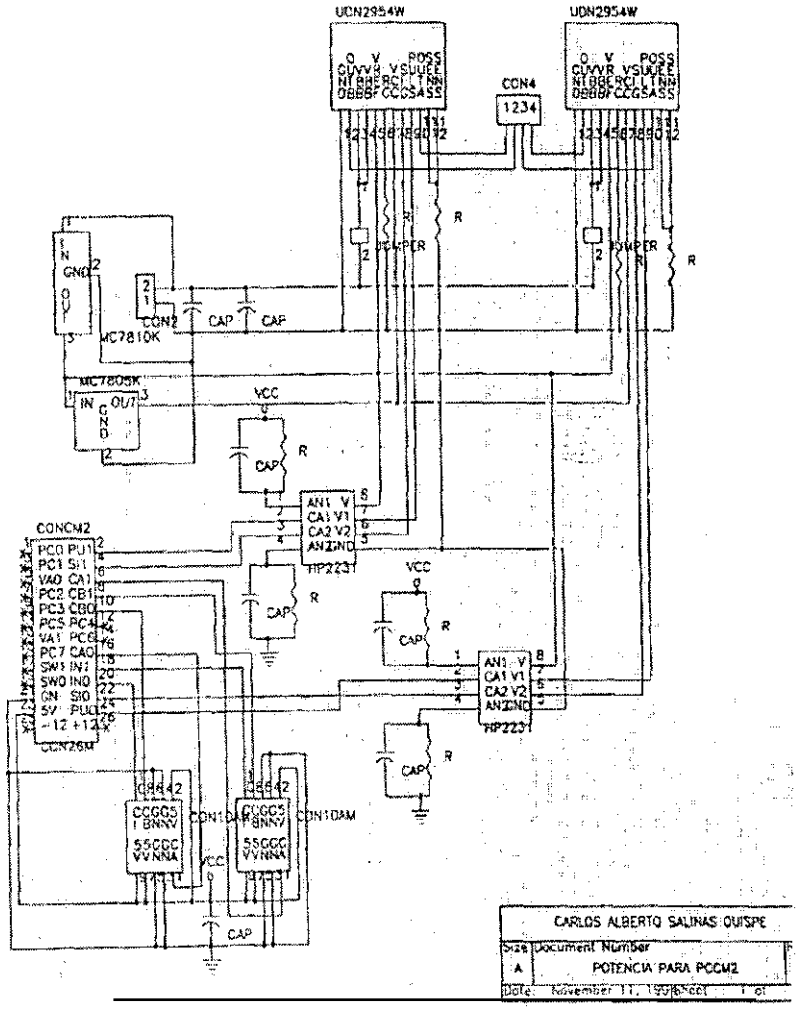


Figura A.10: Esquema electrónico de la etapa de potencia utilizada para amplificar la señal que comanda los motores.

Apéndice B

Medidas experimentales para modelar el telémetro láser

En el capítulo 2 (ec. 2.11), presentamos un modelo inicial mediante el cual se relacionaban parámetros del proceso de medida, como la distancia al blanco D, el ángulo formado entre el eje de medida con la normal a la superficie θ, la reflectividad propia de la superficie ρ, la reflectancia captada $\Re$ y otras constantes conocidas o sin determinar. Este era el modelo:

$$\Re = cte_1 \log(cte_2 \frac{AF_i}{\pi} \frac{\rho \cos\theta}{D^2}) \tag{B.1}$$

Este modelo nos dá la relación general de dependencia que existe entre los parámetros que intervienen en un proceso de medida utilizando telémetros operando mediante principios equivalentes al nuestro. Para completar el modelo es aún necesario determinar las constantes cte_1 y cte_2 que son desconocidas. Con este propósito se realizó un ajuste por mínimos cuadrados con datos experimentales, que nos permitió derivar estas constantes y a la vez comprobar que el modelo planteado es correcto. En este apéndice mostramos el proceso de recogida de datos (fig. B.1), el cual consiste en una toma repetida de valores de reflectancia, $\Re$, mientras se cambian los parámetros de medida. Los parámetros considerados son D y θ, y han sido variados en un rango de 1 a 4 metros a pasos de 200 mm, para el caso de la distancia, y en un rango de 0 a 80 grados en pasos de 10 grados en el caso de θ. Los datos obtenidos mediante este procedimiento se muestran en la tabla B.1.

Igualmente, en este apéndice mostramos los datos experimentales obtenidos tomando valores de *repetitividad en distancia* a diferentes valores de *reflectancia*, $\Re$, siendo obtenidos estos últimos, variando de forma aleatoria el tipo de superficie, la distancia y el ángulo de incidencia, hasta conseguir una muestra representativa en un amplio rango de $\Re$. La tabla B.2 contiene los datos recogidos de esta forma. Gracias a estos datos y a los planteamientos presentados en el capítulo 2, se pudo deducir que existía una clara relación entre $\Re$ y la repetitividad en distancia σ_D, que venía dada por la expresión:

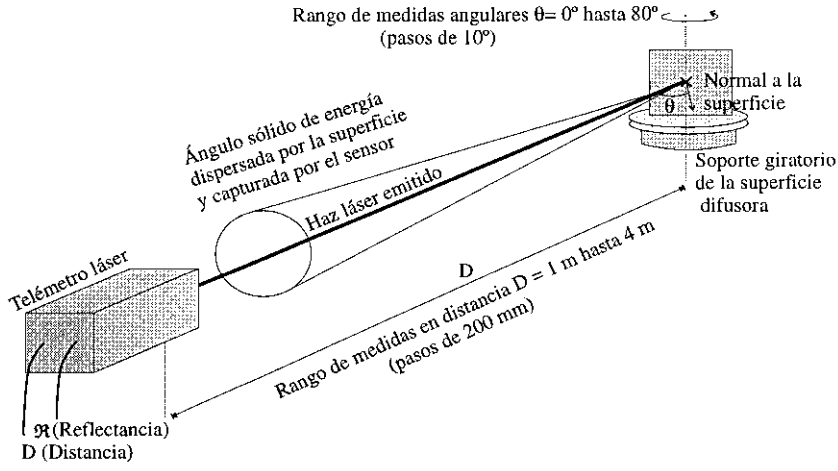


Figura B.1: Montaje experimental para la recogida de los datos que se muestran en la tabla B.1

$$\sigma_D = \sqrt{\frac{k}{T10^{\Re/80.0}} + \left(\frac{D_{max}}{9 \times 10^7 T}\right)^2} \tag{B.2}$$

donde la constante k quedó determinada aplicando un ajuste por mínimos cuadrados.

Ángulo (grados)	Distancia (mm)	Reflectancia
0	1000	145.5
0	1200	148
0	1400	150.5
0	1600	154
0	1800	154
0	2000	154
0	2200	154
0	2400	154
0	2600	153.5
0	2800	153
0	3000	152
0	3200	151
0	3400	149.5
0	3600	148
0	3800	146.5
0	4000	146
10	1000	144
10	1200	146.5
10	1400	150
10	1600	152
10	1800	152
10	2000	152.5
10	2200	153
10	2400	152.5
10	2600	152
10	2800	151.5
10	3000	151
10	3200	150
10	3400	148
10	3600	147
10	3800	145.5
10	4000	145
20	1000	140
20	1200	144
20	1400	147.5
20	1600	148
20	1800	148.5
20	2000	150
20	2200	149.5
20	2400	150
20	2600	149
20	2800	148
20	3000	147.5
20	3200	146.5
20	3400	145.5
20	3600	144
20	3800	142
20	4000	141

Ángulo (grados)	Distancia (mm)	Reflectancia
30	1000	136
30	1200	140
30	1400	142
30	1600	145
30	1800	146
30	2000	145
30	2200	146
30	2400	146
30	2600	145
30	2800	145
30	3000	144
30	3200	142
30	3400	141.5
30	3600	140
30	3800	138.5
30	4000	138
40	1000	132
40	1200	135
40	1400	137
40	1600	140
40	1800	140
40	2000	140.5
40	2200	140.5
40	2400	141
40	2600	140.5
40	2800	139.5
40	3000	138
40	3200	138
40	3400	137
40	3600	136
40	3800	134
40	4000	133
50	1000	124.5
50	1200	127.5
50	1400	131.5
50	1600	133
50	1800	134
50	2000	134
50	2200	134
50	2400	134
50	2600	134
50	2800	133
50	3000	132
50	3200	132
50	3400	130
50	3600	129
50	3800	128
50	4000	126

Ángulo (grados)	Distancia (mm)	Reflectancia
60	1000	116
60	1200	119
60	1400	122.5
60	1600	125
60	1800	125
60	2000	126.5
60	2200	125.5
60	2400	126
60	2600	126
60	2800	125
60	3000	124
60	3200	124
60	3400	123
60	3600	122
60	3800	120
60	4000	119
70	1000	105
70	1200	107
70	1400	111
70	1600	112
70	1800	114
70	2000	114.5
70	2200	114
70	2400	114
70	2600	114
70	2800	113
70	3000	112
70	3200	112
70	3400	110.5
70	3600	110
70	3800	109
70	4000	106.5
80	1000	83
80	1200	84
80	1400	88
80	1600	92.5
80	1800	92
80	2000	92
80	2200	92.5
80	2400	92
80	2600	91
80	2800	91
80	3000	91
80	3200	91
80	3400	90.5
80	3600	88
80	3800	87
80	4000	86

Tabla B.1: Datos experimentales para determinar la dependencia de la *reflectancia* con las variaciones en la distancia de medida y el ángulo de incidencia con la normal a la superficie.

Reflectancia	Desviación estándar (mm)
152	0.32
147	0.33
143	0.34
138	0.37
134	0.40
127	0.44
120	0.52
112	0.58
106	0.66
96	0.77
98	0.74
91	0.81
82	0.93
73	1.05
63	1.20
52	1.45
166	0.26
152	0.31
137	0.36
123	0.48
117	0.54
108	0.62
102	0.69
97	0.74
91	0.79
80	0.94
53	1.40
68	1.10
62	1.17
56	1.29
163	0.28
160	0.29
158	0.30
155	0.30
144	0.34
142	0.35
140	0.37
138	0.38
136	0.39
132	0.42
125	0.48
119	0.50
104	0.66

Tabla B.2: Datos experimentales para captar la relación existente entre la reflectancia captada por el sensor, $\Re$, y la desviación estándar en distancia, intentando de esta forma parametrizar la repetitividad en la medida como función de $\Re$.

Apéndice C

Muestra de imágenes de laboratorio distancia-reflectancia utilizadas en la evaluación empírica

En este apéndice mostramos un subconjunto de las imágenes de laboratorio utilizadas para realizar la evaluación empírica presentada en el capítulo 5 (figs. C.1-C.12). Estas imágenes, que fueron captadas con el sensor láser mostrado en el capítulo 2, vienen dadas en parejas distancia-reflectancia ya que son los datos que genera el sensor y que a la vez requieren los algoritmos de reconocimiento y localización presentados. Junto a cada pareja de imágenes también mostramos las hipótesis de objetos esféricos generadas al analizar las imágenes, incluyendo los indicios parciales que conducen a su formulación. La localización en el plano imagen y la estimación del radio, se muestran superponiendo una circunferencia sobre el mapa de distancias. Todos los resultados mostrados se obtuvieron utilizando un umbral de indicios de 0.15 y un radio de referencia, R, de 40 mm.

La resolución con la que fueron muestreadas las escenas son de 2, 3 y 4 milímetros por punto, por tanto existen tantas versiones de la escena como resoluciones utilizadas para obtener las imágenes distancia-reflectancia. Todas las imágenes se caracterizan porque los objetos de interés están situados a una distancia entre 2000-3000 mm y tienen un diámetro medio variable entre 73 y 82 mm, con excentricidades siempre menores de 1.05. En su generación se utilizó una frecuencia de muestreo en el telémetro de 1500 Hz, con tiempos de captura, de cada pareja de imágenes distancia-reflectancia, entre 10 y 20 segundos, dependiendo del área explorada y de la resolución espacial. El ruido gausiano con el que las imágenes de distancia están contaminadas depende de la reflectancia aparente captada por el sensor, como se mostró al modelar el láser en el capítulo 2, y en estas imágenes varía en un rango de 1 a 2 mm. El nivel de ruido impulsivo trabajando a estas frecuencias relativamente bajas, y con estos tipos de superficies bastante difusoras, es prácticamente inexistente como se puede apreciar.

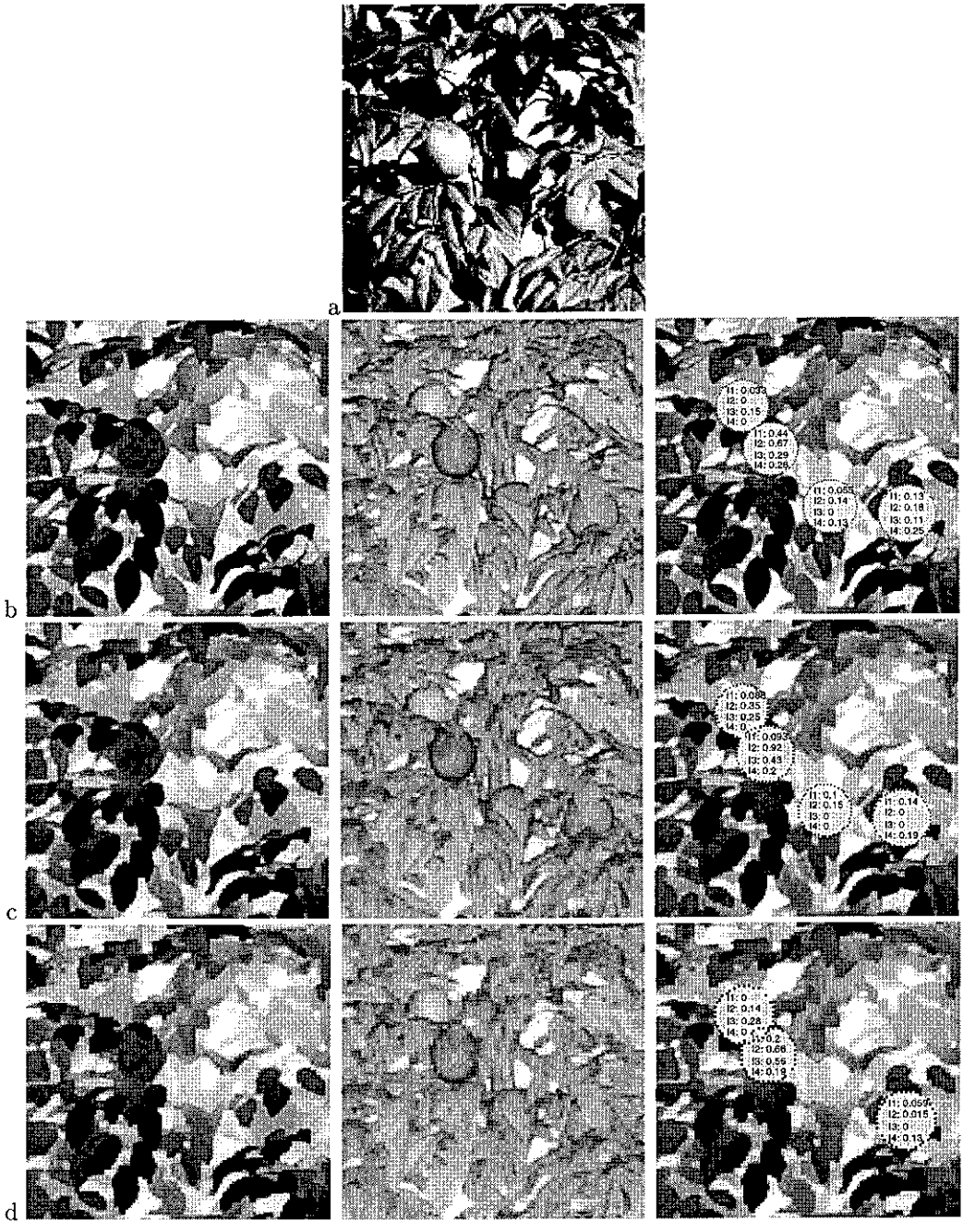


Figura C.1: a) Fotografía en B/N de árbol artificial con cuatro frutos. b),c) y d) Imágenes con resolución espacial de 2, 3 y 4 mm/puntos respectivamente (de izquierda a derecha: mapa de distancia, reflectancia y detecciones encontradas); se indica posición y radio estimados mediante anillos superpuestos en mapa de distancia con indicios parciales corona (I1), contorno (I2), convexo (I3) y reflectividad (I4).

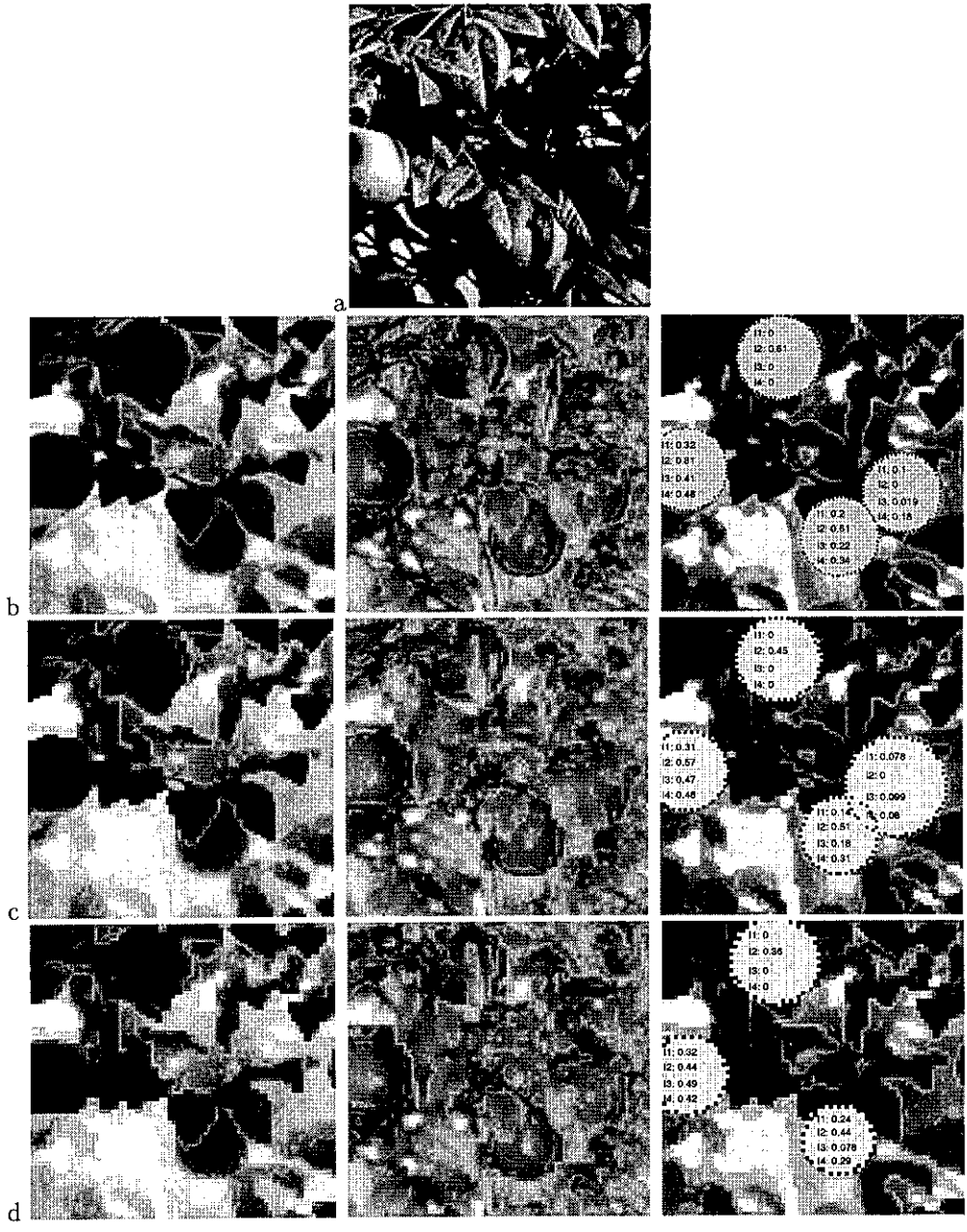


Figura C.2: a) Fotografía en B/N de árbol artificial con cuatro frutos. b),c) y d) Imágenes con resolución espacial de 2, 3 y 4 mm/puntos respectivamente (de izquierda a derecha: mapa de distancia, reflectancia y detecciones encontradas); se indica posición y radio estimados mediante anillos superpuestos en mapa de distancia con indicios parciales corona (I1), contorno (I2), convexo (I3) y reflectividad (I4).

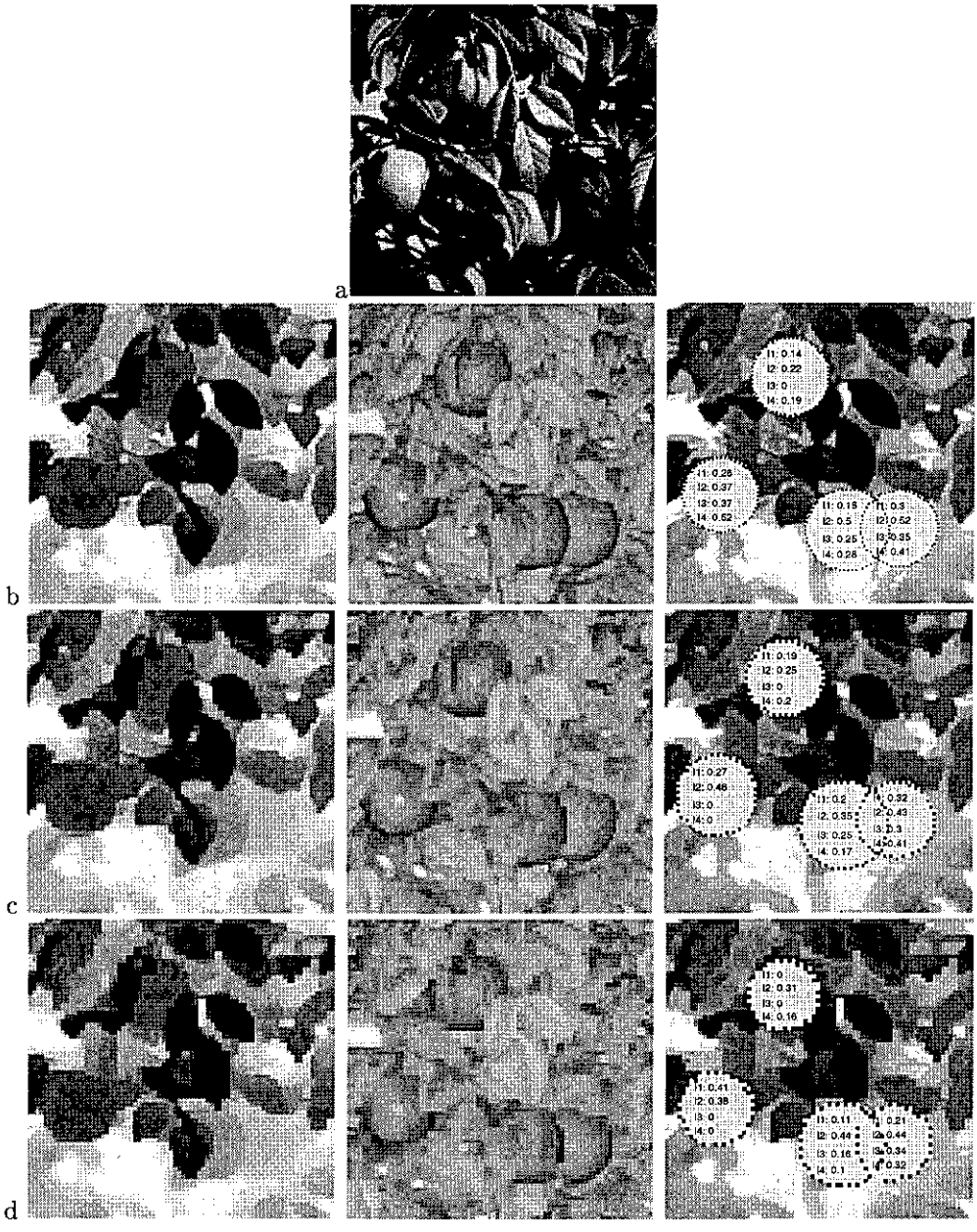


Figura C.3: a) Fotografía en B/N de árbol artificial con cuatro frutos. b),c) y d) Imágenes con resolución espacial de 2, 3 y 4 mm/puntos respectivamente (de izquierda a derecha: mapa de distancia, reflectancia y detecciones encontradas); se indica posición y radio estimados mediante anillos superpuestos en mapa de distancia con indicios parciales corona (I1), contorno (I2), convexo (I3) y reflectividad (I4).

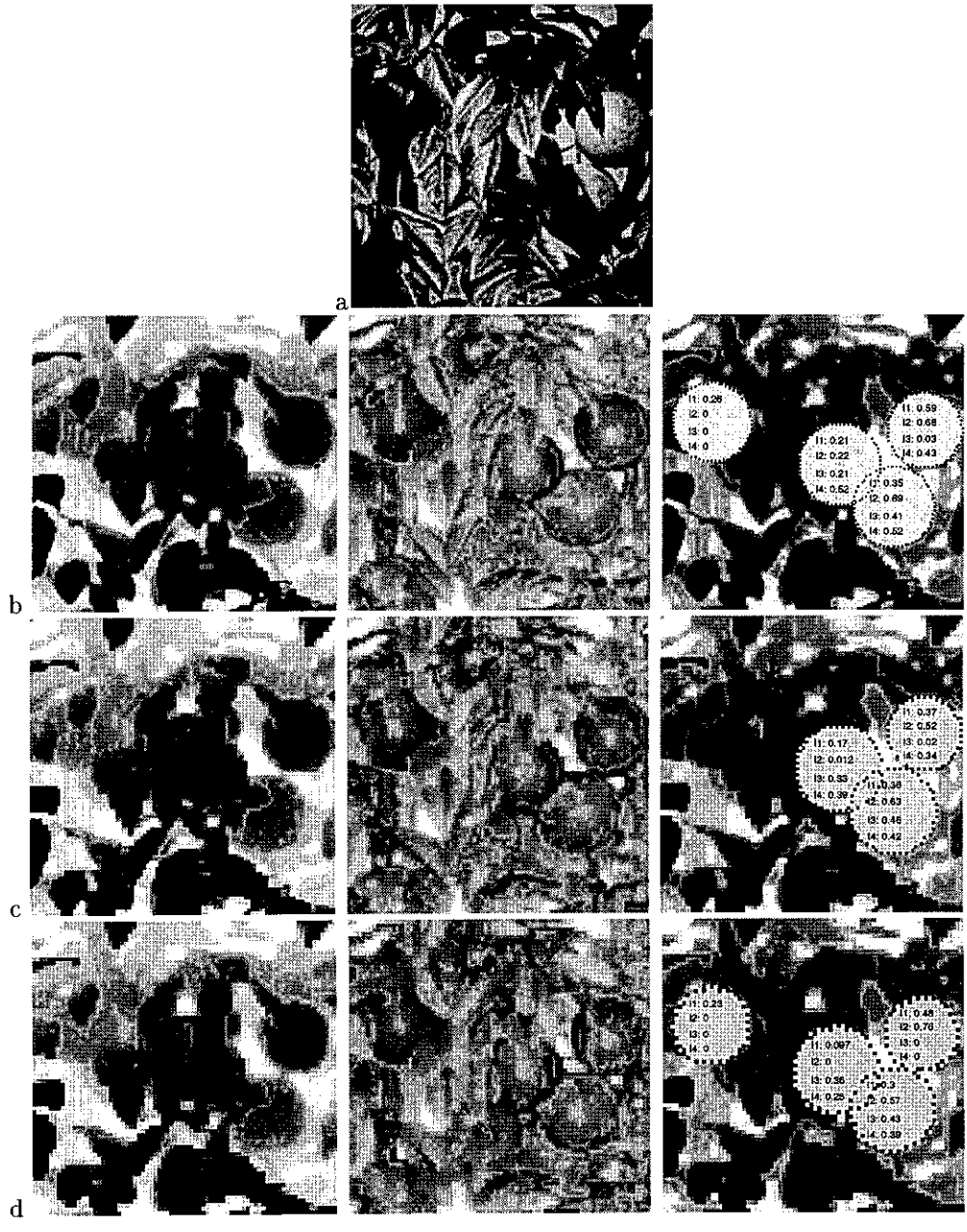


Figura C.4: a) Fotografía en B/N de árbol artificial con cuatro frutos. b),c) y d) Imágenes con resolución espacial de 2, 3 y 4 mm/puntos respectivamente (de izquierda a derecha: mapa de distancia, reflectancia y detecciones encontradas); se indica posición y radio estimados mediante anillos superpuestos en mapa de distancia con indicios parciales corona (I1), contorno (I2), convexo (I3) y reflectividad (I4).

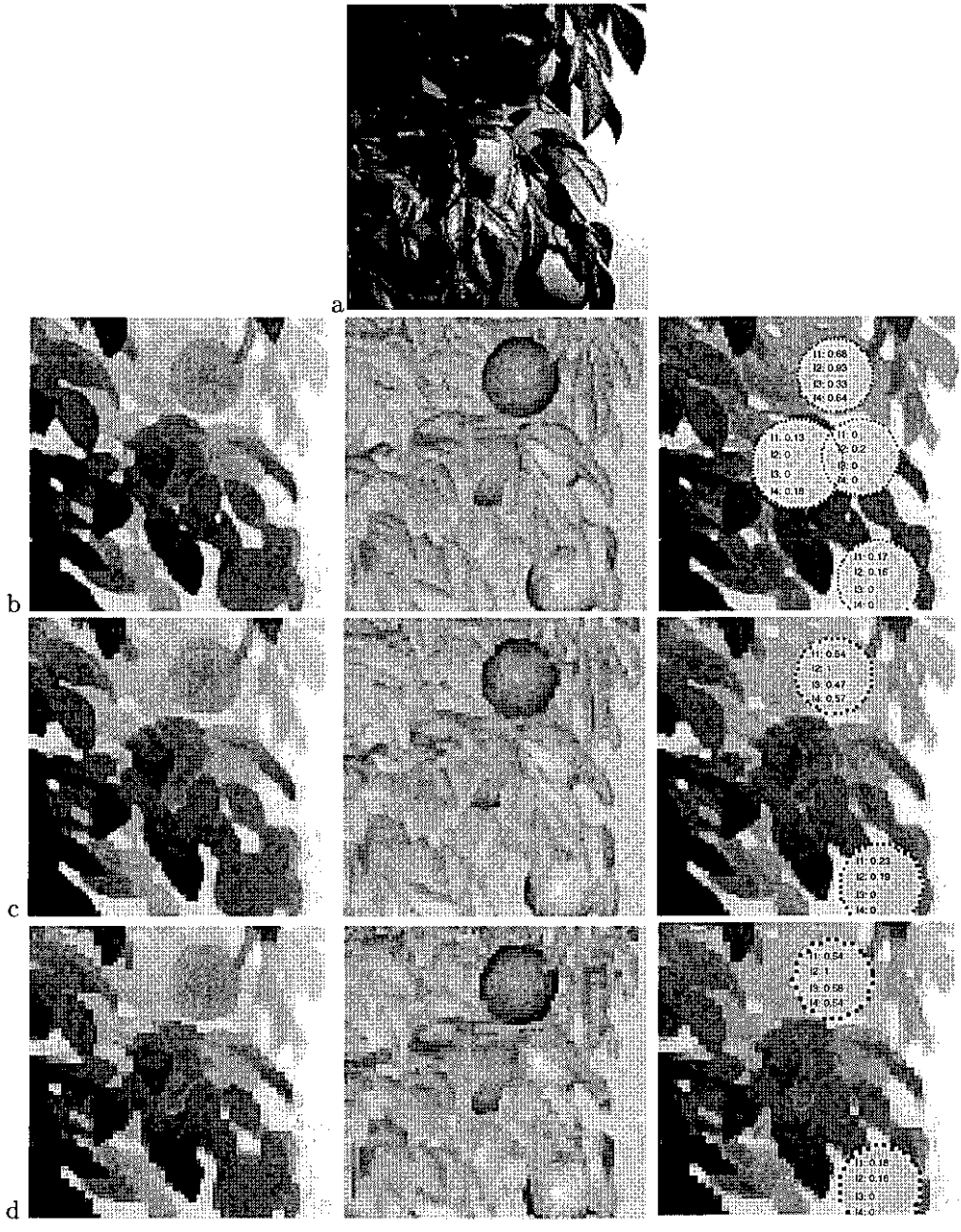


Figura C.5: a) Fotografía en B/N de árbol artificial con cuatro frutos. b),c) y d) Imágenes con resolución espacial de 2, 3 y 4 mm/puntos respectivamente (de izquierda a derecha: mapa de distancia, reflectancia y detecciones encontradas); se indica posición y radio estimados mediante anillos superpuestos en mapa de distancia con indicios parciales corona (I1), contorno (I2), convexo (I3) y reflectividad (I4).

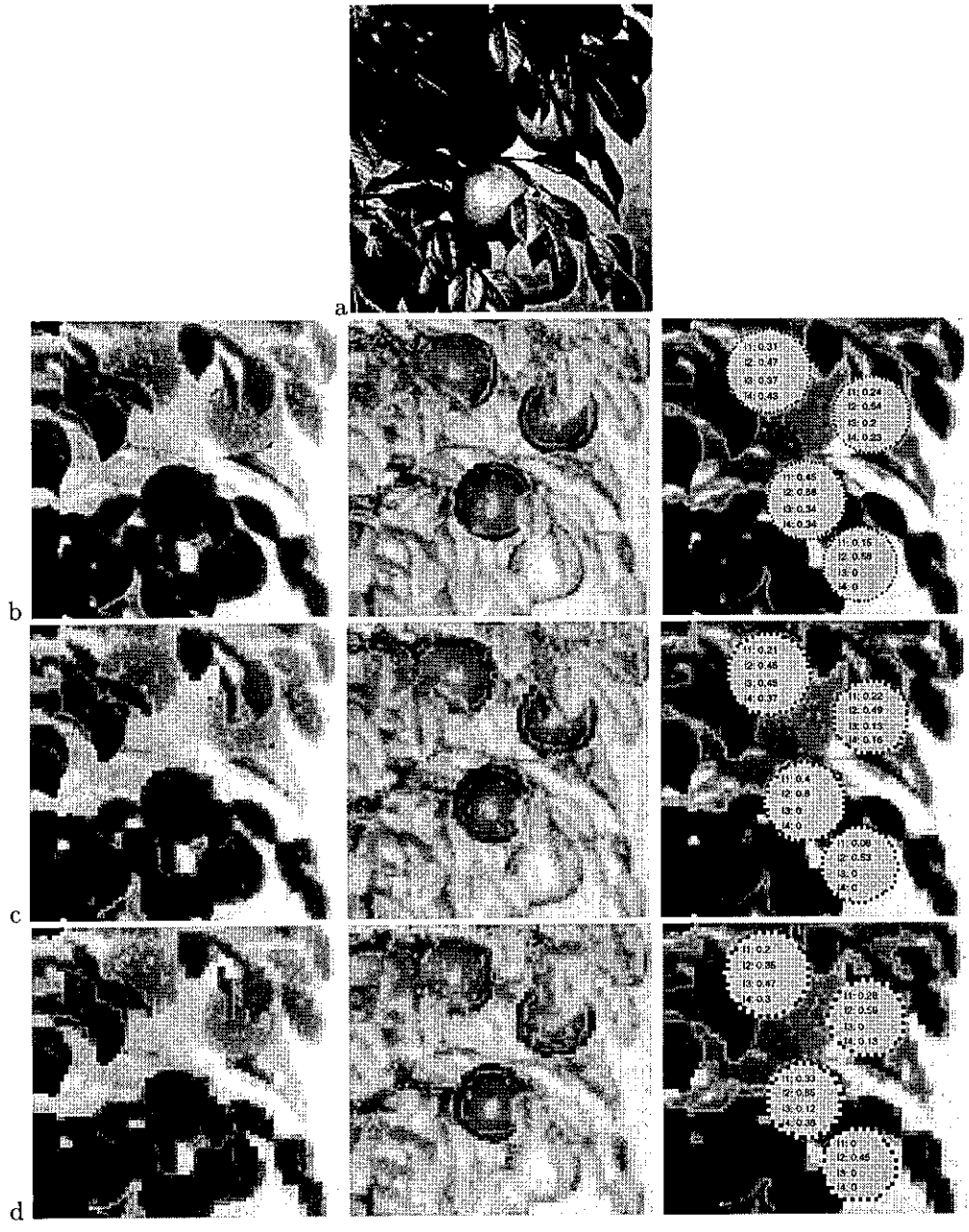


Figura C.6: a) Fotografía en B/N de árbol artificial con cuatro frutos. b),c) y d) Imágenes con resolución espacial de 2, 3 y 4 mm/puntos respectivamente (de izquierda a derecha: mapa de distancia, reflectancia y detecciones encontradas); se indica posición y radio estimados mediante anillos superpuestos en mapa de distancia con indicios parciales corona (I1), contorno (I2), convexo (I3) y reflectividad (I4).

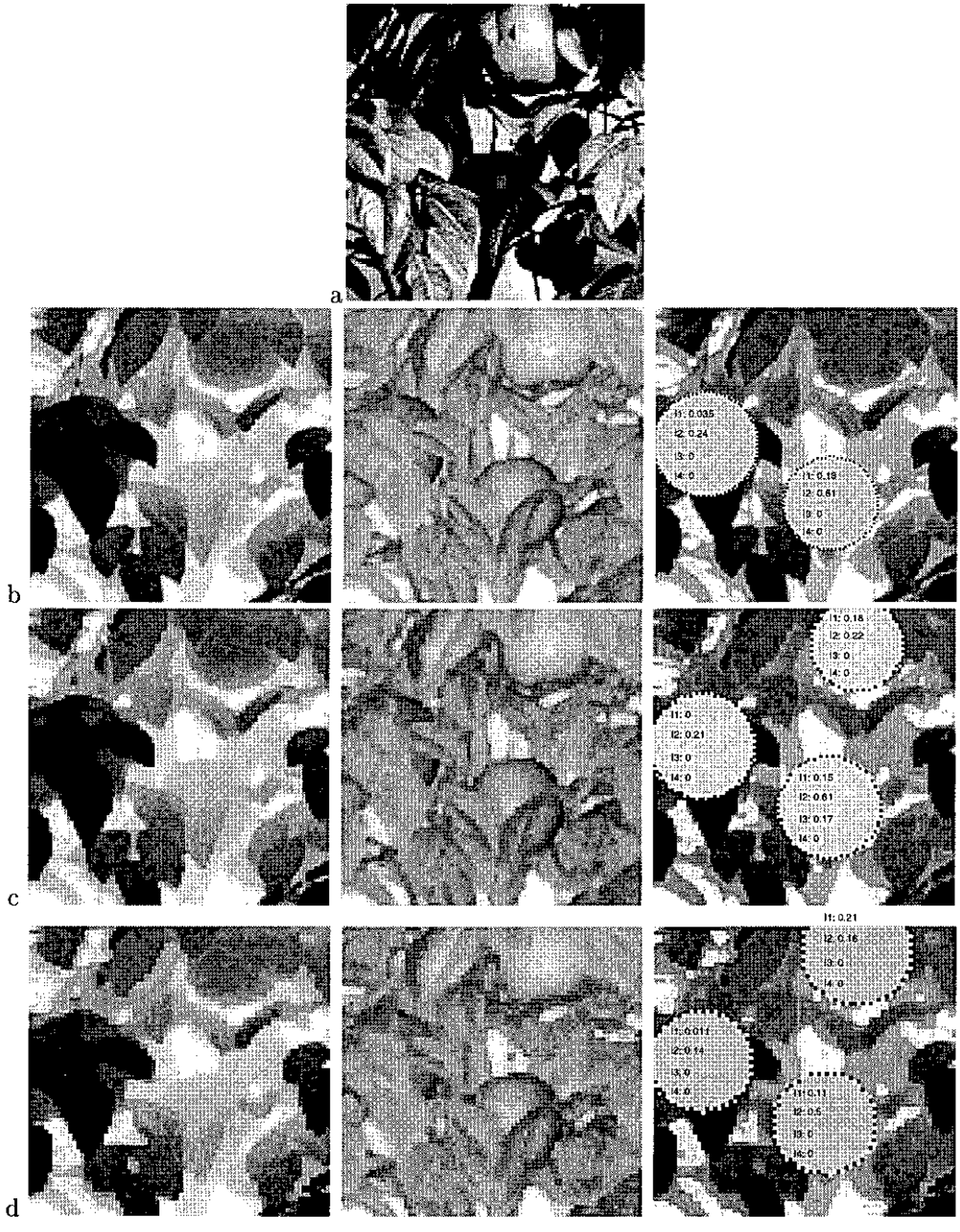


Figura C.7: a) Fotografía en B/N de árbol artificial con cuatro frutos. b),c) y d) Imágenes con resolución espacial de 2, 3 y 4 mm/puntos respectivamente (de izquierda a derecha: mapa de distancia, reflectancia y detecciones encontradas); se indica posición y radio estimados mediante anillos superpuestos en mapa de distancia con indicios parciales corona (I1), contorno (I2), convexo (I3) y reflectividad (I4).

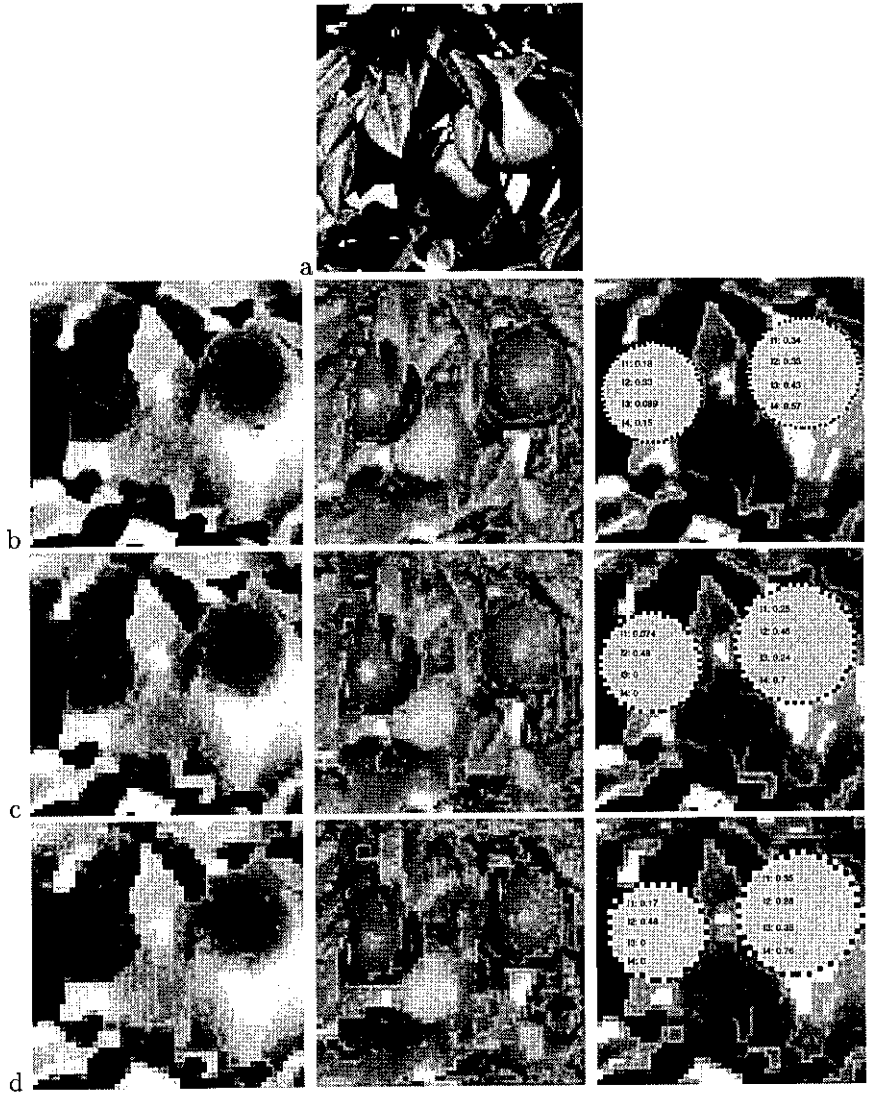


Figura C.8: a) Fotografía en B/N de árbol artificial con tres frutos. b),c) y d) Imágenes con resolución espacial de 2, 3 y 4 mm/puntos respectivamente (de izquierda a derecha: mapa de distancia, reflectancia y detecciones encontradas); se indica posición y radio estimados mediante anillos superpuestos en mapa de distancia con indicios parciales corona (I1), contorno (I2), convexo (I3) y reflectividad (I4).

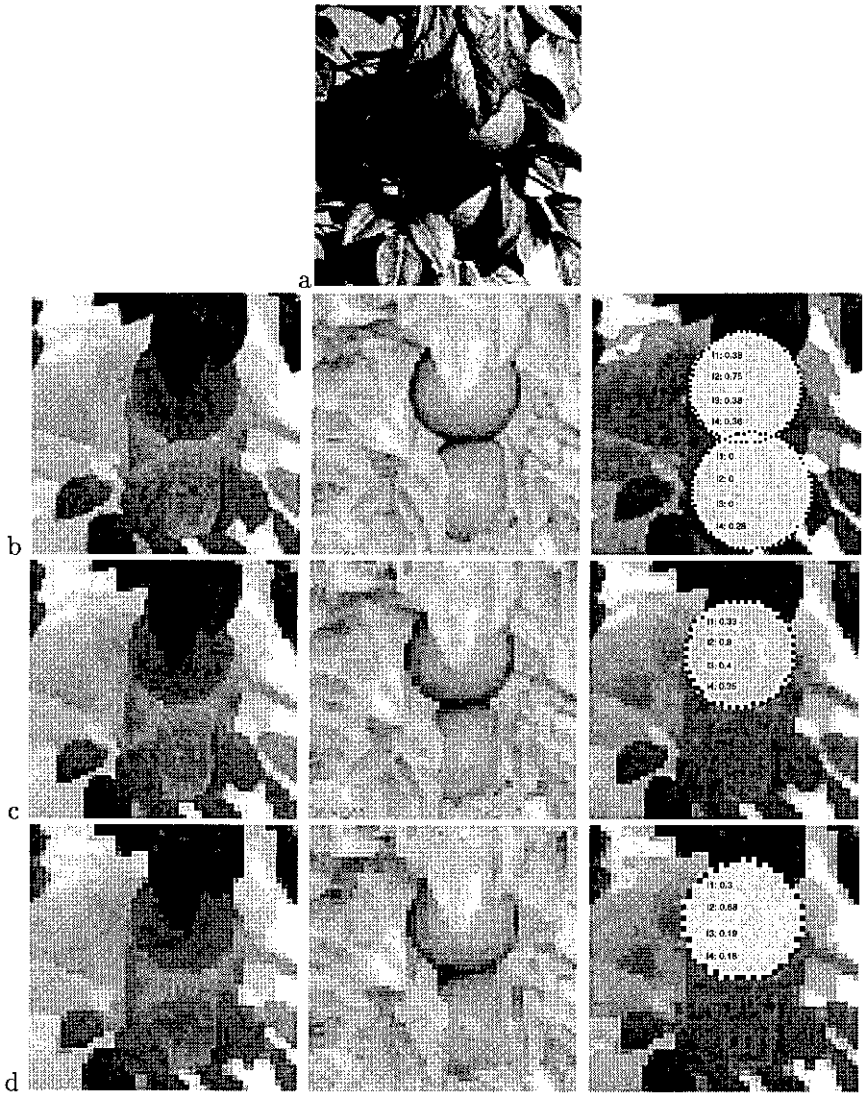


Figura C.9: a) Fotografía en B/N de árbol artificial con dos frutos. b),c) y d) Imágenes con resolución espacial de 2, 3 y 4 mm/puntos respectivamente (de izquierda a derecha: mapa de distancia, reflectancia y detecciones encontradas); se indica posición y radio estimados mediante anillos superpuestos en mapa de distancia con indicios parciales corona (I1), contorno (I2), convexo (I3) y reflectividad (I4).

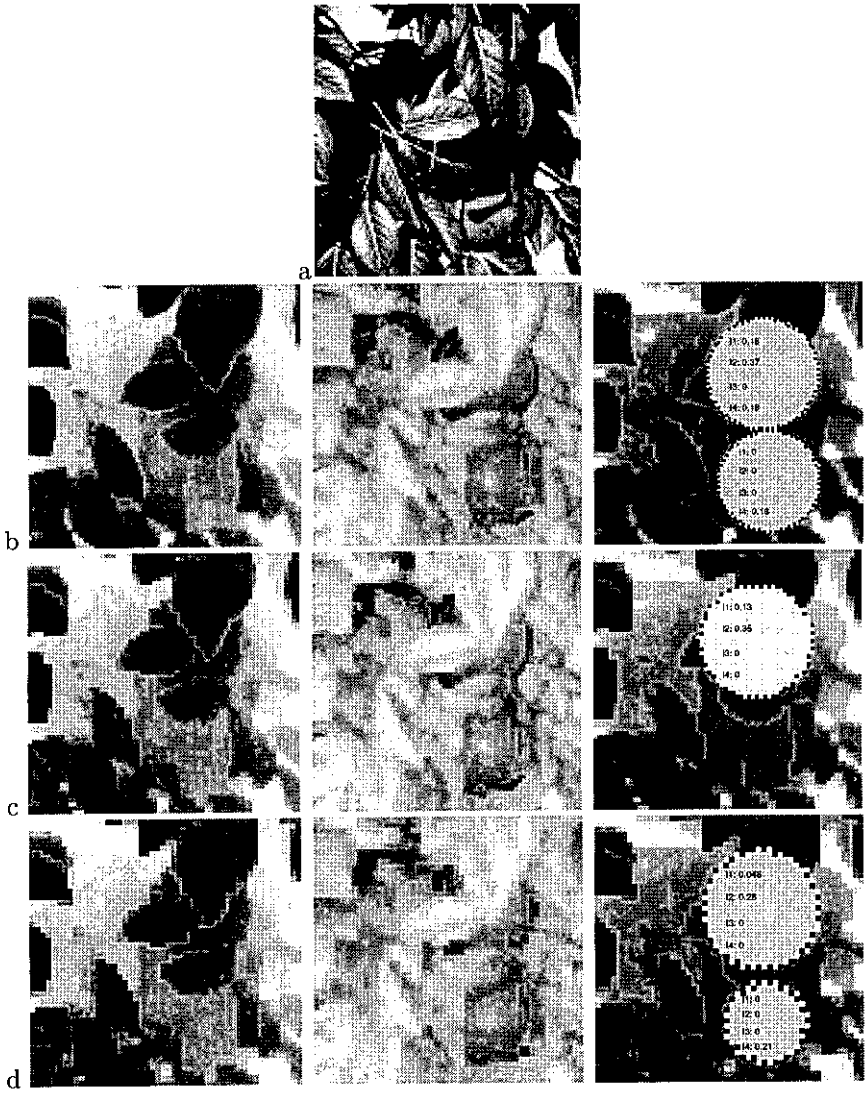


Figura C.10: a) Fotografía en B/N de árbol artificial con dos frutos. b),c) y d) Imágenes con resolución espacial de 2, 3 y 4 mm/puntos respectivamente (de izquierda a derecha: mapa de distancia, reflectancia y detecciones encontradas); se indica posición y radio estimados mediante anillos superpuestos en mapa de distancia con indicios parciales corona (I1), contorno (I2), convexo (I3) y reflectividad (I4).

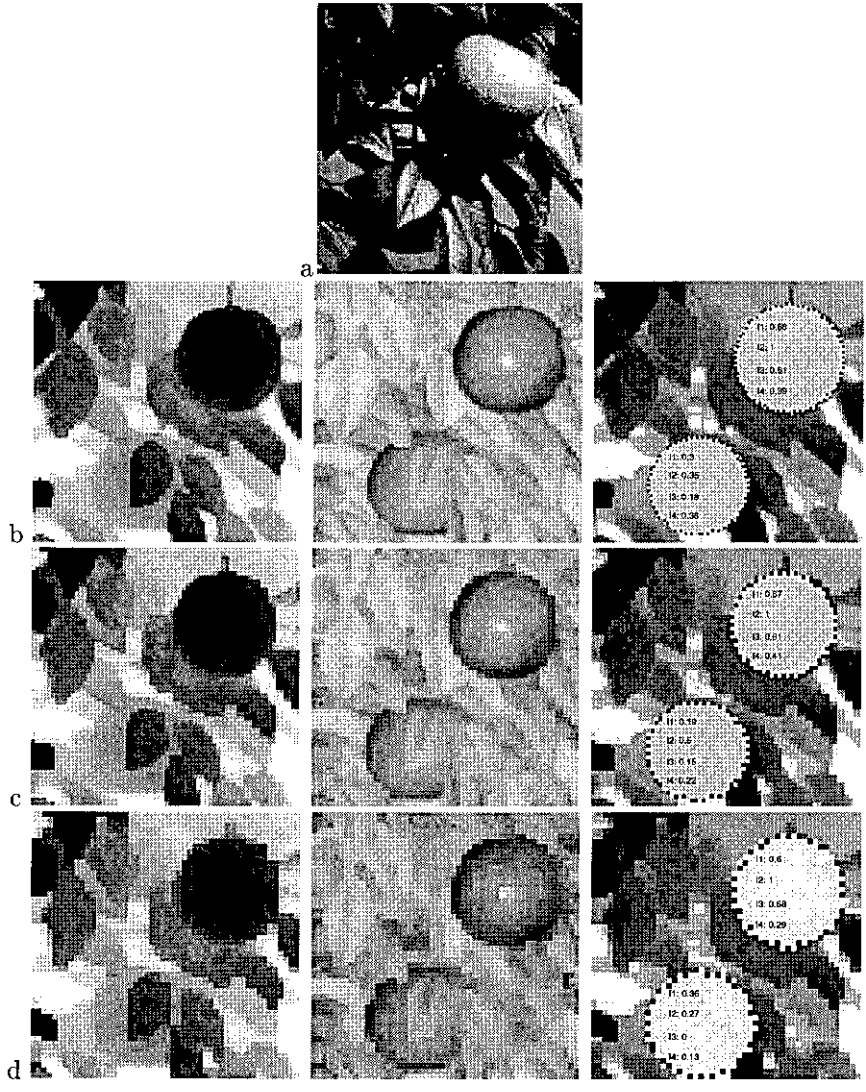


Figura C.11: a) Fotografía en B/N de árbol artificial con dos frutos. b),c) y d) Imágenes con resolución espacial de 2, 3 y 4 mm/puntos respectivamente (de izquierda a derecha: mapa de distancia, reflectancia y detecciones encontradas); se indica posición y radio estimados mediante anillos superpuestos en mapa de distancia con indicios parciales corona (I1), contorno (I2), convexo (I3) y reflectividad (I4).

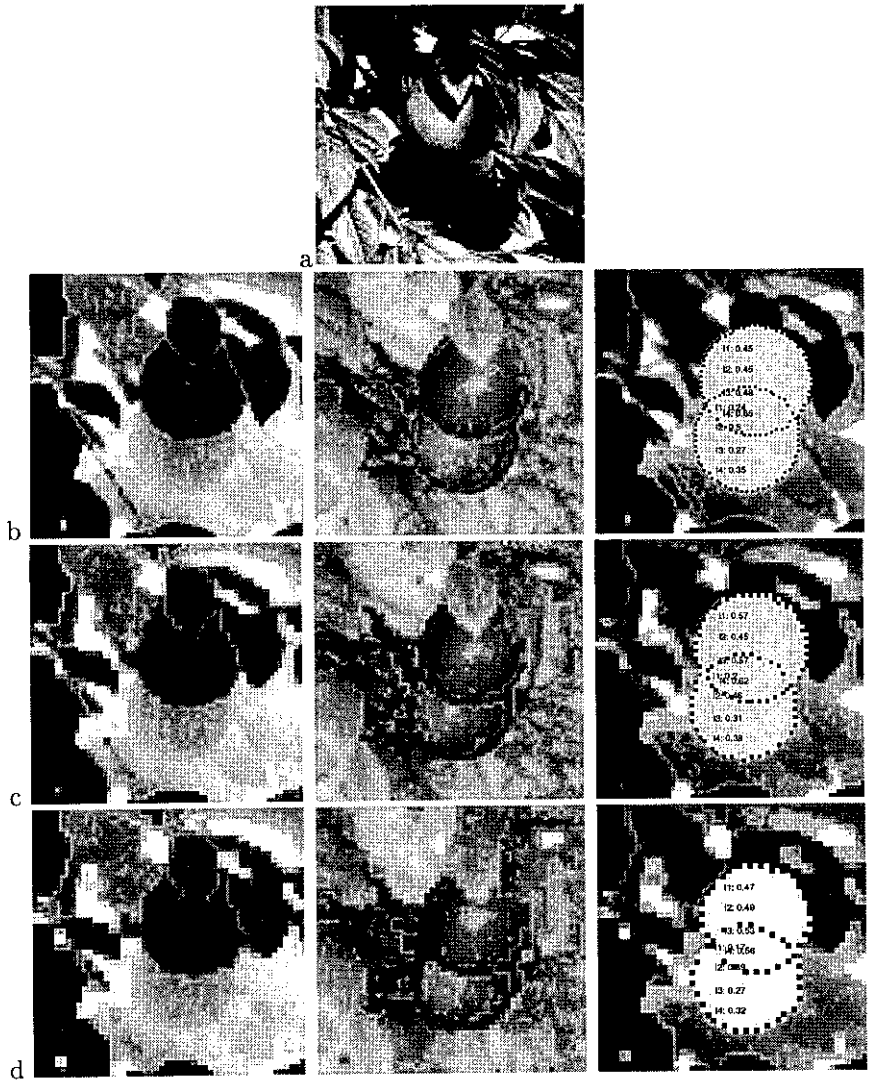


Figura C.12: a) Fotografía en B/N de árbol artificial con dos frutos. b),c) y d) Imágenes con resolución espacial de 2, 3 y 4 mm/puntos respectivamente (de izquierda a derecha: mapa de distancia, reflectancia y detecciones encontradas); se indica posición y radio estimados mediante anillos superpuestos en mapa de distancia con indicios parciales corona (I1), contorno (I2), convexo (I3) y reflectividad (I4).

Bibliografía

[1] L. Agapito. *Estrategias de correspondencia jerárqica y métodos directos de autocalibración para un sistema estereoscópico binocular.* Tesis doctoral. Univ. Complutense de Madrid., 1996.

[2] E. Al-Hujazi and A. Sood. Range image segmentation with applications to robot bin-picking using vacuum gripper. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 20(6):1313–1324, 1990.

[3] I. Amir. Algorithm for finding the center of circular fiducials. *Computer Vision, Graphics and Image Processing*, 49:398–406, 1990.

[4] M.C. Ayuso, F.J. Bravo, M. Ruiz-Altisent, and P. Barreiro. Sensing colour stability and mixtures of powder paprika using optical reflectance and image analysis. *AgENG96-Paper 96F-049*, 1996.

[5] R. Bajcsy. Active perception versus passive perception. *Proceedings Third IEEE Workshop on Vision*, pages 55–59, 1985.

[6] D.H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1991.

[7] R. Baribeau, M. Rioux, and G. Godin. Color reflectance modeling using a polychromatic laser range sensor. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14(2):263–269, 1992.

[8] J.M. Beaulieu and P.Boulanger. Segmentation of range images by piecewise approximaion with shape constraints. *Proc. of Vision Interface 88*, pages 19–24, 1988.

[9] P. Beckmann and A. Spizzichino. The scattering of electromagnetic waves from rough surfaces. *MacMillan, New York*, pages 1–33,70–98, 1963.

[10] M. Benady and G.E. Miles. Locating melons for robotic harvesting using structured light. *Paper ASAE No.:92-7021*, 1992.

[11] J.V. Benlloch, T. Heisel, S. Christensen, and A. Rodas. Image processing techniques for determination of weeds in cereal. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 195–200, 1997.

[12] J. Berkmann and T.Caelli. Computation of surface geometry and segmentation using covariance techniques. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(11):1114–1116, 1994.

[13] P.J. Besl. Active optical range imagining sensors. *Machine Vision and Applications, 1, See also: Advances in Machine Vision: Architectures and Applications, (J.Sanz, Ed.),Springer-Verlag, New York*, pages 127–153, 1988.

[14] P.J. Besl. *Surfaces in range image understanding.* Springer New York (N.V.), 1988.

[15] P.J. Besl. and R.C. Jain. Segmentation through variable-order surface fitting. *IEEE Transactions Pattern Analysis Machine Intelligence*, 10(2), 1988.

[16] J. Billingley and M. Schoenfisch. Automatic guiadance of agricultural vehicles. *Proc. National Conf. Australian Robot Assoc., Sydney*, pages 25–32, 1995.

[17] G. Blandini. First approaches to robot utilisation for automatic citrus harvesting. *Land and Water Use, Dodd Grace (eds.). Balkema, Rotterdam*, 1989.

[18] G. Blandini, E. Cerruto, and G. Manetto. Evaluation of the bruising of citrus fruits during post-harvest processing throught the use of computerized equiment. *AgENG96-Paper 96F-045*, 1996.

[19] A. Bourely. Rosal, a grafting robot for woody plants. pages 145–152, 1995.

[20] J.R. Brandon, S.W. Searcy, and R.J. Babowicz. Distributed control for vision based tractor guidance. *International Summer Meeting of the ASAE*, 89-7517, 1989.

[21] P.N. Brett and R.S. Stone. A tactile sensing technique for automatic gripping of compact shaped non-rigid materials. *IEE Colloquium on Intelligent automation for processing non-rigid products*, pages 1–5, 1994.

[22] J.R. Brodie, A.C. Hansen, and J.F. Reid. Size assesment of stacked logs via the hough transform. *Transactions of the ASAE*, 37(1):303–310, 1994.

[23] F. Buemi, G.M. Calvini, M. Massa, and G. Sandini. Quality control of agro-food products using colour and shape parameters. pages 23–35, 1995.

[24] F. Buemi, M. Massa, and G. Sandini. Agrobot: a robotic system for greenhouse operations. *4th Workshop on robotics in Agriculture, IARP, Tolouse*, pages 172–184, 1995.

[25] L. Calderón. Sensor ultrasónico adaptativo de medida de distancias. aplicación en el campo de la robótica. *Tesis doctoral. Universidad complutense de Madrid.*, 1984.

[26] Canny. A computational approach to edge detection. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.

[27] M. Cardenas, A. Hetzroni, and G.E. Miles. Machine vision to locate melons and guide robotic harvesting. *Paper ASAE No. 91-7006*, 1991.

[28] M.A. Castillo and A. Martínez. Foliage shaker to harvest valencia oranges in cuba. *AgENG96-Paper 96A-077*, 1996.

[29] R. Ceres. La automatización en el sector agroalimentario. *Tercer congreso de la asociación española de robótica AER*, pages 453–461, 1993.

[30] R. Ceres, J.L. Pons, A.R. Jiménez, J.M. Martín, and L. Calderón. Agribot: A robot for aided fruit harvesting. *AgENG96-Paper 96A-107*, 1996.

[31] F. Chaumette, S. Boukir, P. Bouthemy, and D. Juvin. Structure from controlled motion. *IEEE Transactions on Pattern analysis and machine intelligence*, 18 (5):492–504, 1996.

[32] K. Choi, G. Lee, Y.J. Han, and J.M. Bunn. Tomato maturity evaluation using color image analysis. *Transactions of the ASAE*, 38(1):171–176, 1995.

[33] H.I. Christensen and S.W. Frstner. Performance characteristics of vision algorithms. *Machine vision and applications*, 9:215–218, 1997.

[34] P.M. Clarkson and G.A. Williamson. Order statistics and adaptive filtering. *Signal processing methods for audio, images and telecomunication*, pages 109–141, 1995.

[35] R.R. Coifman and D. Donoho. Translation-invariant de-noising. *Wavelets and Statistics*, pages 125–150, 1995.

[36] R.L. Cook and K.E. Torrance. A reflectance model for computer graphics. *ACM Transactions on Graphics*, 1(1):7–24, 1982.

[37] G.E. Coppock, D.B. Churchill, and S.L. Hedden. Shaker stroke affects selective removal of valencia oranges. *Transactions of the ASAE*, 28:1094–1096, 1985.

[38] P. Courtney, N. Thacker, and A. Clark. Algorithmic modeling for performance evaluation. *Workshop on Performance Characteristics of Vision Algorithms. Robin College, Cambridge, April 19*, 1996.

[39] K.C. Cox, G.C. Roman, W.E. Ball, and A.F. Laine. Rapid search for spherical objects in aereal photographs. *IEEE Computer society on Computer Vision and Pattern Recognition*, pages 905–909, 1988.

[40] R.L. Cromwell. Sensors and processors enable robots to see and understand. *Laser focus world*, pages 67–78, 1993.

[41] T.G. Crowe and M.J. Delwiche. Real-time defect detection in fruit- part i: Design concepts and development of prototype hardware. *Transactions of the ASAE*, 39(6):2299–2308, 1996.

[42] T.G. Crowe and M.J. Delwiche. A system for fruit defect detection in real-time. *AgENG96-Paper 96G-023*, 1996.

[43] P. Dario, M. Bergamasco, and A. Fiorillo. Force and tactile sensing for robots. *NATO ASI Series, Vol F43. Sensors and Sensory Systems for Advanced Robots*, pages 153–185, 1988.

[44] P. Dario, A. Bicchi, A. Fiorillo, G. Buttazzo, and R. Francesconi. A sensorised scenario for basic investigation on active touch. *Robot sensors: Tactile and non-vision*, 2:237–245, 1986.

[45] A.G. D'Esnon, G. Rabatel, and R. Pellenc. Magali: A self-propeled robot to pick apples. *ASAE paper 87-1037, ASAE, St Joseph, MI 49085-9659*, 1987.

[46] Y. Dobrousin and Y. Edan. Real-time image processing for robotic melon harvesting. *Paper ASAE No.:92-3515*, 1992.

[47] R.O. Duda and P.E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Comunications of the ACM*, 15(1):11–15, 1972.

[48] R.O. Duda, D. Nitzan, and P.Barrett. Use of range and reflectance data to find planar surface regions. *IEEE Pattern analysis and Machine intelligence*, 1(3):259–271, 1979.

[49] C.S. Dyson, N. Yauilla, and E.S. Kolesar. Object imaging with a piezoelectric robotic tactile sensor. *Proc. of NAECON*, 1:41–47, 1993.

[50] H.R. Everett. *Sensors for mobile robots. Theory and applications.* A.K. Peters, Ltd. Wellesley, Massachusetts, 1995.

[51] O. Fagueras. *Three-Dimensional computer Vision: A geometric viewpoint.* The MIT Press, 1993.

[52] S. Felici, J. Pelegri, G. Sánchez, and J. Pelechano. Virtual centralized control system for intelligent irrigation in distributed agrarian exploitations. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 213–216, 1997.

[53] W. Forstner. 10 pros and cons against performance characterization of vision algorithms. *Workshop on Performance Characteristics of Vision Algorithms. Robin College, Cambridge, April 19*, 1996.

[54] H. Frigui and R. Krishnapuram. Clustering by competitive agglomeration. *Pattern Recognition*, 30(7):1109–1119, 1997.

[55] T. Fujiura, J. Yamashita, and N. Kondo. Agricultural robots: Vision sensing system. *Paper N0. 92 3517*, pages 1–7, 1992.

[56] J.J. Gaffney. Reflectance properties of citrus fruit. *Transactions of the ASAE*, 16(2):310–314, 1969.

[57] M. Galindo, J.A. López, L.A. Contreras, and L.M. Tomás. Defects modeling through artificial vision techniques, applied to satsuma and tangerine slices quality control. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 89–94, 1997.

[58] M.C. García-Alegre, J. Enciso, A. Ribeiro, and D. Guinea. To an automatic visual inspection of egg-shell defects. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 51–56, 1997.

[59] M.C. García-Alegre and D. Guinea. Building an architecture for a farming robot. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 255–260, 1997.

[60] G. Gering. Linking image-space and acumulator-space: A new approach for object recognition. *1st IEEE Int. Conf. On Computer Vision, London*, pages 112–117, 1987.

[61] A. Ghazanfari and J. Irudayaraj. Classification of pistachio nuts using a string matching technique. *Transactions of the ASAE*, 39(3):1197–1202, 1996.

[62] A. Ghazanfari, J. Irudayaraj, A. Kusalik, and M. Romaniuk. Machine vision grading of pistachio nuts using fourier descriptors. *Jounal of Agricultural Engng Res.*, 68:247–252, 1997.

[63] S. Ghosal and R. Mehrotra. Range image segmentation using zernike moment-based generlized edge detector. *IEEE international conference on Robotics and Automation. Nice, France*, pages 1584–1589, 1992.

[64] B. Gil, A. Mitiche, and J.K. Aggarwal. Experiments in combining intensity and range edge maps. *Computer Vision Graphics and Image Processing*, 21(3), 1983.

[65] J. Gómez, F.J. Díaz, and J. López. An approach for determining bidirectional reflectance parameters from range and brightness data. *Proc. of the IEEE Int. Conf. On Image Processing. Lausanne (Switzerland)*, 1:41–44, 1996.

[66] J. Gómez, F.J. Díaz, and J. López. Obtención conjunta de las informaciones tridimensional y cromática. primera aproximación a la caracterización óptica de superficies. *Informática y Automática*, 30(3):19–33, 1997.

[67] F.E. Goodwin. Coherent laser radar 3-d vision sensor. *SME. Society of Manufacturing Engineers. MS85-1005*, page 15, 1985.

[68] A.R. Grahn and L. Astle. Robotic ultrasonic force sensor arrays. *Robot sensors: Tactile and non-vision*, 2:297–315, 1986.

[69] D. Greenhill and E.R. Davies. Relative effectiveness of neural networks for image noise suppression. *Pattern recognition in Practice*, 4:367–378, 1994.

[70] W.E.L. Grimson. Discontinuity detection for visual surface reconstruction. *Computer vision, Graphics and Image Processing*, 30:316–330, 1985.

[71] T. Hague and N.D. Tillet. Navigation and control of an autonomous horticultural robot. *Mechatronics*, 6(2):165–180, 1996.

[72] R.M. Haralick. Computer vision theory: the lack thereof. *Computer Vision: Representation and Control*, pages 113–121, 1985.

[73] R.M. Haralick and L.G. Shapiro. *Computer and robot vision*, volume 1. Addison Wesley, 1992.

[74] R.M. Haralick and L. Watson. A facet model for image data. *Computer Graphics Image Proccesing*, 15:113–129, 1981.

[75] R.C. Harrell. Economic analysis of robotic citrus harvesting in florida. *Transactions of the ASAE*, pages 298–304, 1987.

[76] R.C. Harrell, D.C. Slaughter, and P.D. Adsit. A fruit-tracking system for robotic harvesting. *Machine Vision and Aplications*, 2:69–80, 1989.

[77] R. Harries. Sensing down on the farm. *Sensor review, Abril*, pages 102–104, 1983.

[78] K. Hatou, T. Morimoto, J. De Jager, and Y. Hashimoto. Measurement and recognition of 3-d body in intelligent plant factory. *AgENG96-Paper 96F-027*, 1996.

[79] K. Hatou, T. Sugiyama, and Y. Hashimoto. Three-dimensional measurement of small crops for bio-technological applications. *Proc. Int. Conf. on Virtual Systems and Multimedia*, pages 284–289, 1995.

[80] F.V. Hatzivasiliou and S.G. Tzafestas. Analysis and design of anew piezoresistive tactile sensor system for robotic applications. *Jounal of Intelligent and Robotic Systems*, 10:243–256, 1994.

[81] F. Heijden. Edge and line feature extraction based on covariance models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(1):16–33, 1995.

[82] R. Hoffman and A.K. Jain. Segmentation and classification of range images. *IEEE Transactions on pattern analysis and machine intelligence*, 9(5):608–620, 1987.

[83] A. Hoover. An experimental comparison of range image segmentation algorithms. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 18(7):673–689, 1996.

[84] B.K.P. Horn. Understanding image intensities. *IEEE Computer Society Press. Computer Vision: Principles*, pages 10–35, 1991.

[85] R.D. Howe. Tactile sensing and control of robotic manipulation. *Advanced Robotics*, 8(3):245–261, 1994.

[86] J.N. Huddleston and J. Ben-aire. Grouping edgels into structural entities using circular symmetry, the distributed hough transform, and probabilistic non-accidentalness. *CVGIP: Image Understanding*, 57-2:227–242, 1993.

[87] H. Hwang, S.C. Kim, J.H. Chang, and K.D. Ko. Development of automatic grafting system for fruit vegetables. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 13–19, 1997.

[88] H. Hwang, S.C. Kim, C.H. Lee, and C.K. An. Real-time on-line implementation of automatic grading and sorting of dried oak mushrooms- 2nd prototype. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 223–228, 1997.

[89] IEC-1984. Radiation safety of laser products, equipment classification, requirements and user's guide. *International electrotechnical commision. IEC Standard*, 825:1–137, 1984.

[90] J. Illingworth and J.Kittler. A survey of the hough transform. *Computer Vision, Graphics and Image Processing*, 44:87–116, 1988.

[91] J. Illingworth and J. Kittler. The adaptive hough transform. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 9(5):690–697, 1987.

[92] J. Illingworth, J. Kittler, and J. Princen. Shape detection using the adaptive hough transform. *NATO ASI Series, Sensors devices and Systems for Robotics*, 42:119–142, 1988.

[93] K. Inoue, K. Otsuka, M. Sugimoto, and N. Murakami. Estimation of place of tractor and adaptive control method of autonomus tractor using ins and gps. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 27–32, 1997.

[94] D.W. Jacobs. Robust and efficient detection of salient convex groups. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(1):23–37, 1996.

[95] J. Jager, K. Hatou, T. morimoto, and Y. Hashimoto. Recognition of damaged apples based on 3-dimensional range image. *Proc. of 6th Int. Congress for Computer Technology in Agriculture*, pages 562–566, 1996.

[96] A.K. Jain and P. Flynn. *Three-dimensional Object Recognition Systems.* Elsevier, 1993.

[97] A.K. Jain and P.J. Flynn. Image segmentation using clustering. *IEEE Image understanding*, pages 65–83, 1996.

[98] A.K. Jain, T. Newman, and M. Goulish. Range-intensity histogram for segmenting ladar images. *Pattern Recognition Letters*, 13:41–56, 1992.

[99] R.C. Jain and T.O. Bindford. Ignorance, myopia, and naivete in computer vision systems. *CVGIP: Image Understanding*, 53:112–117, 1991.

[100] R.C. Jain and A.K. Jain. *Analysis and interpretation of range images.* Springer New York (N.V.), 1990.

[101] R.A. Jarvis. A laser time-of-flight range scanner for robotic vision. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 5(5):505–512, 1983.

[102] R.A. Jarvis. A perspective on range finding techniques for computer vision. *Trans. IEEE Pattern Analysis and Machine inteligence*, 5(2):122–139, 1983.

[103] B.V. Jayawant, M.A. Onori, and J.D. Watson. Robot tactile sensing: a new array sensor. *Robot sensors: Tactile and non-vision*, 2:199–205, 1986.

[104] X. Jiang and H. Bunke. Robust edge detection in range images based on scan line approximation. *Proc. 4th Asian Conference on Computer Vision*, 1997.

[105] X.Y. Jiang, U. Meier, and H. Bunke. Fast range image segmentation using high-level segmentation primitives. *Proc. 3rd IEEE Workshop on Applications of Computer Vision. WACV. Sarasota, Florida*, pages 83–88, 1996.

[106] F.J. Jimeno and J.J. López. An artificial vision system for quality control and sorting in citric industry. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 95–100, 1997.

[107] A.R. Jiménez, R. Ceres, and J.L. Pons. Featureless classification of tactile contacts in a gripper using neural networks. *Sensors and Actuators, A*, 62:488–491, 1997.

[108] A.R. Jiménez, R. Ceres, and J.L. Pons. Shape-based methods for fruit recognition and localization using a laser range-finder. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, 1997.

[109] A.R. Jiménez, R. Ceres, and J.L. Pons. A new adaptive filter and a quality evaluation index for image restoration. *Informe interno 001-SAM. IAI-CSIC (submitted to Computer Vision and Image Understanding)*, 1:1–30, 1998.

[110] G. Johnson. Development of a practical scanninig laser radar sensor for robotic bin-picking, autonomous guiadance, and other difficult image analysis tasks. *22nd International Symposium on Industrial Robots. Detroit, Michigan, USA*, pages 21/55–21/71, 1991.

[111] F. Juste. *Memoria de actividades 1993/1994.* Instituto Valenciano Investigaciones Agrarias. IVIA, 1994.

[112] F. Juste, I. Fornes, F. Plá, E. Molto, and F. Blay. Primeros resultados en campo de un prototipo de brazo robotizado para la recolección de cítricos. *CIMA91, 23 Conf. Int. Maquinaria Agrícola, Zaragoza*, pages 433–440, 1991.

[113] F. Juste, I.Fornes, and F. Plá. An approach to robotic harvesting of citrus in spain. *VII Int. Citrus Congress, Acineale, Italia*, 1992.

[114] F. Juste and F. Sevila. Citrus: A european project to study the robotic harvesting of oranges. *Proccedings of the 3rd International Symposium on Fruit, Nut and Vegetable Harvesting Mechanization. Denmark-Sweden-Norway*, 1991.

[115] A.C. Kak and J.L. Edwards. Experimental state of the art in 3d object recognition and localization using range data. *Proc. Workshop on Vision for Robots in IROS'95 Conference, Pittsburgh*, 1995.

[116] H. Kalviainen and P. Hirvonen. An extension to the randomized hough transform exploiting connectivity. *Pattern Recognition Letters*, 18:77–85, 1997.

[117] L. Kassay. Hungarian robotic apple harvester. *Paper ASAE No. 92-7042 St. Joseph, MI 49085*, pages 1–14, 1992.

[118] S. Kaveti, K. Teoh, and H. Wang. Second-order implicit polynomials for segmentation of range images. *Pattern Recognition*, 29(6):937–949, 1996.

[119] N. Kawamura, N. Kiyoshi, F.Tasteshi, and U. Motonobu. Study of fruit harvesting robot and its application on others works. *Proceedings Int. Symp. on Agricultural Mech. and Int. Coop. in High Tech. Era, Japanese Society of Agricultural Machinery*, 1987.

[120] C. Kimme, D.H. Ballard, and J. Sklansky. Finding circles by an array of accumulator. *Comunications of the ACM*, 18(2):120–122, 1975.

[121] N. Kondo, M. Monta, and K.C. Ting. Harvesting robot for inverted single truss tomato production systems. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 149–154, 1997.

[122] E. Krotkov and R. Hoffman. Terrain mapping for a walking planetary rover. *IEEE Transactions on Robotics and Automation*, 10(6):728–740, 1994.

[123] P.J. Kyberd and P.H. Chappell. A force sensor for automatic manipulation based the hall effect. *Meas. Sci. Technol.*, 4:281–287, 1993.

[124] W.C.Y. Lam and S.Y. Yuen. Efficient technique for circle detection using hypothesis filtering and hough transform. *IEE Proc. Vis. Image Signal Process.*, 143-5:292–300, 1996.

[125] C. Lee, R. Haralick, and K. Deguchi. Estimation of curvature from sampled noisy data. *CVPR'93. New York City, June 15-17*, pages 536–541, 1993.

[126] J.S. Lee. Digital image enhacement and noise filtering by use local statistics. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 2:165–168, 1980.

[127] J.S. Lee. Refined filtering of image noise using local statistics. *Computer graphics and image processing*, 15:380–389, 1981.

[128] J.S. Lee. Digital image smoothing and the sigma filter. *Computer Vision, Graphics and Image proceesing*, 24:255–269, 1983.

[129] M. Lefebvre, T. Zimmerman, and C. Baur. Potato operation: automatic detection of potato diseases. *Proceedings of SPIE 2345*, pages 2–9, 1994.

[130] P. Levi, R. Falla, and R. Pappalardo. Image controled robotics applied to citrus fruit harvesting. *Procedures ROVISEC-VII, Zurich*, 1988.

[131] Y. Li and W.C.K. Wong. An efficient and robust method to detect object centre. *Pattern Recognition*, 30-5:659–671, 1997.

[132] F.R. Livingstone, L. King, J-A. Beraldin, and M. Rioux. Development of a real-time laser scanning for object recognition, inspection, and robot control. *SPIE Vol. 2057*, pages 254–461, 1993.

[133] L.C. Lynnworth. *Ultrasonic measurements for process control. Theory, techniques and Applications.* Academic Press, Inc. Harcourt Brace Jovanovich, 1989.

[134] A. Mandow, J. Gomez, A. Ollero, and J.l. Martinez. Geenhouse operation with the aurora mobile robot. pages 185–191, 1995.

[135] C. Maraval. *Reconocimiento de formas y visión artificial.* Ra-ma, 1993.

[136] J.A. Marchant, T. Hague, N. Tillet, and J.M. Sanchiz. Research on an autonomous vehicle for precise plant treatments. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 237–242, 1997.

[137] J.A. Marchant, C.M. Onyango, and M.J. Street. Computer vision for potato inspection without singulation. *Computers and electronics in Agriculture*, 4:235–244, 1990.

[138] B. Marx. Laser and electro-optic based sensors. *Sensor review*, 16(3):12–15, 1996.

[139] J. Matas, Z. Shao, and J. Kittler. Estimation of curvature and tangent direction by median filtered differencing. *8th int. Conf. On Image Analysis and Processing, San Remo 13-15 Sept.*, 1995.

[140] P. Meer, R. Park, and K. Cho. Multiresolution adaptive image smoothing. *CVGIP: Graphical Models and Image processing*, 56(2):140–148, 1994.

[141] D. Mintz. Robust consensus based edge detection. *CVGIP: Image Understanding*, 9(2):137–153, 1994.

[142] A. Misiti. *Neural Network Toolbox User Guide.* The Math Works, 1996.

[143] A. Mitiche and Aggarwal J.K. Detection of edges using range information. *Trans. IEEE Pattern Analysis and Machine Intelligence*, 5(2):174–178, 1983.

[144] E. Moltó, N. Aleixos, L.A. Ruiz, J. Vazquez, and F. Juste. An artificial vision system for fruit quality assessment. *AgENG96-Paper 96F-078*, 1996.

[145] E. Moltó, S. Castillo, and J. Juste. Estudios sobre sistemas para la recolección automática de cítricos. *Fruticultura Profesional*, 24:35–41, 1987.

[146] S. Monchaud. Contribution to range finding techniques for third generation robots. intelligent autonomous systems. *An International conference, Amsterdam, Netherlands, 8-11 December*, pages 459–469, 1986.

[147] J. Monfort, F. Plá, J. Calpe, and D. Gallego. Reliable low-cost vision system for fruit grading. pages 39–45, 1995.

[148] N. Murakami, K. Otsuka, K. Inoue, and Sugimoto. Robotic cabbage harvester. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 131–136, 1997.

[149] H.R. Myler and A.R. Weeks. *The pocket handbook of image processing algorithms in C.* Prentice Hall, 1993.

[150] S.G. Nadabar and A.K. Jain. Edge detection and labeling by fusion of intensity and range images. *SPIE Proc. On Applications of artificial Intelligence Machine Vision and Robotics. Orlando*, 1992.

[151] Y. Nagasaka, R. Otani, K. Shigeta, and K. Taniwaki. Automated operation in paddy fields with fiber optic gyro sensor and gps. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 21–26, 1997.

[152] T. Nagata and H.B. Zha. Determining orientation, location and size of primitive surfaces by a modified hough transform technique. *Pattern Recognition*, 21:481–491, 1988.

[153] L. Najman and M. Schmitt. Geodesic saliency of watershed contours and hierarchical segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1163–1173, 1996.

[154] S.K. Nayar, M. Watanabe, and M. Noguchi. Real-time focus range sensor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1186–1197, 1996.

[155] D. Nitzan, A.E. Brain, and R.O. Duda. The measurement and use of registered reflectance and range data in scene analysis. *Proc. IEEE*, 65:206–220, 1977.

[156] N. Noguchi, K. Ishii, and H. Terao. Development of agricultural mobile robot using geomagnetic direction sensor and image sensors. *AgENG96-Paper 96G-040*, 1996.

[157] A. Ollero, J.L. Martínez, and A. Simón. The autonomous robot for spraying: Aurora. *Fourth Workshop on Robotics in Agriculture and the Food Industry, IARP, Toulouse*, pages 193–198, 1995.

[158] C.F. Olson. Decomposition of the hough transform: Curve detection with efficient error propagation. *Computer Vision ECCV'96. Proceedings 4th European Conference on Computer Vision. Cambridge, U.K.*, 1:263–272, 1996.

[159] D.C. Pao, H.F. Li, and R. Jayakumar. Shapes recognition using the straight line hough transform: Theory and generalization. *IEEE Trans. Pattern Analysis and Machine Intelligence. Vol. 14.n11*, pages 1076–1089, 1992.

[160] E. Parrish and A.K. Goksel. Pictorial pattern recognition applied to fruit harvesting. *Transactions of the ASAE*, 20:822–827, 1977.

[161] K. Peleg. Sorting fruits and vegetables by firmness and maturity. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 57–64, 1997.

[162] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Trans. Int. Conf. PAMI*, 12:629–639, 1990.

[163] D.L. Peterson and G.K. Brown. Mechanical harvester for fresh market quality blueberries. *Transactions of the ASAE*, 39(3):823–827, 1996.

[164] F. Plá. Recognition of partial circular shapes from segmented contours. *CVGIP: Computer Vision and Image Understanding*, 63(2):334–343, 1996.

[165] F. Plá, F. Juste, and F. Ferri. Feature extraction of spherical objects in image analisys: an application to robotic citrus harvesting. *Computers and Electronics in Agriculture*, 8:57–72, 1993.

[166] J.L. Pons. *Sistema asistido de localización tridimensional de blancos mediante telemetría láser.* Máster en tecnologías de la Inf. en Fabricación. Universidad Politécnica de Madrid, Enero, 1995.

[167] J.L. Pons. *Metodologías y estrategias de compensación activa para la mejora del comportamiento cinemático y dinámico de robots específicos. Aplicación al recolector Agribot.* Tesis doctoral. Departamento de Física Aplicada III. F.C.C. Físicas. Universidad Complutense de Madrid, 1996.

[168] J.L. Pons, R. Ceres, and A.R. Jiménez. Mechanical design of a fruit picking manipulator: Improvement of dinamic behavior. *IEEE international conference on Robotics and Automation*, pages 969–974, 1996.

[169] W. K. Pratt. *Digital Image processing.* Wiley, 1991.

[170] G. Rabatel. A vision sistem for magali, the fruit picking robot. *Paper 88293, AGENG88, Int. Conf. Agricultural Engineering, Paris*, 1988.

[171] P. Radeva, J. Serrat, and E. Martí. A snake for model-based segmentation. *Proceedings 5th Int. Conf. On Computer Vision (ICCV'95), MIT, USA*, 1995.

[172] K.R. Rao and J. Ben-Aire. Non-orthogonal expansion for template matching and edge detection. *Signal Processing Method for audio, images and telecommunications. A.P. Edt. Clarkson & H. Stark*, pages 299–331, 1995.

[173] J. Rebman and K.A. Morris. A tactile sensor with electrooptical transduction. *Robot sensors: Tactile and non-vision*, 2:145–155, 1986.

[174] B.E. Roberston and A.J. Walkden. Tactile sensor system for robotics. *Robot Sensors: Tactile and non-vision*, 2:89–97, 1986.

[175] J.M. Roger, E. Moltó, G. Rabatel, and J. Blasco. Design of a robotized, non-chemical weed controller. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 229–236, 1997.

[176] L.A. Ruiz, E. Moltó, F. Juste, F.Plá, and R. Valiente. Location and characterization of the stem-calyx area on oranges by computer vision. *Journal Agricultural Engineering Research*, 64:165–172, 1996.

[177] R.A. Russell. *Robot tactile sensing.* Depart. of Elec. Eng., Monash Univ., Australia. Prentice Hall, 1990.

[178] P. Saint-Marc, J. Chen, and G. Medioni. Adaptative smoothing: A general tool for early vision. *IEEE Transactions on pattern Analysis and Machine Intelligence*, 13(6):514–529, 1991.

[179] P. Saint-Marc, J.L. Jezonin, and Medioni. A versatile pc-based range finding system. *IEEE Transactions on Robotics and Automation, 7(2)*, 1991.

[180] C.A. Salinas. *Control digital de tres ejes y módulos de potencia de motores de corriente continua.* Máster en tecnologías de la información. Universidad Politécnica de Madrid, 1994.

[181] M. Salotti, F.B. Garbay, and C. Garbay. Evaluation of edge detectors: Critics and proposal. *Workshop on Performance Characteristics of Vision Algorithms. Robin College, Cambridge, April 19*, 1996.

[182] Y. Sarig. Robotics of fruit harvesting: A state-of-the-art review. *Journal Agricultural Engineering Research*, 54:265–280, 1993.

[183] Y. Sato and M. Otsuki. Three-dimensional shape reconstruction by active rangefinder. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 15-17 June*, pages 142–147, 1993.

[184] C.E. Schertz and G.K. Brown. Basic considerations in mechanizing citrus harvest. *Transactions of the ASAE*, pages 343–346, 1968.

[185] P. Seitz. The robust recognition of object primitives using local axes of symmetry. *Signal Processing*, pages 89–108, 1989.

[186] Y. Shirai. Application of laser range finder to robot vision. *NATO ASI Series, Vol F52. Sensors devices and Systems for Robotics*, pages 313–322, 1989.

[187] A. Sites and M.J. Delwiche. Computer vision to locate fruit on a tree. *ASAE paper 85-3039, ASAE, St Joseph, MI 49085*, 1988.

[188] D. Slaughter and R.C. Harrel. Color vision in robotic fruit harvesting. *Transactions of the ASAE*, 30(4):1144–1148, 1987.

[189] D. Slaughter and R.C. Harrel. Discriminating fruit for robotic harvest using color in natural outdoor scenes. *Transactions of the ASAE*, 32(2):757–763, 1989.

[190] C. Steger. Extracting curvilinear structures: a differential geometric approach. *Proc. 4th European Conference on Computer Vision. ECCV'96*, 1, 1996.

[191] V. Stepanov. Mobile and manipulation robots for agriculture. pages 137–143, 1995.

[192] T. W. Sze and Y. H. Yang. A simple contour matching algorithm. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 3(6):676–678, 1981.

[193] H.D. Tagare and R.J.P. DeFigueiredo. A framework for the construction of reflectance maps for machine vision. *CVGIP: Computer Vision and Image Understanding*, 57(3):265–282, 1993.

[194] K. Tanie, K. Komoriya, M. Kaneko, S. Tachi, and A. Fujikawa. A high-resolution tactile sensor. *Robot sensors: Tactile and non-vision*, 2:189–205, 1986.

[195] Y. Tao, P.H. Heinemann, and Z. Varghese. Machine vision for color inspection of potatoes and apples. *Transactions of the ASAE*, 38(5):1555–1561, 1995.

[196] Y. Tao, C.T. Morrow, P.H. Heinemann, and H.J. Sommer. Fourier-based separation technique for shape grading of potatoes using machine vision. *Transactions of the ASAE*, 38(3):949–957, 1995.

[197] K. Tate and Z. Li. Depth map construction from range-guided multiresolution stereo matching. *IEEE Transactions on Systems, Man and Cybernetics., vol.24, no. 1*, pages 134–144, 1994.

[198] C. Thorpe, M.H. Hebert, T. Kanade, and S.A. Shafer. Vision and navigation for the carnegie-mellon navlab. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(3):362–373, 1988.

[199] R.D. Tillett. Image analysis for agricultural processes: A review of potential opportunities. *Journal Agricultural Engineering Research*, 50:247–258, 1991.

[200] T. Torii, T. Kanuma, T. Okamoto, and O. Kinati. Image analysis of crop row for agricultural mobile robot. *AgENG96-Paper 96G-038*, 1996.

[201] K.E. Torrance and E.M. Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal Optical Society*, 57:1105–1114, 1967.

[202] E. Trucco. On shape-preserving boundary conditions for diffusion smoothing. *IEEE International conference on Robotics and Automation. Nice, France*, pages 1690–1694, 1992.

[203] E. Trucco and R.B. Fisher. Experiments in curvature-based segmentation of range data. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(2):177–181, 1995.

[204] M. Umasuthan and A.M. Wallace. Outlier removal and discontinuity preserving smoothing of range data. *IEE Proceeding Vis. Image Signal Process.*, 143(3):191–200, 1996.

[205] P. Uthaisombut. *Detecting defects in cherries using machine vision.* Master of science. Dept. Computer Science, Michigan State University, 1996.

[206] J. Valente and C. Couto. intelligent soil moisture employing thermal and capacitive methods for irrigation control. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 177–182, 1997.

[207] L.M. van Kollengburg, P. Wennekes, and C. Wekhoven. Developement of a mechatronic system for automatic harvesting of cucumbers. *Bio-Robotics-97. International Workshop on robotics and automated machinery for bio-productions*, pages 143–148, 1997.

[208] Z. Varghese, C.T. Morrow, and P.H. Heinemann. Automated inspection of golden delicious apples using color computer vision. *ASAE paper 91-7002, ASAE, St Joseph, MI*, 1991.

[209] D. Varona. *Diseño de un sistema prensor para el desprendimiento de frutos.* Máster en Tecnologías de la Inf. en Fabricación. Universidad Politécnica de Madrid, Enero, 1997.

[210] J. Varona and J.J. Villanueva. Neural networks as spatial filters for image processing: Neurofilters. *CVC Tech. Report no. 007*, pages 1–11, 1996.

[211] G. Vosselman and R.M. Haralick. Performance analysis of line and circle fitting in digital images. *Workshop on Performance Characteristics of Vision Algorithms. Robin College, Cambridge, April 19*, 1996.

[212] M.A. Wani and B.G.Batchelor. Edge-region-based segmentation of range images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(3):314–319, 1994.

[213] Whittaker, Miles, Mitchell, and Gaultney. Fruit location in a partially ocluded image. *Transactions of the ASAE, 30*, pages 591–597, 1987.

[214] P.M. Will and K.S. Pennington. Grid coding: A preprocessing technique for robot and machine vision. *Proc. 2nd Int. Joint Conf. Pattern recognition*, pages 66–68, 1971.

[215] A.P. Witkin. Scale-space filtering. *IEEE Computer Society Press. Computer Vision: Principles*, pages 108–111, 1991.

[216] M. Worring and A.W.M. Smeulders. Digital curvature estimation. *CVGIP: Image understanding*, 58:366–382, 1993.

[217] Q. Yang. Finding stalk and calyx of apples using structured lighting. *Computers and electronics in Agriculture*, 8:31–42, 1993.

[218] X. Yu, D. Bui, and A. Krzyzak. Robust estimation for range image segmentation and reconstruction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(5):530–537, 1994.

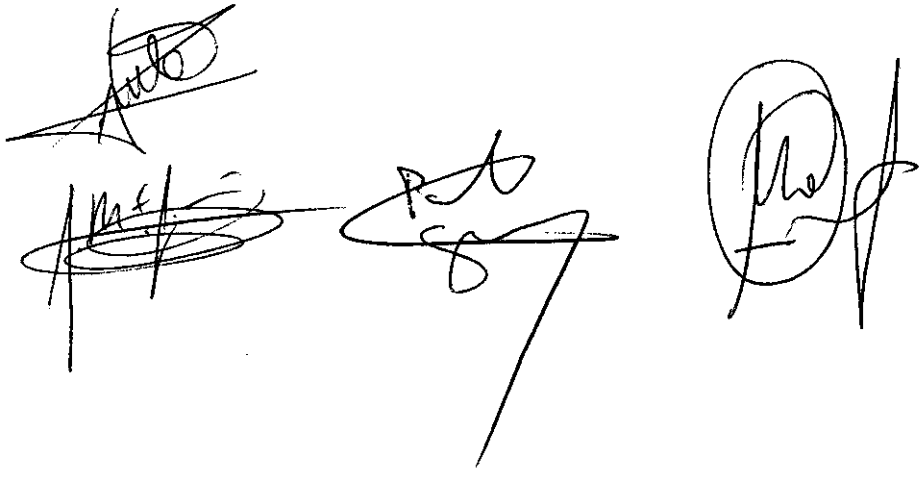
[219] P.C. Yuen and G.C. Feng. A novel method for parameter estimation of digital arc. *Pattern Recognition Letters*, 17:929–938, 1996.

[220] G. Zhang and A. Wallace. Physical modeling and combination of range and intensity edge data. *CVGIP: Image Understanding*, 58(2):191–220, 1993.

[221] Z. Zhang. Parameter estimation techniques: a tutorial with application to conic fitting. *Image and Vision Computing*, 15:59–76, 1997.

Reunido el Tribunal que suscribe en el día de la fecha acordó calificar la presente Tesis Doctoral con la censura de - Sobresaliente Cum laude por unanimidad

Madrid, 5 OCT 1998



* FIRMAN TODOS LOS MIEMBROS DEL TRIBUNAL