



Registro de Imágenes mediante Transformaciones Lineales por Trozos

Tesis Doctoral para optar al grado de Doctor en Informática
presentada por:

Vicente M. Arévalo Espejo

Dirigida por:

Javier González Jiménez

Málaga, Diciembre 2007



Publicaciones y
Divulgación Científica

AUTOR: Vicente Manuel Arévalo Espejo
 <http://orcid.org/0000-0003-0622-207X>

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



Esta obra está sujeta a una licencia Creative Commons:
Reconocimiento - No comercial - SinObraDerivada (cc-by-nc-nd):
[Http://creativecommons.org/licenses/by-nc-nd/3.0/es](http://creativecommons.org/licenses/by-nc-nd/3.0/es)
Cualquier parte de esta obra se puede reproducir sin autorización
pero con el reconocimiento y atribución de los autores.
No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer
obras derivadas.

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de
Málaga (RIUMA): riuma.uma.es



El Dr. Javier González Jiménez, director de la tesis titulada: **“Registro de Imágenes mediante Transformaciones Lineales por Trozos”** realizada por Vicente M. Arévalo Espejo certifica su idoneidad para la obtención del título de **Doctor en Informática**.

Málaga, Diciembre 2007

Javier González Jiménez

Lo que con mucho trabajo se adquiere, más se ama.

— Aristóteles (384 AC-322 AC) Filósofo griego.

A Carmen y Pablo.

Agradecimientos

Dedico estas líneas a todas aquellas personas, compañeros y amigos, que me han ayudado durante estos últimos años a llegar a donde estoy. A muchos de ellos, olvidaré citarlos, y no por ello, son menos importantes para mí. A todos ellos, mucha gracias.

A Javier, por confiar en mí hace casi 5 años, por su inestimable y gran ayuda, por su apoyo y comprensión en los malos momentos, que los ha habido y muchos, y sobre todo, por su paciencia. Sólo él y yo sabemos lo que significa esta tesis y el trabajo que ha supuesto para ambos. Sin él, no sería una realidad, por todo ello, Javier, muchas gracias.

A todos mis compañeros, los que están y los que se fueron, por los buenos momentos que hemos pasado juntos, por su apoyo y ayuda en estos años. A Juan Antonio, Ana, Goyo, Jose Luis, Paco, Tony, Cabello, etc. y en particular, a Cipri, amigo además de compañero, por sus consejos y saber escuchar, por aguantarme en las horas bajas. A él y a todos los demás. Muchas gracias.

A mis padres y hermana, por su amor y apoyo incondicional, por su sacrificio cuando las cosas no iban bien, por enseñarme a ser humilde y responsable, por estos últimos años, tan duros para todos. Os quiero y muchas gracias.

A Carmen y Pablo, las dos personas más importantes del mundo, mi universo, sin ellos, sin su amor, paciencia y sacrificio esta tesis no sería hoy una realidad. Os quiero mucho, perdón por estos últimos meses y muchas gracias.

A todos vosotros, esta tesis también es vuestra :-).

Resumen

Multitud de aplicaciones de la visión artificial necesitan comparar o integrar imágenes de un mismo objeto pero obtenidas en instantes de tiempo diferentes, con distintos dispositivos (cámaras), desde distintas posiciones, bajo distintas condiciones, etc. Estas diferencias en la captura dan lugar a imágenes con importantes diferencias geométricas relativas que impiden que éstas “encajen” con precisión unas sobre otras.

El registro elimina estas diferencias geométricas de forma que píxeles situados en las mismas coordenadas se correspondan con el mismo punto del objeto y, por tanto, ambas imágenes se puedan comparar o integrar fácilmente. El registro de imágenes es esencial en disciplinas como la teledetección, radiología, visión robótica, etc.; campos, todos ellos, que superponen imágenes para estudiar fenómenos medio-ambientales, monitorizar tumores cancerígenos o para reconstruir la escena observada.

En esta tesis se aborda la problemática asociada a este proceso, se analizan experimentalmente las técnicas de registro no-rígido más representativas. También se estudian diferentes medidas de similitud utilizadas para medir su consistencia y se propone un novedoso procedimiento para mejorar la precisión del registro lineal por trozos. Concretamente:

- Se analizan experimentalmente los elementos que influyen en la estimación de distribuciones de probabilidad de los niveles de intensidad de las imágenes. Estas distribuciones son la base para el cálculo de medidas de similitud basadas en la entropía como la información mutua (MI) o el coeficiente de correlación de entropía (ECC). Por tanto, la efectividad de estas medidas depende críticamente de su correcta estimación.

- Se compara experimentalmente las técnicas de registro no-rígido más representativas, en particular: funciones globales polinomiales (de diversos órdenes), lineal por trozos y funciones de base radial, en particular, la función *thin-plate-spline*. En esta comparativa se analizan la influencia en la precisión del registro de aspectos tales como el número y distribución de los puntos de control (CP) utilizados, la geometría de la escena o las posiciones de observación.
- Se formaliza el registro lineal por trozos mediante complejos simpliciales. Los complejos simpliciales son estructuras matemáticas basadas en triángulos que se emplean típicamente en topología algebraica y que ofrecen ventajas sobre las redes triangulares empleadas habitualmente en el registro lineal por trozos. Esta formalización plantea el registro más allá de un mero enfoque 2D, puesto que abarca, no sólo la generación de redes triangulares 2D óptimas (desde el punto de vista de la consistencia del registro), sino también formaliza la reconstrucción 3D de la superficie de la escena observada.
- Se propone un procedimiento de optimización de la consistencia del registro que modifica topológica y geoméricamente la red triangular (complejo simplicial) utilizada en el registro lineal por trozos. El método propuesto se compara experimentalmente con diversas funciones de transformación no-rígidas y procedimientos similares propuestos en el campo de la reconstrucción 3D. Por otro lado, para ilustrar los beneficios de utilizar redes triangulares consistentes en la reconstrucción de la escena observada, se incluyen diversos ejemplos generados a partir de las redes iniciales y las optimizadas.

En la realización de las pruebas y experimentos de esta tesis se han utilizado imágenes de distintos tipos: imágenes de alta resolución del satélite QuickBird (proporcionadas por Decasat Ingeniería www.decasat.com), imágenes sintéticas del repositorio ALOI [36], imágenes captadas con cámaras digitales domésticas (por ejemplo, de fachadas de edificios u objetos), etc.

Índice general

Agradecimientos	i
Resumen	iii
Índices	v
General	v
Figuras	ix
Tablas	xiii
Acrónimos	xv
1 Introducción	1
1.1 Registro de imágenes	1
1.2 La precisión en el registro	2
1.3 Definición del problema	6
1.3.1 Las bases del registro	8
1.3.2 La función de transformación	9
1.3.3 La medida de consistencia	11
1.3.4 La función de interpolación	12
1.3.5 El método de estimación	13
1.4 Contribuciones de la Tesis	14
2 Medidas de consistencia del registro	17
2.1 Medidas basadas en puntos	18
2.2 Medidas basadas en intensidad	20
2.3 Medidas basadas en la entropía	24

2.3.1	Información mutua	26
2.4	Estimación de la distribución de probabilidad	29
2.4.1	Evaluación experimental	31
2.4.2	La función de interpolación	34
2.5	Conclusiones	37
3	Registro basado en puntos	39
3.1	Transformaciones rígidas	40
3.2	Transformaciones no-rígidas	43
3.3	Comparativa experimental	47
3.3.1	Conjuntos de datos	49
3.3.2	Experimentos y resultados	52
3.4	Conclusiones	56
4	Registro lineal por trozos	59
4.1	Optimización de redes	61
4.2	Planteamiento del problema	64
4.3	Formalización del problema	67
4.3.1	Definiciones previas	68
4.4	Generación de las redes	73
4.5	Selección del conjunto de correspondencias	76
4.5.1	Detección y emparejamiento de esquinas	77
4.5.2	Selección robusta de correspondencias	79
4.6	Conclusiones	82
5	Optimización de redes triangulares	85
5.1	Definiciones	86
5.1.1	Elementos de la red	86
5.1.2	Acciones topológicas	88
5.2	Aplicación de las acciones	91
5.3	Descripción del método	96
5.3.1	La función de coste local	96
5.3.2	Método de optimización	99
5.4	Pruebas experimentales y comparativas	105

5.4.1	Datos y metodología	106
5.4.2	Comparativa con métodos similares	109
5.4.3	Comparativa con métodos tradicionales	111
5.4.4	Reconstrucción 3D	117
5.5	Conclusiones	117
6	Conclusiones y trabajos futuros	121
6.1	Conclusiones	121
6.2	Trabajos futuros	123
	Apéndices	126
A	Entropía, entropía relativa e información mutua	127
A.1	Entropía	127
A.2	Entropía conjunta y entropía condicional	130
A.3	Entropía relativa e Información mutua	133
A.4	Relación entre entropía e información mutua	135
B	Métodos de regresión robusta	139
B.1	LMedS	141
B.2	RANSAC	142
B.3	MAPSAC	143
C	Topología algebraica	145
C.1	Simplex	145
C.2	Complejo simplicial abstracto	147
C.3	Complejo simplicial o poliedro	148
C.3.1	Invariantes topológicos	152
C.3.2	Representación	153
D	Detectores de esquinas: Harris y Lowe	157
D.1	El detector de Harris	158
D.2	El detector de Lowe (SIFT)	160
	Bibliografía	164

Índice de figuras

1	Introducción	1
1.1	Aplicaciones del registro de imágenes	3
1.2	Más aplicaciones del registro de imágenes	4
1.3	Imágenes captadas desde distintos puntos de observación . . .	6
1.4	Imágenes de distinta naturaleza (multi-modalidad)	7
1.5	Esquema de un proceso genérico de registro de imágenes. . . .	8
1.6	Clasificación de las funciones de transformación.	10
1.7	Funciones de interpolación	13
2	Medidas de consistencia del registro	17
2.1	Medidas de similitud basadas en puntos	19
2.2	Histograma conjunto y dependencia estadística	24
2.3	Histograma de una imagen en escala de grises.	25
2.4	Información mutua y correlación cruzada normalizada	27
2.5	Estimación de una distribución de probabilidad mediante la ventana de Parzen	30
2.6	Número de bins y tamaños de muestra. Comparativa	32
2.7	Método paramétrico y la ventana de Parzen. Comparativa . .	33
2.8	Función densidad de probabilidad conjunta de dos imágenes .	34
2.9	Métodos de interpolación. Comparativa	35
2.10	Funciones de interpolación	36
3	Registro basado en puntos	39

3.1	Familia de transformaciones rígidas	41
3.2	Transformaciones polinomiales. Algunos ejemplos.	44
3.3	Registro lineal por trozos	45
3.4	Observación fuera del nadir	48
3.5	Imágenes de prueba utilizadas en el estudio comparativo . . .	51
3.6	Procedimiento utilizado en la selección de los CPs e ICPs . . .	52
3.7	Resultados obtenidos fijando el número de CPs, el ángulo de observación y la geometría de la escena	54
3.8	Resultados agrupados por ángulo de observación y geometría de la escena	55
4	Registro lineal por trozos	59
4.1	Ilustración de las transformaciones <i>thin-plate-spline</i> y lineal por trozos	60
4.2	Redes compatibles e incompatibles con la geometría de la escena	61
4.3	Acciones basadas en aristas empleadas para modificar la topología/geometría de una red dada.	63
4.4	Proyección para-perspectiva de la cámara	65
4.5	Registro lineal por trozos de imágenes afectadas de distorsión proyectiva y afín	66
4.6	Configuración escena-cámara sintética	69
4.7	Una red triangular en \mathbb{R}^3 que aproxima la escena de la figura 4.6.	70
4.8	Redes triangulares en \mathbb{R}^2 que aproximan las proyecciones de la escena de la figura 4.6 sobre los planos imagen de las cámaras.	71
4.9	Registro lineal por trozos de dos realizaciones geométricas . .	73
4.10	Cambios topológicos propuestos para eliminar <i>patch reversal</i> en las redes iniciales	75
4.11	Inconsistencias de la red producidas por el método de generación de redes propuesto	76
4.12	Proceso de selección robusta de correspondencias	83

5 Optimización de redes triangulares 85

5.1 Conjuntos de símlices característicos: star, closure, bound, etc 87

5.2 Acción *intercambiar arista* 88

5.3 Acción *dividir arista* 89

5.4 Acción *eliminar arista* 90

5.5 *Patch reversal* producido la eliminación de la arista 92

5.6 Inserción óptima de nuevos vértices en las redes triangulares . 93

5.7 Proceso de inserción de nuevos vértices en las redes triangulares 95

5.8 Comparación entre “intercambiar arista” y “dividir arista” . . . 97

5.9 Algoritmo *greedy* propuesto para mejorar el registro lineal por trozos de dos imágenes. 100

5.10 Ilustración del proceso de optimización propuesto 102

5.11 Partición binaria de una red triangular. 104

5.12 Imágenes reales de escenas poliédricas, sus correspondientes redes triangulares de Delaunay y las redes optimizadas 107

5.13 Evolución de la consistencia del registro global para las imágenes de la figura 5.12(a-d) 110

5.14 Red triangular optimizada con el método propuesto (todas las acciones) 112

5.15 Imágenes de prueba utilizadas en el estudio comparativo . . . 113

5.16 Evolución de la consistencia global del registro para las imágenes de la figura 5.15 116

5.17 Reconstrucciones 3D generadas a partir de las redes: iniciales y optimizadas, de la figura 5.12 118

5.18 Reconstrucción 3D generada a partir de la red optimizada de la zona residencial de la figura 5.15 119

Apéndices 126

A Entropía, entropía relativa e información mutua 127

A.1 Comparativa $H(X)$ versus p 130

A.2 Relación entre entropía e información mutua. 136

B	Métodos de regresión robusta	139
B.1	Comparativa del ajuste mínimo cuadrático y un procedimiento de estimación robusta.	140
C	Topología algebraica	145
C.1	n -símplices, con $n = 0, 1, 2, 3$	146
C.2	Transformación canónica de un 2-simplex estándar a un 2-simplex cualquiera.	147
C.3	Ilustración de un complejo simplicial abstracto de dimensión 2.	149
C.4	Ilustración de un complejo simplicial de dimensión 3.	150
C.5	Realización topológica y geométrica de un complejo simplicial.	151
C.6	Simplex abierto $ \sigma \subset \mathbb{R}^3$ vs. Simplex estándar $\Delta^n \subset \mathbb{R}^2$	152
D	Detectores de esquinas. Harris y Lowe	157
D.1	Proceso de detección de esquinas propuesto por Harris.	159
D.2	Proceso de construcción del espacio de escalas.	162
D.3	Proceso de búsqueda de extremos en las octavas construidas.	162
D.4	Descriptor de longitud $2 \times 2 \times 8 = 32$ obtenido a partir de regiones de 8×8 muestras.	163

Índice de tablas

5 Optimización de redes triangulares	85
5.1 Comparación del método propuesto con otros métodos de similares. Datos de las redes y tiempos	108
5.2 Comparación del método propuesto con otros métodos de registro no-rígidos	114
5.3 Datos de la red y tiempos empleados en el proceso de optimización	115
Apéndices	126
B Métodos de regresión robusta	139
B.1 Datos utilizados en la comparativa de la figura B.1.	140
C Topología algebraica	145
C.1 Matriz de incidencia de un complejo simplicial.	155
C.2 Matriz de 1-conexión (γ_1) de un complejo simplicial.	155

Acrónimos

CP	Punto de Control
DDT	Triangulación Dependiente de los Datos
DEM	Modelo Digital de Elevación
DTM	Modelo Digital del Terreno
ECC	Coefficiente de Correlación de Entropía
GCP	Punto de Control Terrestre
GIS	Sistema de Información Geográfica
GPS	Satélite de Posicionamiento Global
GPU	Unidad de Procesamiento Gráfico
ICP	Punto Independiente de Control
MI	Información Mutua
MLE	Estimación de Máxima Probabilidad
MR	Resonancia Magnética
NCC	Correlación Cruzada Normalizada
PCA	Análisis de Componentes Principales
PET	Tomografía por Emisión de Positrones
PIU	<i>Partitioned Intensity Uniformity</i>
RIU	<i>Ratio Image Uniformity</i>

RPC	Coeficientes de Polinomios Racionales
SAD	Suma de Diferencias en Valor Absoluto
SAR	Radar de Apertura Sintética
SIFT	<i>Scale Invariant Feature Transformation</i>
SSD	Suma de Diferencias al Cuadrado
SURF	<i>Speeded Up Robust Features</i>
TC	Tomografía Computarizada
TM	Mapa Temático

Capítulo 1

Introducción

1.1 Registro de imágenes

El *registro de imágenes* consiste en superponer dos imágenes de la misma escena adquiridas en diferentes instantes de tiempo (análisis multi-temporal), desde distintos puntos de vista (análisis multi-vista) y/o usando distintos sensores (análisis multi-modal). En este proceso, una de las imágenes permanece sin modificar (imagen de *referencia* o *fija*), mientras que la otra (imagen de *entrada* o *móvil*) se transforma geoméricamente hasta que se ajusta a la de referencia.

Este proceso es crucial en todas aquellas aplicaciones que necesitan combinar, comparar o fusionar información visual. Uno de los campos de aplicación es la teledetección donde, para poder monitorizar fenómenos medio-ambientales, como los efectos de unas inundaciones, o humanos, como el impacto del desarrollo urbano sobre una determinada región es necesario superponer con precisión imágenes de fechas diferentes para estudiar el nivel de cambio antes y después del fenómeno a evaluar.

Algo similar ocurre en medicina. Por ejemplo, para analizar diversos procesos biológicos, como la evolución de un tumor cancerígeno o una lesión muscular, se comparan resonancias magnéticas (MR) de la zona afectada adquiridas en distintas fechas (antes y después del tratamiento). También, para generar atlas médicos en los que se combinan imágenes de diferentes sujetos

de un determina área del cerebro.

En robótica es muy habitual utilizar cámaras para extraer información visual del entorno para detectar y/o manipular objetos, localizarse, desplazarse sin colisionar, reconstruir la escena, etc. Para realizar estas tareas se capturan imágenes desde distintas posiciones que se combinan o se comparan con otras adquiridas previamente.

Numerosos sistemas de seguridad basados en visión comparan imágenes adquiridas en distintos instantes de tiempo y/o desde distintos puntos de vista para detectar posibles intrusos en un área restringida. Por otro lado, la reciente aparición de sistemas de bajo coste capaces de captar imágenes retinales¹ ha favorecido la aparición de sistemas (biométricos) de control de acceso. Estos sistemas comparan la imagen retinal de un sujeto con las almacenadas en el sistema antes de autorizar su acceso.

Estas son sólo algunas de las muchas aplicaciones donde interviene el registro de imágenes. En las figuras 1.1 y 1.2 el lector puede encontrar otros ejemplos: generación de mosaicos, diagnóstico de enfermedades, compensación del desenfoque, reconstrucción 3D, etc.

1.2 La precisión en el registro

La precisión del registro es vital en la mayoría de las aplicaciones. Sirva como ilustración el siguiente ejemplo.

Supóngase que se pretende monitorizar el grado de desertización en una determinada región de interés mediante la comparación de mapas temáticos (TM) del satélite Landsat. Puesto que la resolución espacial de estas imágenes es de 25 m./píxel, cualquier error en el ajuste superior a un píxel desvirtuaría por completo los resultados obtenidos e induciría claramente a medidas incorrectas.

Aunque conceptualmente la superposición de dos imágenes es un problema relativamente sencillo, la precisión, y por tanto el éxito, de este proceso se ve condicionado por las condiciones de adquisición de las imágenes (el

¹Al igual que las huellas dactilares, las retinas (más concretamente, la disposición de los capilares retinales) de un individuo son únicos.

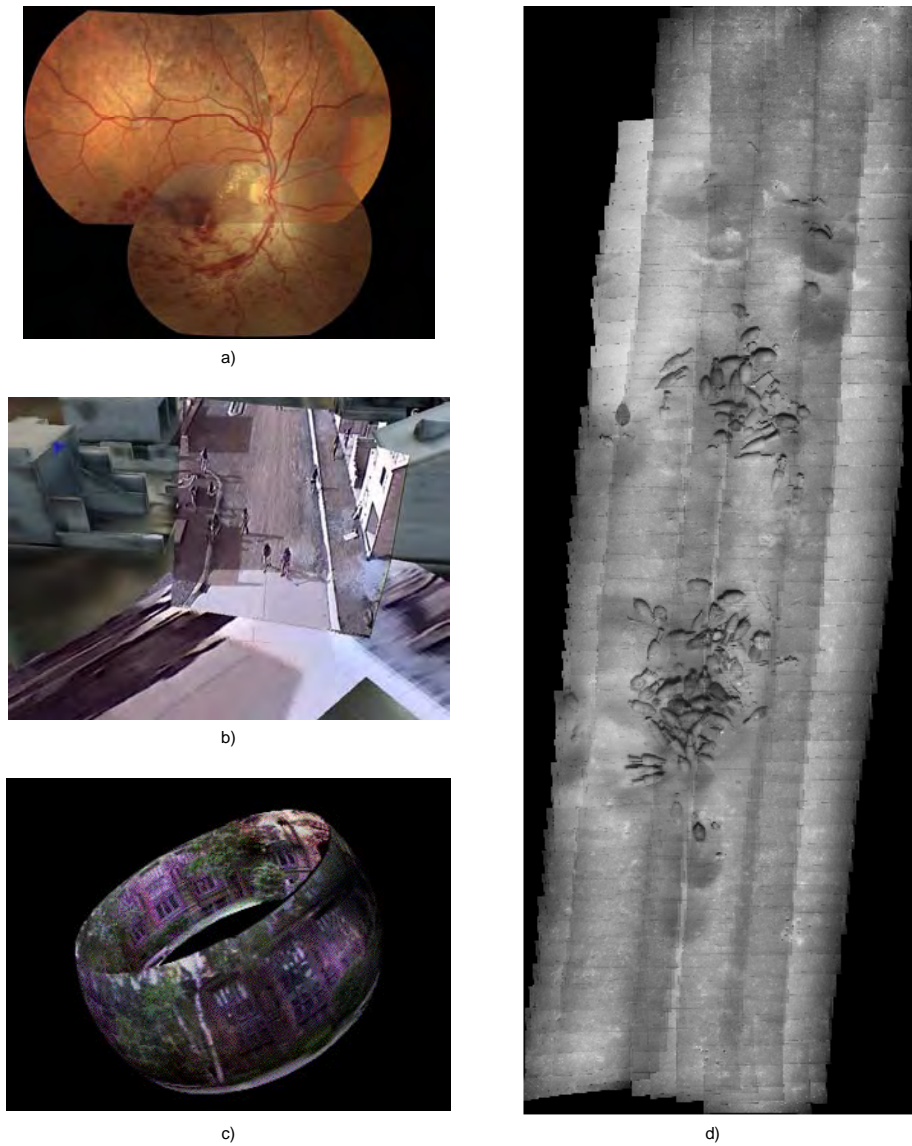


Figura 1.1: Aplicaciones del registro de imágenes: (a) Mosaico de imágenes retinales para el diagnóstico de enfermedades oculares; (b) Para ganar en realismo, los video juegos de reciente aparición proyectan vídeo sobre modelos 3D de la escena; (c) Mosaicos esféricos utilizados, típicamente, en realidad virtual y teleoperación; (d) Mosaicos de lechos submarinos para la realización de inventarios arqueológicos.

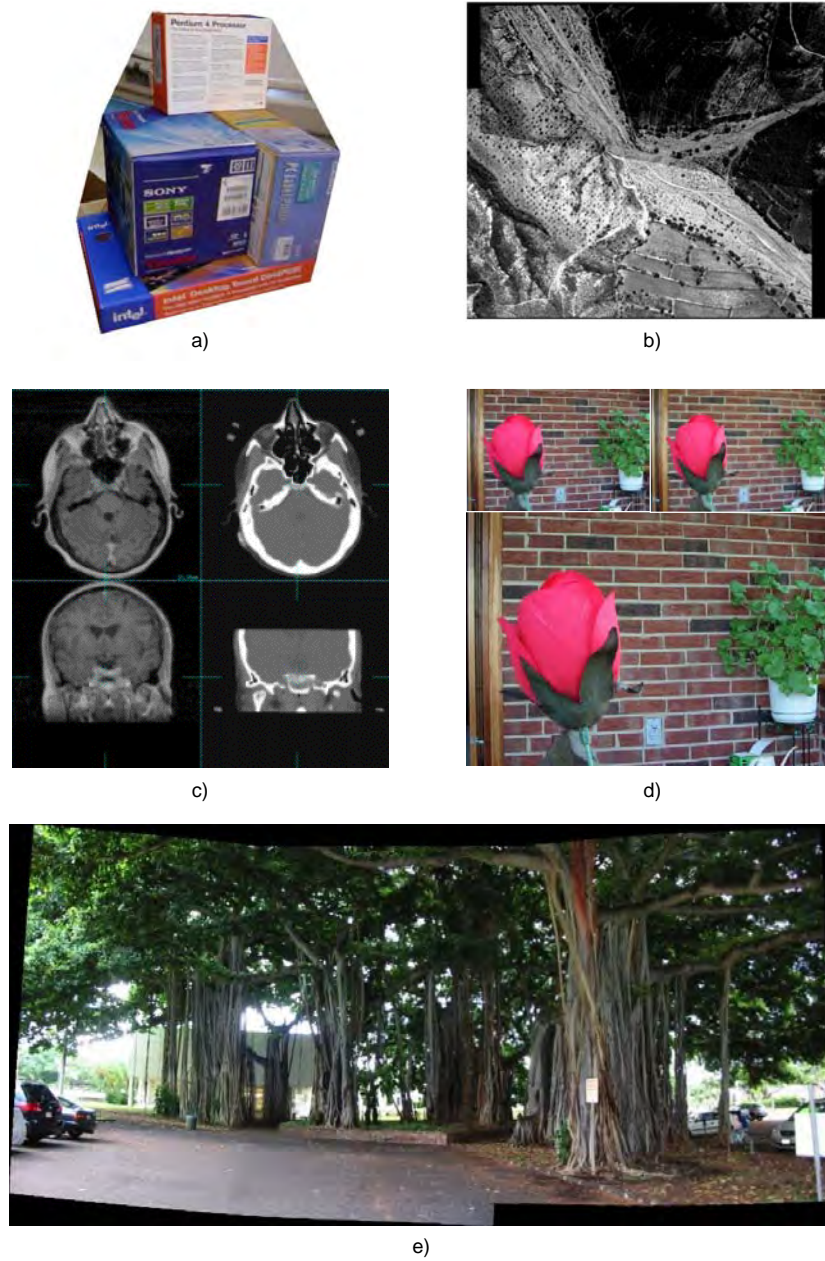


Figura 1.2: Más aplicaciones del registro de imágenes: (a) Reconstrucción 3D de escenas u objetos; (b) Fusión de imágenes de diferentes sensores, típica en teledetección y astrofísica; (c) Generación de atlas médicos para estudios médicos; (d) Compensación del desenfoque; (e) Generación de imágenes panorámicas.

ángulo de observación y los cambios en la iluminación), por la geometría de la escena, e incluso, por la propia naturaleza de las imágenes utilizadas, a veces, capturadas con sensores diferentes.

Entre estos factores destaca, por la relevancia de las diferencias geométricas relativas que inducen, la observación de una escena con geometría no-plana desde posiciones muy diferentes. Por ejemplo, imágenes de satélite de una región montañosa captadas en el *nádir*² pueden ser registradas con transformaciones relativamente simples (por ejemplo, homografías o funciones polinomiales [37]), mientras que si, por el contrario, son captadas desde ángulos diferentes, las diferencias geométricas entre las imágenes se acentúan, requiriendo transformaciones más complejas, de influencia local, como las funciones de base radial o las lineales por trozos [2].

La figura 1.3 muestra varios ejemplos de distorsiones inducidas por la observación de una escena de geometría no-plana desde diferentes puntos de observación. Obsérvense las importantes diferencias geométricas locales que se producen entorno al puente (ver fig. 1.3(a)). Estos efectos pueden agravarse cuando aparecen oclusiones, cambios en el contenido debidos a la distinta iluminación, sombras, etc. (ver fig. 1.3(b)).

En la literatura se pueden encontrar una amplia variedad de métricas que cuantifican la precisión del registro, sin embargo, su comportamiento se ve afectado, entre otros, por la diferente naturaleza de las imágenes, las condiciones de iluminación, los cambios en el contenido de las imágenes, sombras, saturaciones, etc. En radiología, por ejemplo, cuando se registra una imagen de resonancia magnética (MR) y una tomografía computarizada (TC) de un mismo paciente, las medidas de similitud tradicionales como la correlación cruzada normalizada (NCC) o la suma de diferencias al cuadrado (SSD) no son efectivas, debiéndose utilizar medidas basadas en conceptos tales como entropía relativa, información mutua (MI), etc. propias de Teoría de la Información, y que ofrecen un comportamiento más eficaz.

La figura 1.4 muestra diversos ejemplos de imágenes multi-modales. Ob-

²Se denomina *nadir* a la intersección entre la vertical del observador y la esfera celeste. Si se considera la recta imaginaria que pasa por el centro de la Tierra y por nuestra ubicación en su superficie, el *nadir* se encuentra sobre esa recta.



Figura 1.3: Imágenes captadas desde distintos puntos de observación: (a) Observación de la cobertura terrestre fuera del nadir. (b) Imágenes de una escena de interior típica.

sérvese la diferente naturaleza de cada una de las imágenes médicas³, a pesar de que todas ellas corresponden al mismo paciente (ver fig. 1.4(a)). Algo similar ocurre cuando la cámara incluye sensores de diferentes tipos, por ejemplo, color e infrarrojos (ver fig. 1.4(b)).

A continuación se introduce formalmente el problema del registro y las etapas y elementos que intervienen en él.

1.3 Definición del problema

Dado un punto cualquiera $X = (x, y, z)^T \in \mathbb{R}^3$ que se proyecta en los planos imagen de dos cámaras en las coordenadas $x = (x, y)^T \in \Omega$ y $x' = (x', y')^T \in$

³Las imágenes médicas, a menudo, son matrices tridimensionales de medidas de resonancia magnética, tomografía computarizada, etc. generadas a partir de “cortes” o secciones transversales de algún órgano del paciente.

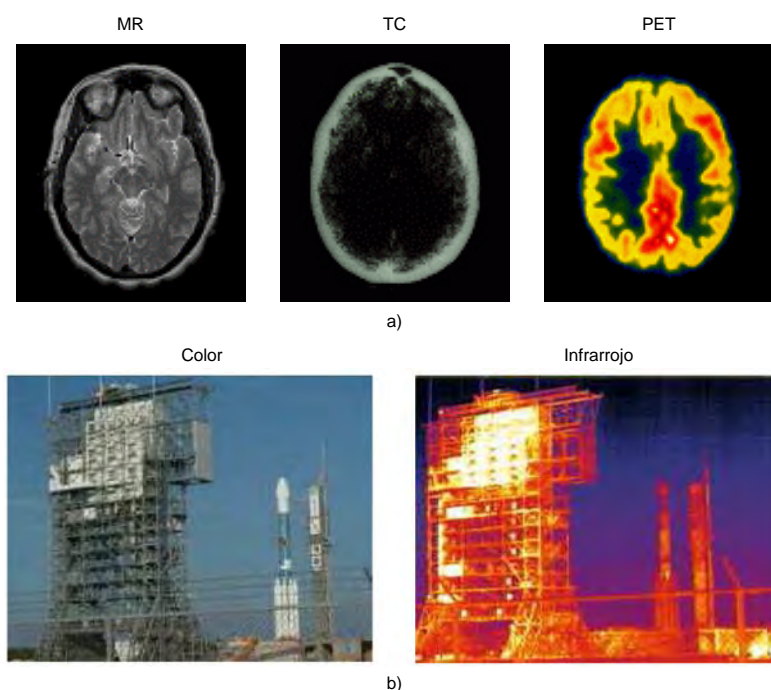


Figura 1.4: Imágenes de distinta naturaleza (multi-modalidad): (a) MR, TC y tomografía por emisión de positrones (PET) de un mismo paciente y (b) Imagen en color (RGB) e infrarrojo (IR) del Spitzer Space Telescope's Delta II.

Ω , respectivamente, con $\Omega \subset \mathbb{N}^2$, se define la aplicación $f : \Omega \mapsto \Omega$, biyectiva, que relaciona geoméricamente las coordenadas x y x' como:

$$x = f(x') \equiv \begin{cases} x = f_x(x', y') \\ y = f_y(x', y') \end{cases} \quad (1.1)$$

donde f_x e f_y son las funciones de transformación en x e y , respectivamente.

Consecuentemente, el *registro de imágenes* consiste en determinar el tipo y parámetros de la función f que superpone x' sobre x . La figura 1.5 muestra un esquema simplificado de este proceso.

Sean I e I' la imagen de referencia y la de entrada, respectivamente. Considérese también una función r de interpolación y una medida s de consistencia del registro. El proceso de registro se aborda típicamente del siguiente modo:

1. se extraen elementos distintivos de las imágenes y se establecen correspondencias entre ellos;

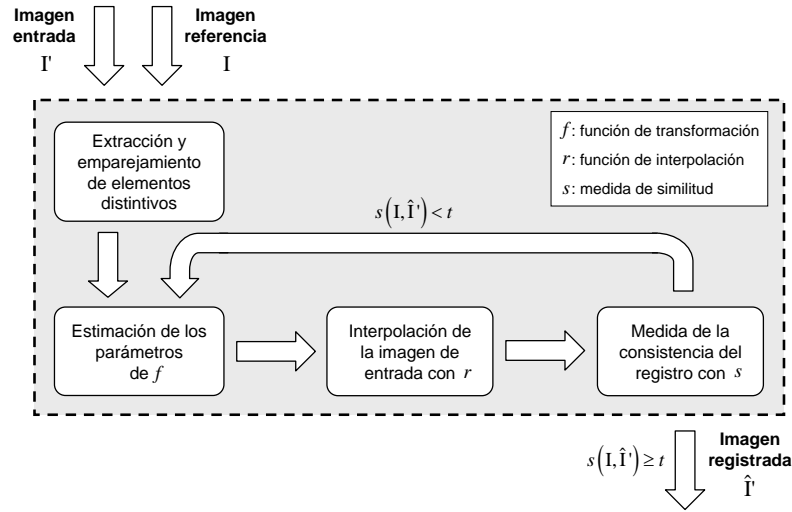


Figura 1.5: Esquema de un proceso genérico de registro de imágenes.

2. utilizando las correspondencias establecidas, se determinan los parámetros de la función de transformación f ;
3. para generar la imagen registrada, se transforman las coordenadas de la imagen de entrada y se transfieren los valores de intensidad mediante la función de interpolación r ;

Este proceso se repite iterativamente hasta que la medida de similitud s alcanza un valor dado t .

A continuación se describen detalladamente los elementos que intervienen en este proceso.

1.3.1 Las bases del registro

Las *bases del registro* son aquellos elementos utilizados para establecer la relación geométrica entre las imágenes. Estos elementos pueden ser pares de características de ambas imágenes o las imágenes completas, esto es, todos los niveles de intensidad de ambas imágenes. Dependiendo de las bases utilizadas, los métodos de estimación se pueden agrupar en dos categorías:

Basados en puntos: Los parámetros de la función de transformación se determinan a partir de un conjunto de correspondencias $\{(x_i, x'_i), i = 1, \dots, n\}$,

llamados *puntos de control* (CP), identificadas en ambas imágenes. En este proceso suelen utilizarse métodos robustos de estimación como el LMedS o RANSAC (métodos que se analizan detalladamente en el apéndice B). Formalmente, el problema consiste en resolver:

$$\arg \min_{\alpha, \{(x_j, x'_j), j=1, \dots, m\}} (x_i - f(x'_i)) \quad (1.2)$$

donde α son los parámetros de f y $\{(x_j, x'_j), j = 1 \dots m\}$, con $m \leq n$, un conjunto de correspondencias libre de espurios.

Basados en intensidad: Los parámetros de la función de transformación se obtienen como resultado de un proceso de optimización consistente en maximizar cierta medida de similitud de intensidad. Este proceso, a diferencia del anterior, involucra la totalidad de los píxeles de ambas imágenes lo que requiere normalmente un mayor esfuerzo computacional. Formalmente, hay que resolver:

$$\arg \max_{\alpha} s(I(x_i), I'(f^{-1}(x_i))) \quad (1.3)$$

donde $i = 1, \dots, N$, siendo N el número de píxeles de la imagen de referencia y s una medida de similitud.

1.3.2 La función de transformación

La *función de transformación* determina la relación geométrica entre los píxeles de ambas imágenes (ver fig. 1.6). En general, se puede hablar de dos grupos:

Rígidas: Transformaciones que aseguran la invarianza de las rectas. Las transformaciones Euclídea, afín o proyectiva son algunos ejemplos de esta categoría.

No-rígidas: Transformaciones que no aseguran la invarianza de las rectas. En esta categoría se pueden citar, entre otras, las funciones polinomiales, de base radial o las lineales por trozos.

Diversos autores utilizan el término elástico para referirse a las transformaciones no-rígidas [85, 33]. En esta tesis se utilizarán indistintamente ambos términos.

Las funciones rígidas o elásticas, a su vez, se pueden clasificar atendiendo a su alcance o ámbito de influencia en:

Globales: La transformación afecta a la totalidad de los píxeles de la imagen.

Locales: La influencia de la transformación depende de la posición de cada píxel en la imagen.

En la literatura también se pueden encontrar enfoques *híbridos*, como el propuesto en [41], donde se propone combinar una afinidad global y un conjunto de n funciones locales de base radial.

		Imagen			
		Euclídea	Afín	Proyectiva	Elástica
Global	Global				
	Local				

Figura 1.6: Clasificación de las funciones de transformación.

En el capítulo 3 se hace una revisión de las funciones más representativas, así como un análisis comparativo de diversas transformaciones no-rígidas (sección 3.3). Los parámetros de las funciones se estiman mediante puntos de control (CP) identificados en las imágenes. El procedimiento utilizado para su identificación robusta se describe en la sección 4.5.1. Por completitud, en

el apéndice D se describen las técnicas de detección de esquinas utilizadas en este proceso.

Esta tesis se centra, en especial, en la función lineal por trozos, una transformación local que aborda el registro de las imágenes mediante su división en regiones triangulares conjugadas que se registran individualmente mediante transformaciones afines. En los capítulos 4 y 5 se formaliza el registro lineal por trozos y se propone un procedimiento basado en la optimización de redes triangulares para mejorar su precisión. Para formalizar el problema del registro lineal por trozos se proponen unas estructuras matemáticas denominadas *complejos simpliciales*. Estas estructuras, desarrolladas en el campo de la topología algebraica, se describen formalmente en el apéndice C.

1.3.3 La medida de consistencia

La *medida de consistencia del registro* cuantifica cómo de bien se superponen la imagen registrada y la imagen de referencia. Para ello se puede utilizar una amplia variedad de métricas, las cuales también se pueden agrupar atendiendo a las bases utilizadas:

Basadas en puntos: La precisión del registro se determina a partir de los errores geométricos de ajuste (distancia) de un conjunto de correspondencias identificados en las dos imágenes, denominados *puntos independientes de control* (ICP). Una transformación inadecuada conlleva importantes desajustes, y viceversa. Para medirlos se emplea típicamente el error cuadrático medio (RMSE) y/o el error circular con 90 % de confianza (CE90). Ambos se comentan con más detalle en la sección 2.1.

Basadas en intensidad: Este tipo de medidas, a diferencia de las anteriores, cuantifican la precisión del registro comparando el contenido de las dos imágenes. Existen tantas posibilidades para ello como enfoques para evaluar la similitud de dos series numéricas de datos como, por ejemplo, el coeficiente de correlación de Pearson, la información mutua (MI), el análisis de componentes principales (PCA), etc. Nótese que, en

este enfoque, la normalización radiométrica de las imágenes es crucial, ya que cambios en el nivel de intensidad debidos, por ejemplo, a una distinta iluminación, serían considerados como errores de ajuste.

La elección de la medida de consistencia idónea, crucial en cualquier procedimiento de registro, depende de una variedad de factores como la disponibilidad de ICPs fiables, la naturaleza de las imágenes (esto es, mono-modales o multi-modales), su sensibilidad a las posibles diferencias radiométricas e incluso su coste computacional. En el capítulo 2 se realiza una revisión de las medidas de similitud más representativas.

El procedimiento de registro propuesto en el capítulo 5 de esta tesis emplea el coeficiente de correlación de entropía (ECC), una variante normalizada de la MI, para dirigir eficientemente el proceso de optimización. El ECC, al igual que la MI, se estima a partir de las funciones de distribución de probabilidad marginal y conjunta de las intensidades de ambas imágenes. En la sección 2.4 se analiza experimentalmente diversos procedimientos utilizados para la correcta estimación de estas distribuciones. Muchos de los conceptos y términos que se emplean en esta sección son descritos con más detalle en el apéndice A.

1.3.4 La función de interpolación

En la mayoría de las ocasiones, el resultado de transformar geoméricamente las coordenadas de un píxel no es un par de coordenadas discretas, es decir, las coordenadas transformadas no coinciden con un píxel. Se requiere, por tanto, una *función de interpolación*. El proceso de interpolación se aborda, típicamente, del siguiente modo (ver fig. 1.7):

1. para cada píxel $x = (x, y)^\top$ de la imagen interpolada se obtienen las coordenadas origen en la imagen de entrada mediante $f^{-1}(x)$,
2. dependiendo de la función de interpolación, se determina la intensidad/color a transferir a la imagen interpolada,
3. finalmente, se “rellena” el píxel x con dicho valor.

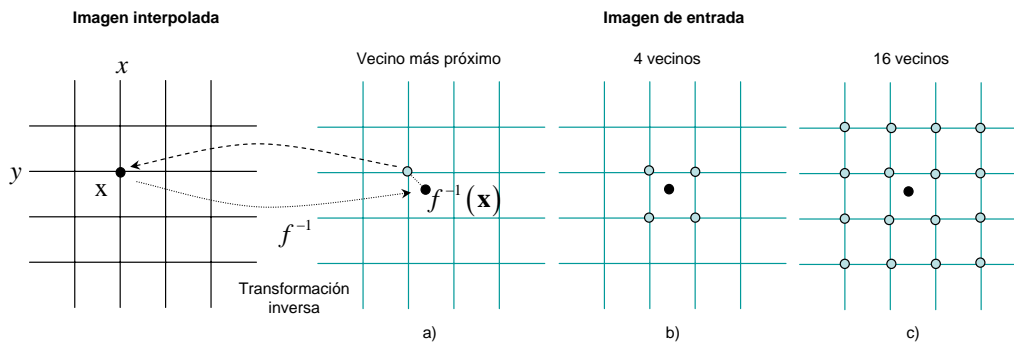


Figura 1.7: Funciones de interpolación: a) “vecino más próximo”, b) bilinear y c) bicúbica.

En la literatura se proponen diferentes funciones de interpolación, algunas poco costosas computacionalmente, como “el vecino más próximo”, pero que produce imágenes con contornos fragmentados (efecto escalón). Otras más costosas, pero que producen imágenes de una mayor calidad visual, como la interpolación bilinear, bicúbica (que utilizan, respectivamente, los 4 o 16 píxeles más cercanos para estimar el valor del píxel transformado), *splines*, etc. Algunos paquetes de procesamiento de imágenes incorporan técnicas mucho más elaboradas, como la propuesta por Intel en [53], denominada *super-sampling*, especialmente apropiada para registrar imágenes que presentan cambios de escala muy importantes.

1.3.5 El método de estimación

El proceso de registro consiste en determinar los parámetros óptimos de una función de transformación geométrica, esto es, resolver los problemas de minimización o maximización propuestos en (1.2) o (1.3), respectivamente. En este sentido se han propuesto una amplia variedad de procedimientos de optimización y estimación robusta. Así, se pueden encontrar trabajos en los que dada la simplicidad de la función objetivo, por ejemplo, Euclídea o polinomial de orden bajo, se resuelven con formulaciones cerradas [65]. Si la dimensión del problema crece, como ocurre normalmente cuando se emplean funciones de transformación no-rígidas y/o locales, ya no es posible emplear soluciones cerradas y se recurre a técnicas de optimización basadas en algo-

ritmos iterativos (como el descenso del gradiente o la búsqueda *greedy* [77]), modelos bayesianos [105], el recocido simulado [91], etc. El lector puede dirigirse al trabajo de Maes [68] para un estudio completo sobre diferentes estrategias de optimización.

1.4 Contribuciones de la Tesis

Esta tesis aborda el registro de imágenes que presentan diferencias geométricas locales. En este caso sólo las funciones de transformación elásticas y/o locales ofrecen soluciones aceptables en términos de precisión. Cómo medir esta precisión es justamente uno de los aspectos claves en el procedimiento de registro, ya que esta medida es la que guía el proceso de estimación de los parámetros de la función de transformación.

Aunque algunas de las contribuciones de este trabajo ya se han apuntado con anterioridad, a continuación se detalla cada una de ellas así como las publicaciones que se han derivado de este trabajo.

- Análisis experimental sobre la estimación robusta de funciones distribución de probabilidad de intensidades de la imagen y su aplicación en el cálculo del ECC, una variante normalizada de la MI. Este análisis ha sido publicado parcialmente en [6, 11].
- Evaluación de las técnicas de registro no-rígidas más representativas. Este trabajo ha sido publicado en [2, 4]. En [1, 7, 37] se pueden encontrar aplicaciones de este tipo de transformaciones en el campo de la teledetección.
- Formalización del registro lineal por trozos mediante complejos simpliciales. Esta formalización ha sido parcialmente publicada en [3].
- Optimización topológica y geométrica de redes triangulares para mejorar la precisión del registro lineal por trozos. La optimización topológica ha sido publicada en [5, 6, 11] (el primero en proceso de revisión).

En la implementación del método de registro lineal por trozos propuesto, así como en las diferentes pruebas experimentales, se ha utilizado una

librería *open-source* de visión por computador denominada OpenCV [100]. Se han realizado diversos trabajos en los que se describen las características de esta librería y sus posibles aplicaciones en el campo de la investigación y la docencia [8, 9, 10].

Por otro lado, para poder extraer robustamente pares de correspondencias en imágenes de satélite de alta-resolución en las que proliferan las sombras, se ha desarrollado una novedosa técnica que permite detectar y delimitar con precisión las zonas de la imagen afectadas [3, 12].

En los trabajos [38, 64, 76], aún no estando directamente relacionados con esta tesis, se aplican conceptos e ideas adquiridos a lo largo de ella para la detección de olivos en imágenes de satélite, la estimación del movimiento mediante imágenes adquiridas con un sistema estéreo y la detección de matrículas, respectivamente.

Capítulo 2

Medidas de consistencia del registro

La estimación de los parámetros de la función de transformación se realiza típicamente mediante un proceso iterativo dirigido por una medida de consistencia del registro. Estas medidas tratan de cuantificar la precisión con la que la imagen de entrada, una vez registrada, se ajusta a la de referencia. En la literatura se han propuesto diversas métricas para este fin, las cuales miden cómo de bien se ajustan ambas imágenes en base a:

1. las diferencias radiométricas de la totalidad sus píxeles (medidas basada en intensidad) o
2. los errores de ajuste de un conjunto de correspondencias identificadas en ellas (medidas basadas en puntos).

La elección de la medida de similitud idónea es, por tanto, crucial en cualquier procedimiento de registro y depende de diversos factores: la disponibilidad de correspondencias fiables, la diferente naturaleza de las imágenes (esto es, monomodales o multimodales), las diferencias radiométricas, e incluso, su coste computacional. En este capítulo se revisan y comparan las más representativas, haciendo especial hincapié en la información mutua (MI), una medida de la entropía relativa de dos variables aleatorias y que fue propuesta originariamente para el registro de imágenes médicas [67, 104]. La

MI se calcula a partir de las funciones de distribución de probabilidad marginales y conjunta de los niveles de intensidad de las imágenes. La medida MI de dos imágenes depende, por tanto, de la correcta estimación de estas distribuciones.

Esta tesis contribuye con una evaluación experimental en la que se comparan diferentes procedimientos para la estimación eficaz de las distribuciones de probabilidad. También se analiza la influencia en esta estimación de parámetros tales como el tamaño de la muestra, el número de intervalos (o *bins*) del histograma conjunto, la función de interpolación utilizada, etc. En el apéndice A se describen conceptos utilizados en este capítulo tales como la entropía, entropía relativa y la MI.

2.1 Medidas basadas en puntos

Las medidas basadas en puntos cuantifican la precisión del registro en base al error en el ajuste de un conjunto de correspondencias, denominados puntos independientes de control (ICP), seleccionadas en ambas imágenes. Estas medidas se caracterizan por su reducido coste computacional y su robustez a las diferencias radiométricas o diferente modalidad de las imágenes. Sin embargo, su efectividad depende críticamente del número de ICPs utilizados, así como de la precisión con la que han sido seleccionados y de lo representativos que sean de las diferencias geométricas de ambas imágenes (esto es, su distribución).

Un claro ejemplo de utilización de estas medidas tiene lugar en el campo de la teledetección, y se debe fundamentalmente a lo siguiente: el tamaño de las imágenes, de varios cientos de megabytes en ocasiones, desaconsejan el empleo de cualquier técnica basada en intensidad, mucho más costosas computacionalmente.

A continuación se definen las medidas basadas en puntos más utilizadas.

Dadas dos imágenes I e I' del mismo tamaño; un conjunto de n correspondencias $\{(x_i, x'_i), i = 1, \dots, n\}$ identificadas en ellas y la función de transformación geométrica f , se definen la siguientes medidas de consistencia:

- Error cuadrático medio:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n d_i^2} \quad (2.1)$$

- Error circular con 90 % de confianza:

$$P(d_i \leq CE90) = 90\% \quad (2.2)$$

donde

$$d_i = \|x_i - f(x'_i)\|.$$

La principal diferencia entre ambas radica en lo siguiente: mientras el RMSE considera los errores de todos los ICPs, incluyendo los correspondientes a las posibles correspondencias espurias, el CE90 proporciona el valor que acota superiormente el 90 % de los errores, esto es, no proporciona información sobre cómo de mal se ajusta es el 10 % restante. La figura 2.1 ilustra gráficamente el significado de ambas medidas.

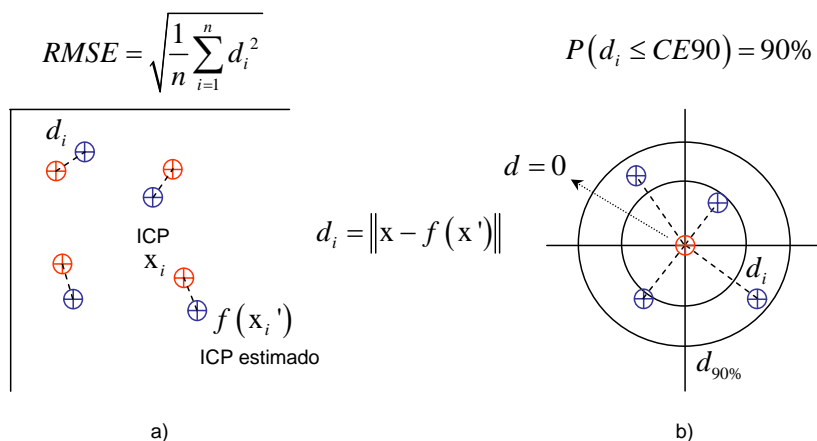


Figura 2.1: Medidas de similitud basadas en puntos: (a) RMSE mide el error de ajuste medio de todos los ICPs, incluidos los correspondientes a los espurios. (b) CE90 mide el error máximo que acota superiormente el 90 % de los errores de ajuste.

2.2 Medidas basadas en intensidad

Las medidas basadas en intensidad cuantifican la precisión del registro mediante la comparación píxel a píxel del contenido de ambas imágenes, requiriendo, por tanto, un mayor esfuerzo computacional para su evaluación. A esto hay que sumarle otro inconveniente: la dificultad para discernir cuando la diferencia de intensidad que presenta un píxel en ambas imágenes se debe a un mal ajuste o a un cambio radiométrico. Esta circunstancia ocurre típicamente cuando se registran imágenes de diferente modalidad o de igual modalidad, pero que presentan cambios radiométricos importantes. Su ventaja fundamental es que no requieren seleccionar correspondencias.

En la literatura se pueden encontrar una variedad de medidas: algunas simples, basadas en la diferencia de intensidades [34, 43, 44] como la suma de diferencias al cuadrado (SSD) o la suma de diferencias en valor absoluto (SAD); otras con un claro enfoque estadístico como es el caso del coeficiente de correlación (CC) [62], la correlación cruzada normalizada (NCC) [48] o el ratio de uniformidad de la imagen¹ (RIU) [106, 107].

Estas medidas han demostrado su eficacia con imágenes monomodales no afectadas de importantes diferencias radiométricas, como es el caso de cambios en la iluminación o del contenido. Cuando el perfil radiométrico de las imágenes varía significativamente o son captadas con diferentes sensores, algo típico en medicina y teledetección, su eficacia se ve seriamente comprometida. Con objeto de minimizar esta limitación, diversos autores han propuesto técnicas más robustas, como la uniformidad de intensidad particionada² (PIU) [108] propuesta para el registro de imágenes MR-PET, o medidas que miden la “información compartida” como la entropía conjunta [98], la información mutua (MI) [67, 104], la información mutua normalizada (NMI) [97] o el coeficiente de correlación de entropía (ECC) [13].

A continuación se definen las medidas basadas en intensidad más representativas.

Sean I e I' dos imágenes de tamaño $M \times N$ y con valores de intensidad

¹Traducción de los términos ingleses *Ratio Image Uniformity*.

²Traducción de los términos ingleses *Partitioned Intensity Uniformity*.

$i \in [0, L-1]$ y $j \in [0, L-1]$, respectivamente. Sea la función de transformación geométrica f y la imagen transformada $\hat{I}' = I'(f^{-1})$. Se definen las siguientes métricas para medir la similitud, esto es, la consistencia del registro, entre las imágenes I e \hat{I}' :

- Distancia Euclídea:

$$DE(I, \hat{I}') = \sqrt{\sum_{y=1}^M \sum_{x=1}^N (I(x, y) - \hat{I}'(x, y))^2} \quad (2.3)$$

- Suma de diferencias al cuadrado (el cuadrado de la DE):

$$SSD(I, \hat{I}') = \sum_{y=1}^M \sum_{x=1}^N (I(x, y) - \hat{I}'(x, y))^2 \quad (2.4)$$

La SSD, al igual que la DE, se estima a partir de las diferencias en los niveles de intensidad de los píxeles de ambas imágenes. Así, cualquier diferencia de intensidades, por ejemplo, por cambios en el brillo o el contraste afectará a la medida, aún estando correctamente registradas. Este inconveniente limita notablemente su uso, restringiéndolo a imágenes monomodales sin diferencias radiométricas de relevancia.

- Coeficiente de correlación:

$$CC(I, \hat{I}') = \sum_{y=1}^M \sum_{x=1}^N (I(x, y) - \bar{I}) (\hat{I}'(x, y) - \bar{\hat{I}'}) \quad (2.5)$$

donde

$$\bar{I} = \frac{1}{M \times N} \sum_{y=1}^M \sum_{x=1}^N I(x, y)$$

$$\bar{\hat{I}'} = \frac{1}{M \times N} \sum_{y=1}^M \sum_{x=1}^N \hat{I}'(x, y)$$

El CC se deriva a partir de la SSD, pero a diferencia de ésta, es capaz de

compensar cambios en el brillo medio de las imágenes, restando al nivel de cada píxel el nivel de intensidad medio de la imagen. Pese a esta mejora, el CC continúa mostrando problemas cuando las diferencias entre imágenes no se restringen a este tipo de cambios.

- Correlación cruzada normalizada (coeficiente de correlación lineal de Pearson):

$$NCC(I, \hat{I}') = \frac{\sum_{y=1}^M \sum_{x=1}^N (I(x, y) - \bar{I}) (\hat{I}'(x, y) - \bar{\hat{I}'})}{\sqrt{\sum_{y=1}^M \sum_{x=1}^N (I(x, y) - \bar{I})^2} \sqrt{\sum_{y=1}^M \sum_{x=1}^N (\hat{I}'(x, y) - \bar{\hat{I}'})^2}} \quad (2.6)$$

La NCC es la variante normalizada del CC y toma valores en el intervalo cerrado $[-1, +1]$, donde un valor $+1$ indica que ambas imágenes encajan (se registran) perfectamente. Esta medida compensa, además de los cambios en el brillo medio, posibles variaciones en el contraste, para lo cual normaliza el CC con las desviaciones típicas de los niveles de ambas imágenes. Estas compensaciones mejoran sustancialmente su eficacia respecto a las medidas definidas arriba, aunque sólo ofrece un comportamiento adecuado cuando las diferencias en las paletas de las imágenes son lineales.

- Información mutua:

$$MI(I, \hat{I}') = \sum_i \sum_j P_{I, \hat{I}'}(i, j) \log \left(\frac{P_{I, \hat{I}'}(i, j)}{P_I(i) P_{\hat{I}'}(j)} \right) \quad (2.7)$$

donde $P_I(i)$, $P_{\hat{I}'}(j)$ y $P_{I, \hat{I}'}(i, j)$ son las estimaciones de las distribuciones de probabilidad a partir del histograma conjunto h de I e \hat{I}'

$$\begin{aligned} P_{I, \hat{I}'}(i, j) &= \frac{1}{M \times N} h_{I, \hat{I}'}(i, j) \\ P_I(i) &= \sum_j P_{I, \hat{I}'}(i, j) \\ P_{\hat{I}'}(j) &= \sum_i P_{I, \hat{I}'}(i, j) \end{aligned} \quad (2.8)$$

donde

$$h_{\mathbf{I}, \hat{\mathbf{I}}'} = \begin{pmatrix} h(0,0) & h(0,1) & \dots & h(0,L-1) \\ h(1,0) & h(1,1) & \dots & h(1,L-1) \\ \dots & \dots & \dots & \dots \\ h(L-1,0) & h(L-1,1) & \dots & h(L-1,L-1) \end{pmatrix} \quad (2.9)$$

siendo $h_{\mathbf{I}, \hat{\mathbf{I}}'}(i, j)$ el acumulado del par de niveles (i, j) formado tomando el valor i de la imagen \mathbf{I} y j de la imagen $\hat{\mathbf{I}}'$. El cálculo del histograma conjunto como se detalla aquí recibe el nombre de método paramétrico [104].

La MI mide la dependencia estadística de los niveles de intensidad de las dos imágenes, no su similitud radiométrica, luego no se ve afectada por la diferente naturaleza de las imágenes o las diferencias radiométricas que éstas pudieran presentar. Esta medida se analiza detalladamente en la sección 2.3.1.

Obsérvese en la expresión (2.7) que, cuando los valores de intensidad de las dos imágenes son independientes (esto es, sin correlación), el argumento de los logaritmos tiende a uno, alcanzando su valor mínimo en cero. Para ilustrar cómo el histograma conjunto captura el concepto de dependencia estadística, la figura 2.2 muestra los histogramas de dos pares de imágenes sintéticas. Nótese que, cuando la imagen de entrada está rotada, la dispersión del histograma es mayor (menor correlación) que cuando ambas están perfectamente alineadas.

La elección de la medida idónea depende del tipo de aplicación, de los requisitos computacionales y, fundamentalmente, de la modalidad y del tipo de diferencias radiométricas que exhiban las imágenes. Entre todas las definidas, las más utilizadas son la NCC y la MI, destacando la capacidad de la segunda para trabajar con imágenes multimodales y diferencias radiométricas no lineales. Desde la perspectiva de la Teoría de Información, la MI explota el concepto de entropía, el cual se describe a continuación junto con otros aspectos relacionados con la estimación de la MI.

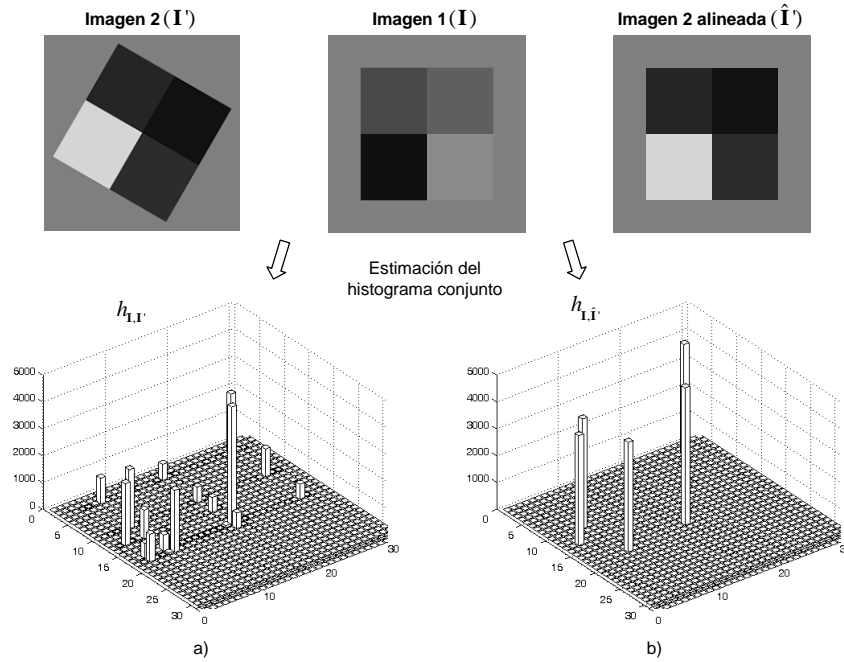


Figura 2.2: Histograma conjunto de dos pares de imágenes sintéticas (cuadrados de 4 tonos): (a) un par no alineado y (b) un par perfectamente alineado. Obsérvese como el histograma conjunto (de 32 *bins*) presenta una menor dispersión (mayor correlación) cuando las imágenes están alineadas.

2.3 Medidas basadas en la entropía

La *entropía* de una variable aleatoria X , denotada como $H(X)$, es una medida de la incertidumbre de X . Por ejemplo, dada un señal a la que se le supone un comportamiento aleatorio, la entropía mide la cantidad de información contenida en dicha señal (ver def. A.1). Matemáticamente se define como:

$$H(X) = \sum_{x \in X} p(x) \log \frac{1}{p(x)} \quad (2.10)$$

donde $p(x)$ es la distribución de probabilidad de X .

La figura 2.3 muestra una imagen I en escala de grises y su correspondiente histograma de intensidades h_I . Considérese el histograma h_I normalizado como una estimación de la probabilidad de ocurrencia de un determinado

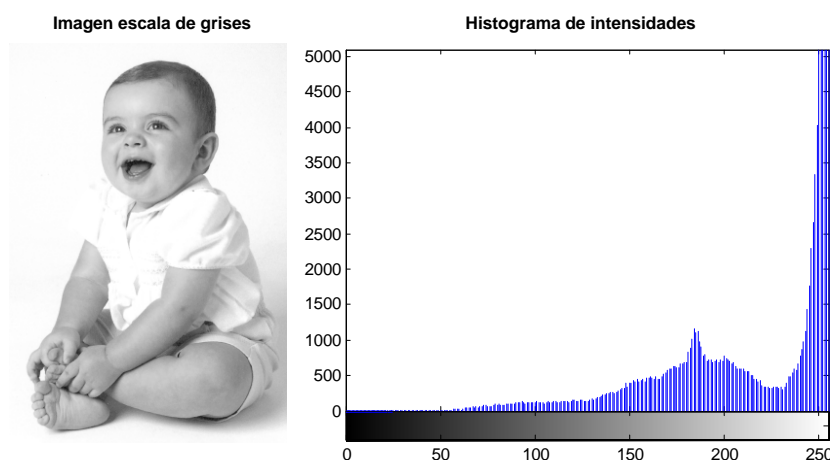


Figura 2.3: Histograma de una imagen en escala de grises.

nivel de intensidad, esto es,

$$p(i) = \frac{h_I(i)}{N} \quad (2.11)$$

donde $h_I(i)$ denota la entrada i -ésima del histograma de intensidades de I y N el número de píxeles de la imagen.

Se define la entropía de una imagen como:

$$H(I) = \sum_{i \in I} h_I(i) \log \frac{N}{h_I(i)} \quad (2.12)$$

La entropía de una imagen proporciona información sobre la ocurrencia de sus niveles de intensidad, de tal forma que la entropía es máxima si la distribución de niveles de gris es uniforme y mínima si la imagen tiene un sólo valor de intensidad.

Otra medida de incertidumbre, íntimamente relacionada con el concepto de entropía, es la *entropía relativa* (ver def. A.4). La entropía relativa o distancia de Kullback-Leibler mide la “distancia” entre dos distribuciones de probabilidad, esto es, dadas dos distribuciones cualesquiera de una variable aleatoria, la entropía relativa mide el error de asumir una distribución cuando la distribución verdadera es la otra. A pesar de que siempre es positiva

(con distribuciones diferentes) o cero (con distribuciones iguales), no es una distancia en el sentido más estricto de la palabra ya que no es simétrica ni verifica la desigualdad del triángulo, circunstancias que dificultan su empleo como medida de consistencia del registro [26].

2.3.1 Información mutua

La *información mutua* (MI) (ver def. A.5) mide la dependencia estadística o redundancia de información de dos variables aleatorias. Se deriva a partir de la entropía relativa pero, a diferencia de ésta, es simétrica y verifica la desigualdad del triángulo. En 1997, Viola [104] y Maes [67] proponen, por primera vez y casi simultáneamente, el empleo de la MI como medida de consistencia del registro.

Tradicionalmente, el parecido o similitud entre dos imágenes se ha medido con la SSD o su variante normalizada, la NCC que, cómo se ha comentado anteriormente, admite diferencias lineales entre las paletas de las dos imágenes [23]. Cuando esta relación no se verifica para toda la paleta, la NCC es incapaz de captar el parecido de éstas. La MI, sin embargo, no asume una relación funcional (normalmente lineal) entre los niveles de intensidad de las imágenes a comparar, sino su relación estadística. Es decir, no asume que la naturaleza radiométrica de las imágenes sea la misma o muy similar, por tanto, se puede utilizar con imágenes multimodales o monomodales con perfiles radiométricos muy dispares.

Para ilustrar la ventaja de la MI frente a la NCC se presentan dos experimentos. En el primero, se alinean (mediante una rotación) los imágenes sintéticas de la figura 2.2, midiendo la consistencia del registro con ambas métricas. Como se observa en la figura 2.4(a), la MI proporciona un máximo en cero grados (esto es, cuando ambas imágenes encajan perfectamente), a pesar de que las intensidades de las imágenes no coinciden; este tipo de correlación pasa completamente desapercibida para la NCC.

Lo mismo ocurre con el ejemplo que se muestra en la figura 2.4(b). En este caso, un trozo de una imagen de satélite QuickBird es desplazada sobre otra mayor pero adquirida seis meses después (esto es, diferentes condiciones de

iluminación y cambios en el contenido debidos al cambio estacional, cambios temporales en la escena, etc.). Estos resultados ponen de relieve la robustez y efectividad de la MI cuando se utiliza con pares de imágenes con diferencias radiométricas no-funcionales.

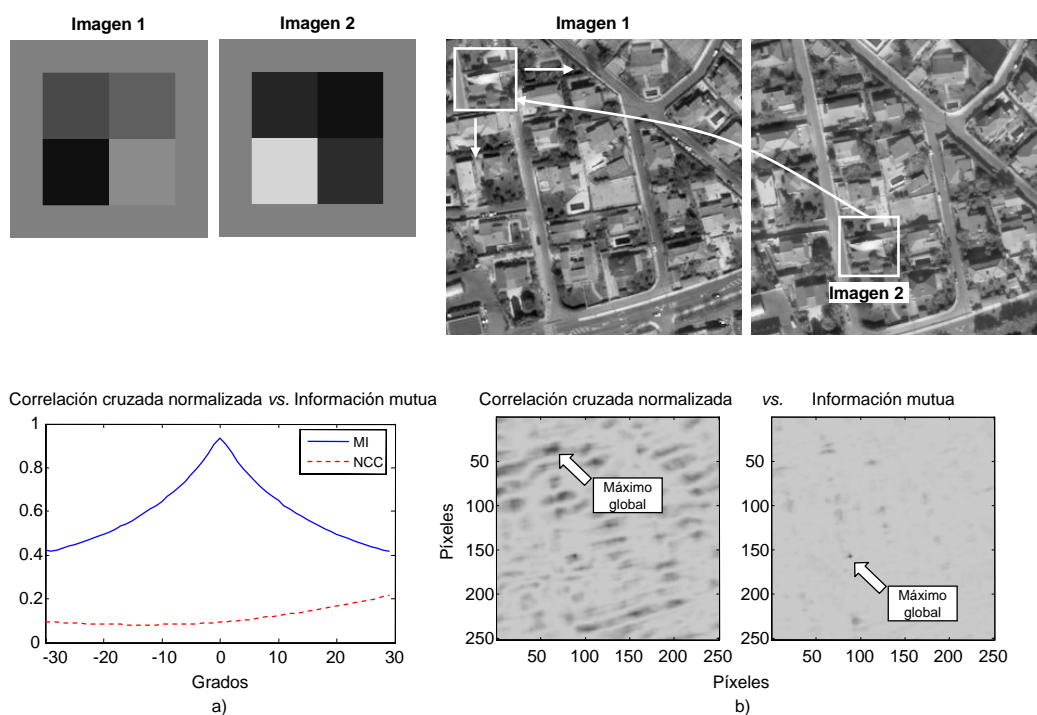


Figura 2.4: Experimentos que ilustran la idoneidad de la MI frente a la NCC como medida de consistencia. (a) Un patrón sintético es rotado (desde -30° hasta $+30^\circ$) sobre otro con la misma estructura (esto es, un cuadro de igual tamaño) pero con distintos niveles de gris; y (b) un pequeño trozo de una imagen de satélite QuickBird es desplazado sobre otra imagen de la misma escena adquirida en otra fecha. Obsérvese como, a diferencia de la MI, la NCC falla en ambos experimentos, esto es, no detecta un máximo global en (a) y proporciona un máximo global erróneo en (b).

La MI se relaciona con la entropía mediante las siguientes expresiones (ver sec. A.4 para un desarrollo completo):

$$MI(I, \hat{I}') = H(I) + H(\hat{I}') - H(I, \hat{I}') \quad (2.13)$$

donde $H(I)$ y $H(\hat{I}')$ son las entropías de I y \hat{I}' , respectivamente, y $H(I, \hat{I}')$

la entropía conjunta.

En la práctica son frecuentes los trabajos que utilizan derivaciones normalizadas de la MI, por ejemplo en Maes [67] y Chen [23]. A continuación se definen las dos más representativas:

- Información mutua normalizada [97]:

$$NMI(I, \hat{I}') = \frac{H(I) + H(\hat{I}')}{H(I, \hat{I}')} \quad (2.14)$$

- Coeficiente de correlación de entropía [13]:

$$ECC(I, \hat{I}') = 2 - \frac{2H(I, \hat{I}')}{H(I) + H(\hat{I}')} \quad (2.15)$$

que toman valores en los intervalos cerrados $[0, 2]$ y $[0, 1]$, respectivamente.

Estas variantes normalizadas aportan ciertas ventajas respecto a sus correspondiente no normalizada. La más importante es que permite comparar cómo de bien se ajustan dos regiones conjugadas de la imagen con respecto a otras dos cuando el tamaño de ambos pares es diferente. Esta circunstancia es clave en el procedimiento de registro propuesto en el capítulo 5.

El cálculo de la MI depende de forma crucial de la correcta estimación de las funciones de distribución marginales y conjunta de los niveles de intensidad de ambas imágenes, convirtiéndose éste en el principal escollo para la utilización de esta medida. A continuación se analiza esta problemática, así como diferentes procedimientos para la correcta estimación de las distribuciones. También se analiza la influencia del procedimiento de interpolación en la estimación de estas funciones.

2.4 Estimación de la distribución de probabilidad

Una estimación imprecisa de las distribuciones de probabilidad introduce errores en la MI que derivan en una medida errónea de la similitud (expresiones (2.7) y (2.13)). Existen diversos factores que influyen en la estimación de estas distribuciones como, por ejemplo, la elección del número de intervalos del histograma conjunto (esto es, el número de *bins*) y el tamaño de las imágenes (número de píxeles considerados). Estos elementos determinan cómo de representativo es el histograma conjunto obtenido a partir de los niveles de intensidad de las imágenes.

Así, por ejemplo, si se desea medir la similitud de una imagen de 32×32 píxeles, de 256 niveles de gris, superpuesta sobre otra de mayor tamaño, el histograma conjunto será una matriz de 256×256 donde se acumulan sólo 32×32 pares, lo que da lugar a un histograma poco representativo. En estos casos, si se reduce el número de *bins* del histograma conjunto (expresión 2.8) utilizando, por ejemplo, 64 ó 32 en lugar de 256 (en el caso de imágenes en escala de gris) se obtiene un histograma más representativo y, por tanto, mejores estimaciones de las distribuciones de probabilidad. Con esta actuación también se logra reducir el tiempo de computo (menos términos en el sumatorio de la expresión (2.7)) y dotar de robustez al proceso frente al ruido.

Reducir el número de intervalos del histograma en función del número de muestras disponibles, como se propone arriba, no siempre garantiza una estimación apropiada de las funciones de distribución. Existen métodos de estimación no-paramétricos, más costosos computacionalmente, que tratan de minimizar esta limitación infiriendo funciones de distribución a partir de un conjunto limitado de muestras. Este es el caso de la ventana de Parzen [104], que propone acumular un kernel gaussiano K , centrado en cada par (i, j) , en lugar de unidades discretas. Más concretamente, la ventana de Parzen consiste en una acumulación ponderada, función de σ , con aportación máxima en la celda (i, j) y que decrece a medida que se aleja de ella, de modo que, cuanto más pequeña sea la desviación más se asemejará a la dis-

tribución obtenida con el método paramétrico. La estimación vendría dada por la siguiente expresión:

$$p(i, j) \approx p^*(i, j) = \frac{1}{N} \sum_{(a,b) \in S} K(i - a, j - b) \quad (2.16)$$

donde S es el subconjunto de muestras utilizado.

Naturalmente, este procedimiento asume que el subconjunto de muestras utilizado en la estimación es representativo de la distribución real, en cuyo caso, se obtiene una mejor estimación de ésta. La figura 2.5(a) ilustra gráficamente este procedimiento mediante la estimación de una distribución Gaussiana a partir de un conjunto reducido de muestras.

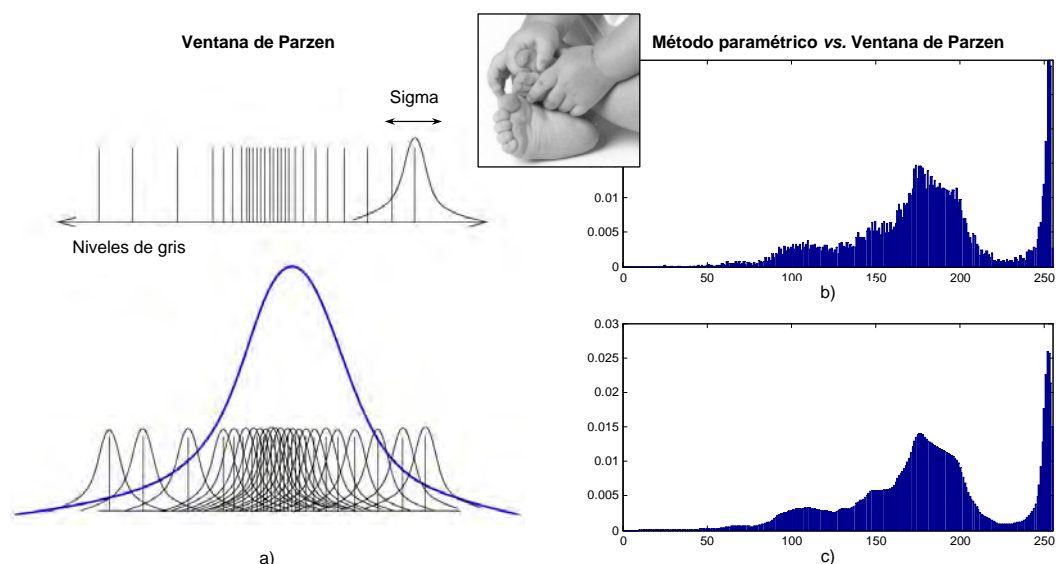


Figura 2.5: Estimación de una distribución de probabilidad mediante la ventana de Parzen. (a) Estimación de una Gaussiana a partir de un conjunto reducido de muestras (ejemplo extraído de [104]). (b) Histograma normalizado de niveles de gris (expresión (2.11)). (c) Estimación obtenida con la ventana de Parzen.

Las figuras 2.5(b,c) muestran las distribuciones de probabilidad de una imagen en escala de gris estimadas acumulando directamente los niveles de gris (histograma normalizado) y mediante la ventana de Parzen, respectivamente. Obsérvese cómo la estimación obtenida con la ventana de Parzen es mucho menos ruidosa que la obtenida a partir del histograma.

2.4.1 Evaluación experimental

Independientemente del método utilizado, paramétrico o la ventana de Parzen, la estimación de la función de distribución depende, entre otros, del número de *bins* y del tamaño de la imagen y , en el caso de la ventana de Parzen, de la desviación típica del kernel gaussiano. En esta sección se analiza la influencia de estos parámetros en las estimaciones proporcionadas por ambos procedimientos.

Para realizar este análisis se han efectuado diversas pruebas consistentes, básicamente, en desplazar horizontalmente y píxel a píxel ($\Delta_x = 1$), trozos de imágenes de satélite IRS-1D ($L = 256$) sobre otra imagen de la misma zona y de mayor tamaño (imagen de referencia), calculando el ECC en cada posición. Recuérdese que, de acuerdo con la expresión (2.15), para calcular el ECC se requieren las distribuciones de probabilidad conjunta y marginales de los niveles de intensidad de ambas imágenes (expresión (2.8)).

Los resultados de estas pruebas se recogen en las figuras 2.6 y 2.7. Del análisis de estas gráficas se extraen diversas conclusiones, algunas de ellas vienen a corroborar algunos aspectos comentados anteriormente:

1. El método paramétrico no produce distribuciones de probabilidad representativas cuando el tamaño de la imagen, en comparación con la dimensión del histograma, no es importante (fig. 2.6 abajo). Cuando el número de muestras utilizado en la estimación es grande (fig. 2.6 arriba), las curvas muestran un pico único y distintivo, independientemente del número de *bins* considerado. Obsérvese, también, como una reducción del número de intervalos del histograma (32 en lugar de 64, 128 ó 256) produce mejores resultados (fig. 2.6 abajo).
2. La ventana de Parzen produce mejores estimaciones de las distribuciones de probabilidad incluso cuando el número de muestras, en relación con el tamaño del histograma, es pequeño (fig. 2.7(a) abajo), mostrando un mejor comportamiento que el método de estimación tradicional en idénticas condiciones (fig. 2.7(a) arriba). Obsérvese en la figura 2.7(b) cómo, lógicamente, cuando σ decrece la respuesta de la ventana de Parzen se asemeja a la obtenida con el método paramétrico.

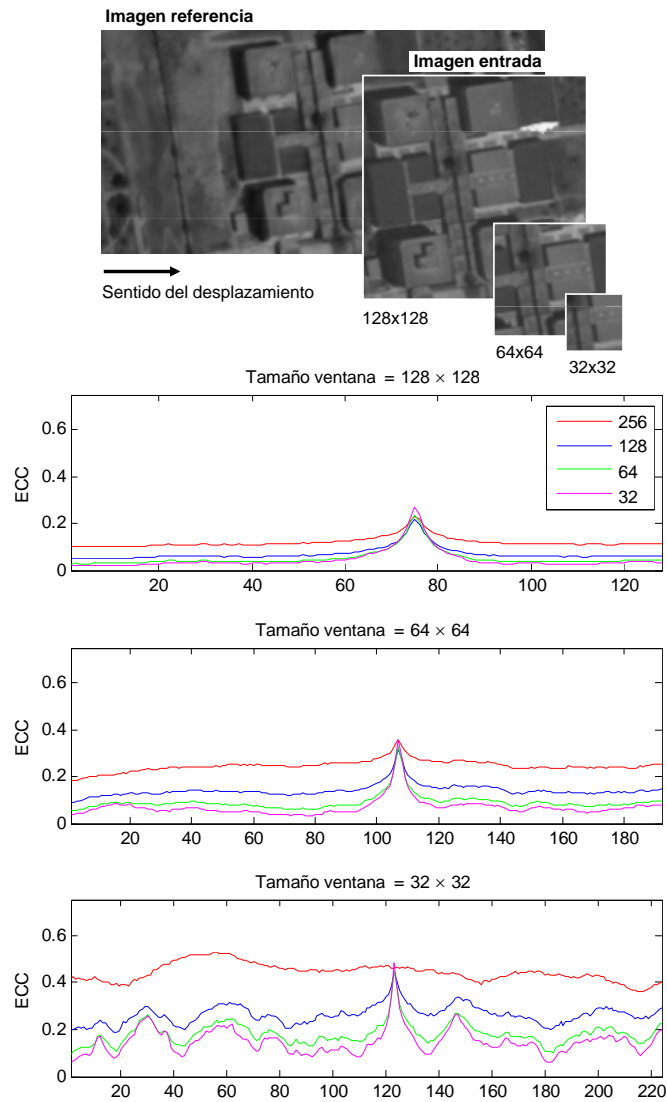


Figura 2.6: Valores de ECC obtenidos al desplazar imágenes cuadradas de 128, 64 y 32 píxeles de lado sobre la imagen de referencia. La estimación de las funciones de distribución se ha realizado con el método paramétrico modificando el número de *bins* y el tamaño de la ventana.

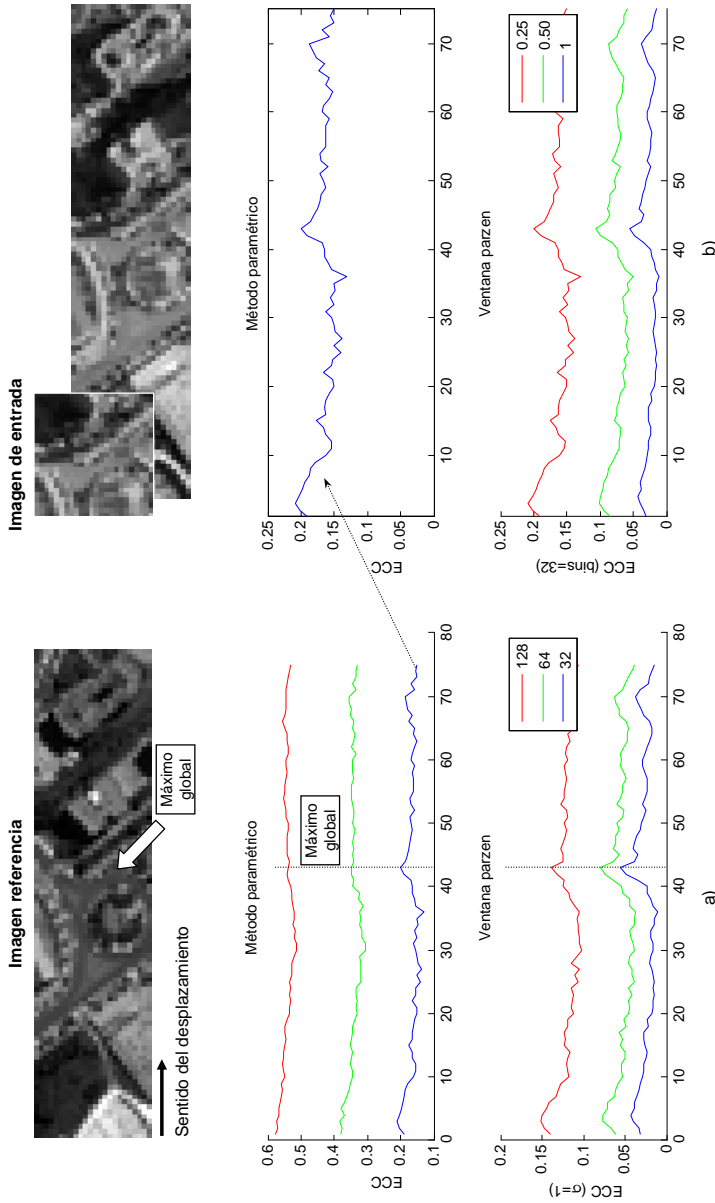


Figura 2.7: Valores de ECC obtenidos al desplazar un fragmento cuadrado de la imagen de entrada (de 32 píxeles de lado) sobre la imagen de referencia. (a) Estimación de las funciones de distribución realizada con el método paramétrico (arriba) y la ventana de Parzen (abajo) modificando el número de *bins* $L = 128, 64, 32$. (b) Estimación de la función de distribución con la ventana de Parzen utilizando diferentes desviaciones típicas $\sigma = 0,25, 0,5, 1$.

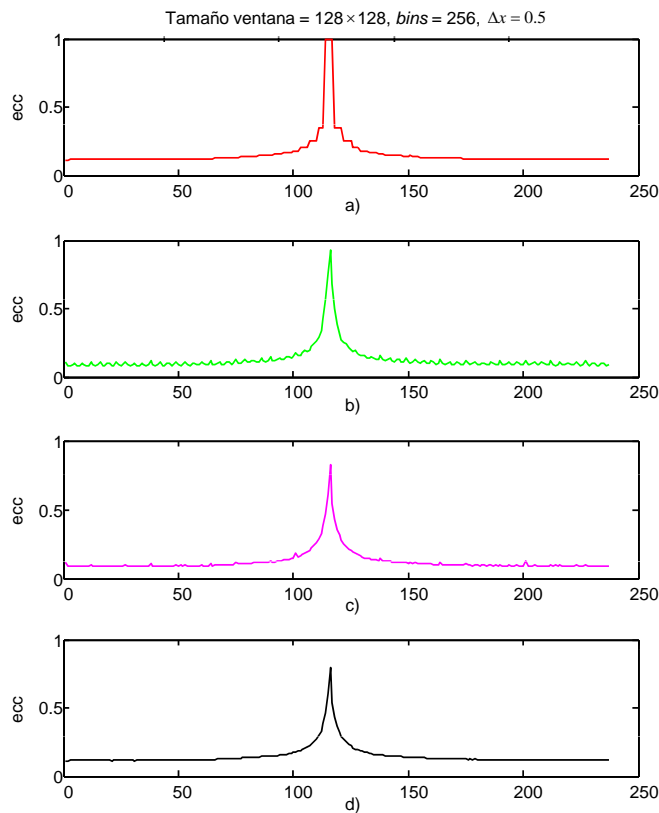


Figura 2.9: Valores de ECC obtenidos al desplazar cada 0.5 píxeles una imagen cuadrada de 128 píxeles de lado sobre la imagen de referencia. La interpolación de los niveles de gris se ha realizado con diferentes funciones: (a) “el vecino más próximo” (NN), (b) bilineal (B), (c) volumen parcial (PV) y (d) interpolación parcial (PI).

Las gráficas 2.9(a-b) recogen los valores de ECC obtenidos al repetir el experimento de la figura 2.6, pero en esta ocasión, en lugar de píxel a píxel, la imagen (de 128 de lado) se desplaza cada 0,5 píxeles ($\Delta x = 0,5$) lo que obliga a interpolar. Como se observa en la figura, las curvas correspondientes a las funciones de interpolación clásicas, el “vecino más próximo” (NN) y bilineal (BI), aunque muestran un pico claro y distintivo, exhiben comportamientos bastantes ruidosos.

Supóngase, por ejemplo, un procedimiento de registro basado en intensidad dirigido por el ECC y que utiliza NN o BI para interpolar los niveles de intensidad de la imagen registrada. El método de optimización actúa sobre los

parámetros de la transformación en función de los valores proporcionados por la medida de consistencia. Si los valores proporcionados por el ECC presenta un comportamiento ruidoso, como el ilustrado en la figura, la optimización podría no converger a los parámetros óptimos.

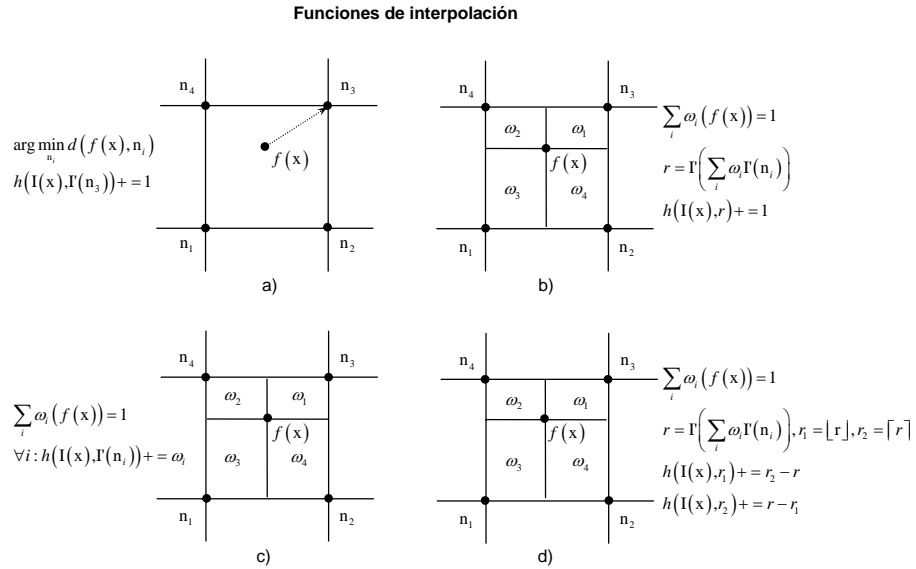


Figura 2.10: Funciones de interpolación: (a) “vecino más próximo” (b) bilineal, (c) volumen parcial y (c) interpolación parcial.

Esta problemática se analiza en profundidad en [67], el cual propone dos nuevos procedimientos de interpolación ideados específicamente para minimizar estos inconvenientes, estos son: el volumen³ parcial (PV) y la interpolación parcial (PI) (ver fig. 2.10). A diferencia del enfoque tradicional, donde las funciones de interpolación son utilizadas para determinar el nivel de intensidad de cada píxel de la imagen registrada, el autor propone un procedimiento análogo para obtener histogramas conjuntos más representativos. En el caso del PV (fig. 2.10(c)) se procede del siguiente modo: una vez determinada la posición del píxel transformado, se toman los vecinos n_i más cercanos, con $i = 1, \dots, 4$; se determinan los coeficientes de ponderación ω_i , para, a continuación, acumularlos en las posiciones $h(I(x), I'(n_i))$ del

³Ambos métodos se propusieron para el registro de volúmenes MR (imágenes 3D), de ahí el nombre de volumen. En este trabajo se analiza la variante 2D de esta función.

histograma conjunto. El procedimiento a seguir con el PI es muy similar al descrito anteriormente. Las gráficas 2.9(c-d) muestran las curvas obtenidas en el experimento anterior utilizando PV y PI. Obsérvese como, con estas funciones, las curvas muestran un comportamiento mucho menos ruidoso.

2.5 Conclusiones

Las medidas de consistencia del registro cuantifican la precisión con la que la imagen registrada se ajusta a la de referencia. En este capítulo se revisan diferentes tipos de medidas utilizadas para este propósito, algunas de ellas miden el error de ajuste de un conjunto de puntos de control localizados en las imágenes (basadas en puntos), otras por el contrario, miden el parecido radiométrico de ambas imágenes comparando píxel a píxel su contenido (basadas en intensidad). En este último grupo cabría destacar por su robustez la MI y sus distintas variantes normalizadas (NMI y ECC). Estas medidas se basan en el concepto de entropía y se caracterizan por medir eficazmente la similitud de imágenes de la misma escena que presentan diferencias radiométricas que no obedecen a una función (diferencias no-funcionales) o han sido captadas con diferentes sensores (análisis multimodal).

La MI se calcula a partir de las funciones de distribución de probabilidad conjunta y marginales de ambas imágenes, por tanto, su eficacia es muy sensible a la correcta estimación de éstas. La estimación de las distribuciones se realiza a partir de un histograma conjunto que se obtiene a partir de los niveles de intensidad de ambas imágenes. De lo representativo que sea este histograma dependerá lo fidedignas que éstas sean y, por tanto, la eficacia de la MI. En este proceso intervienen diferentes parámetros dependiendo del procedimiento utilizado: el tamaño de la muestra (es decir, el tamaño de las imágenes) y el número de acumuladores o *bins* utilizados (estimación paramétrica), la desviación típica (en el caso de la ventana de Parzen) y también, de la función utilizada para interpolar los niveles de intensidad de la imagen registrada.

Este capítulo aporta también una evaluación experimental en la que se analiza la influencia de los parámetros enumerados arriba en la estimación

de las funciones de distribución de probabilidad empleadas en el cálculo de la MI. Para ello se realizan diferentes pruebas experimentales consistentes en registrar diversos pares de imágenes utilizando como medida de consistencia el ECC. De estas pruebas experimentales se pueden extraer diversas conclusiones: el método paramétrico, consistente en acumular directamente los “pares” niveles de intensidad, es eficaz sólo si el número de muestras es suficientemente importante. Cuando el número de muestras decrece produce estimaciones poco precisas, circunstancia que se puede compensar con la reducción del número de acumuladores; la ventana de Parzen ofrece estimaciones mucho menos ruidosas que el método anterior, incluso cuando el número de muestras es pequeño, aunque para ello requiera un mayor coste computacional; y finalmente, respecto a los métodos de interpolación evaluados (NN, BI, PV y PI), los resultados de los tests muestran que los métodos PV y PI ofrecen estimaciones más precisas que los métodos tradicionales.

Capítulo 3

Registro basado en puntos

Las diferencias geométricas que presentan un par de imágenes de la misma escena (denominadas conjugadas) dependen, fundamentalmente, de la geometría de la ésta (normalmente, no-plana) y de las posiciones desde las que éstas son adquiridas. Así, por ejemplo, si las imágenes son capturadas desde prácticamente la misma posición o después de un movimiento de rotación puro [72], las diferencias no son relevantes y su registro se puede efectuar mediante transformaciones simples (rígidas). Por el contrario, si la cámara se aleja de la posición inicial, las diferencias se acentúan y, salvo que la escena sea plana, las funciones rígidas ya no ofrecen resultados aceptables, debiéndose emplear funciones más complejas (no-rígidas o elásticas) que dificultan el proceso de registro. Este capítulo revisa las funciones de transformación más representativas, prestando especial interés a las transformaciones no-rígidas.

Una vez determinada la transformación idónea para abordar el registro se procede a estimar los parámetros de ésta, para lo cual se pueden emplear correspondencias identificadas en ellas (registro basado en puntos) o a la totalidad de sus píxeles (registro basado en intensidad). Este capítulo se centra en el primer enfoque, contribuyendo con evaluación experimental en la que se analiza la precisión de tres técnicas no-rígidas ampliamente utilizadas: polinomial, lineal por trozos y de base radial. Para evaluar su comportamiento se utilizan imágenes de satélite de alta-resolución de una región con una orografía del terreno variada, adquiridas desde diferentes puntos de observación.

Otro de los aspectos analizados en esta comparativa es la influencia en la precisión del registro del número y distribución de las correspondencias utilizadas en la estimación, para lo cual se realizan experimentos con conjuntos de diversos tamaños.

Antes de proceder a describir distintos tipos de funciones de transformación y mostrar los resultados de la evaluación experimental se detallan brevemente las etapas que conforman el registro basado en puntos:

1. se identifica un conjunto de correspondencias o puntos de control (CP) en ambas imágenes. Los CPs, en número y distribución, deben ser representativos de las distorsiones relativas que éstas presentan;
2. se estiman los parámetros de la función de transformación a partir del conjunto de correspondencias seleccionadas. La transformación ha de proporcionar los grados de libertad suficientes para modelar las distorsiones relativas; finalmente,
3. la imagen de entrada se transforma geoméricamente utilizando la transformación estimada, utilizando para el “trasvase” de sus niveles de intensidad alguna función de interpolación (ver sec. 1.3.4).

3.1 Transformaciones rígidas

Las transformaciones rígidas, también denominadas homografías, definen la relación espacial entre proyecciones 2D de un mismo plano 3D, es decir, transformaciones entre planos. Las transformaciones rígidas preservan las rectas (de ahí su nombre), esto es, las rectas en la imagen de entrada continúan siéndolo una vez que ésta es registrada. Sin embargo, pueden no asegurar la invarianza de otras propiedades tales como el paralelismo, la proporción de ángulos, la proporción de segmentos, etc. A continuación se detalla la familia de transformaciones rígidas, así como también, sus respectivas propiedades invariantes (ver fig. 3.1).

Dadas dos imágenes I e I' del mismo tamaño y un conjunto de n correspondencias $\{(x_i, x'_i), i = 1, \dots, n\}$ identificadas en ellas, se definen las siguientes

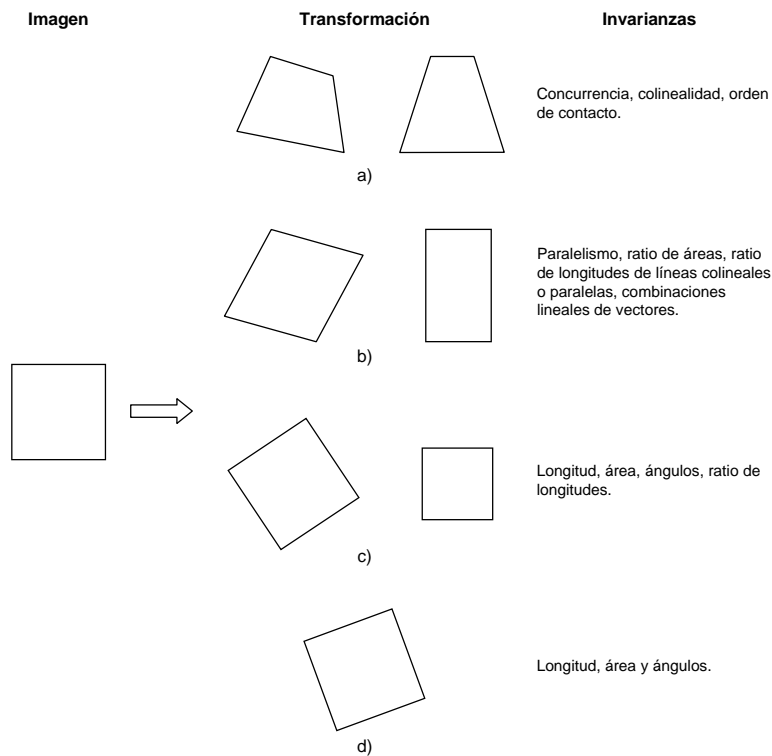


Figura 3.1: Familia de transformaciones rígidas: (a) proyectiva, (b) afín, (c) similitud y (d) Euclídea (isometría).

funciones de transformación:

Similitud: Transformación con 4 grados de libertad dada por la expresión:

$$\begin{aligned} x &= s x' \cos \theta - s y' \sin \theta + t_x \\ y &= s x' \sin \theta + s y' \cos \theta + t_y \end{aligned} \quad (3.1)$$

donde s es un factor de escala; θ una rotación y $t = (t_x, t_y)^\top$ un vector de traslación, respectivamente.

La similitud requiere de al menos 2 correspondencias para su estimación y se caracteriza por preservar la longitud, el área, el ángulos y el ratio de longitudes (ver fig. 3.1(c)). Para el caso particular en que $s = 1$, la similitud se denomina transformación Euclídea o isometría

(ver fig. 3.1(d)).

Afin: La transformación afín viene dada por la expresión:

$$\begin{aligned}x &= x'a_1 + y'a_2 + a_3 \\y &= x'b_1 + y'b_2 + b_3\end{aligned}\tag{3.2}$$

donde los términos de la forma a_i y b_i , $i = 1, 2, 3$, son los coeficientes de la transformación en x e y , respectivamente.

Esta es una transformación con 6 grados de libertad que se caracteriza por preservar el paralelismo, el ratio de áreas, el ratio de longitudes de líneas colineales o paralelas y las combinaciones lineales de vectores (por ejemplo, centroides). Se requieren al menos 3 correspondencias no colineales para su estimación (ver fig. 3.1(b)).

Las funciones afines se utilizan en el registro lineal por trozos para transformar las diferentes regiones triangulares en las que se dividen las imágenes conjugadas. El lector encontrará información más detallada sobre estas transformaciones en las secciones 3.2 y 4.2.

Proyectiva: Transformación con 8 grados de libertad dada por la expresión:

$$\begin{aligned}x &= \frac{x'a_1 + y'a_2 + a_3}{1 + x'c_1 + y'c_2} \\y &= \frac{x'b_1 + y'b_2 + b_3}{1 + x'c_1 + y'c_2}\end{aligned}\tag{3.3}$$

donde los términos de la forma a_i , b_i , $i = 1, 2, 3$ y c_j , $j = 1, 2$ son los coeficientes de la transformación en x e y , y el vector de distorsión perspectiva, respectivamente.

La transformación proyectiva preserva la concurrencia, la colinealidad, el orden de contacto, etc., pero a diferencia de las transformaciones anteriores, no preserva el paralelismo. Para su estimación se necesitan como mínimo 4 correspondencias no colineales (ver fig. 3.1(a)).

Normalmente, el número de correspondencias identificadas en ambas imágenes es superior al mínimo requerido por cada transformación. Esta circunstancia es explotada por diferentes procedimientos de estimación para minimizar el error de ajuste, como por ejemplo el ajuste de mínimo error cuadrático (RMSE) o los estimadores robustos LMedS, RANSAC o MAPSAC, que se describen en el apéndice B.

Por otro lado, obsérvese como cada transformación extiende a la anterior, es decir, la similaridad extiende la Euclídea (similaridad = euclídea + cambio de escala), la afinidad extiende la similaridad, y así sucesivamente. El lector puede dirigirse al libro de Hartley y Zisserman [47] para un estudio completo de este tipo de transformaciones y sus invarianzas geométricas.

3.2 Transformaciones no-rígidas

Salvo en casos muy concretos, tales como imágenes de escenas planas o imágenes de escenas no-planas tomadas desde la misma posición o después de un movimiento de rotación puro, las transformaciones rígidas rara vez son capaces de modelar las distorsiones, típicamente no lineales, que presentan un par de imágenes conjugadas. Son en estas situaciones donde las transformaciones no-rígidas o elásticas son requeridas. A continuación se detallan las transformaciones elásticas más representativas:

Polinomial: Las transformaciones polinomiales se utilizan típicamente con imágenes que presentan diferencias geométricas que afectan a toda la imagen por igual (alcance global) (ver fig. 3.2). Este sería el caso en el que la imagen de referencia y la de entrada son adquiridas desde la misma posición o la escena es prácticamente plana. Este tipo de transformaciones se emplean con frecuencia en el campo de la teledetección para el registro de imágenes que son adquiridas desde una órbita fija y sin cabeceo del satélite [30, 37, 80]. Las funciones polinomiales tienen

$$N = (m + 1)(m + 2) \quad (3.4)$$

grados de libertad, siendo m el grado de las funciones polinomiales.

Requieren, por tanto, de $N/2$ correspondencias como mínimo para su estimación. Lógicamente, funciones polinomiales con grados elevados (esto es, más grados de libertad) modelan diferencias geométricas más severas y viceversa.

Matemáticamente, una transformación polinomial se escribe como:

$$\begin{aligned} x &= \sum_{i=0}^m \sum_{j=0}^i a_{ij} (x')^{i-j} (y')^j \\ y &= \sum_{i=0}^m \sum_{j=0}^i b_{ij} (x')^{i-j} (y')^j \end{aligned} \quad (3.5)$$

donde a_{ij} , b_{ij} son los coeficientes de los polinomios en x e y , respectivamente.

Este tipo de transformaciones no ofrecen resultados aceptables con imágenes que presentan diferencias locales, o sea, que no afectan a toda la imagen por igual, lo que limita significativamente su campo de aplicación. Este hecho queda patente en los resultados de la evaluación experimental que se presenta en la sección 3.3.

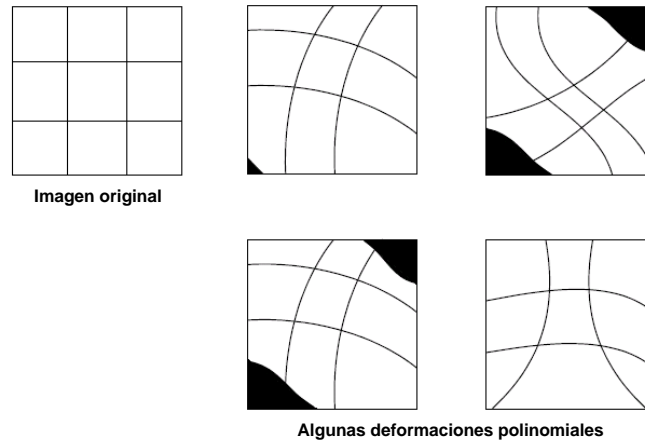


Figura 3.2: Transformaciones polinomiales. Algunos ejemplos.

Lineal por trozos: Las funciones lineales por trozos abordan el registro dividiendo las imágenes en regiones triangulares conjugadas (por ejem-

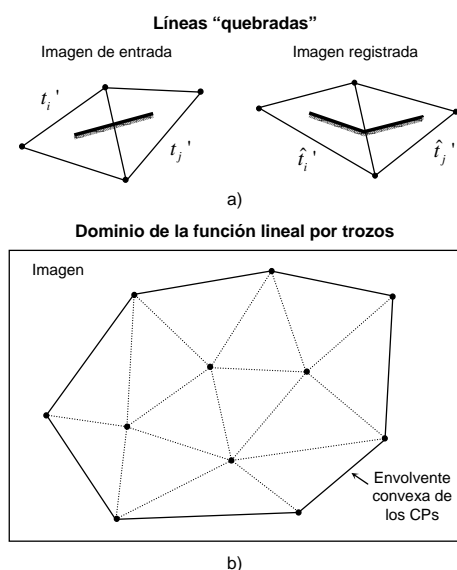


Figura 3.3: Registro lineal por trozos. (a) Efecto de líneas “quebradas” en las transiciones de los triángulos. (b) Dominio de la función lineal por trozos. Esta función sólo está definida dentro de la envolvente convexa del conjunto de puntos identificados en la imagen.

plo, mediante el método de triangulación de Delaunay [93]) que son registradas individualmente mediante transformaciones lineales [6, 39]. Aunque este enfoque garantiza la continuidad en triángulos adyacentes, no produce transiciones suaves entre ellos, lo que puede causar efectos no deseados en la imagen transformada, tales como el “quebramiento” de las líneas, esto es, las rectas pueden no preservarse en las transiciones (ver fig. 3.3(a)). Matemáticamente se define como:

$$\begin{aligned}
 x = f_x(x', y') &= \begin{cases} a_{11} + a_{12}x' + a_{13}y' & \text{si } (x', y') \in t_1 \\ \vdots & \\ a_{n1} + a_{n2}x' + a_{n3}y' & \text{si } (x', y') \in t_n \end{cases} \\
 y = f_y(x', y') &= \begin{cases} b_{11} + b_{12}x' + b_{13}y' & \text{si } (x', y') \in t_1 \\ \vdots & \\ b_{n1} + b_{n2}x' + b_{n3}y' & \text{if } (x', y') \in t_n \end{cases}
 \end{aligned} \tag{3.6}$$

donde t_i es un elemento triangular construido sobre tres corresponden-

cias; a_{ij} y b_{ij} , con $j = 1, 2, 3$, son los coeficientes de la transformación lineal correspondiente a t_i ; y n es el número de elementos que configuran la malla triangular.

Nótese que las funciones lineales por trozos sólo están definidas dentro de la envolvente convexa del conjunto de correspondencias (ver fig. 3.3(b)). Aunque es posible extrapolar fuera de esta región, en la práctica no es aconsejable ya que pueden introducir errores que distorsionan la imagen registrada. En el capítulo 4 se analizan en profundidad este tipo de transformaciones.

Base radial: Las transformaciones de base radial son métodos de interpolación de datos dispersos donde la transformación geométrica viene dada por la combinación lineal de funciones de base radial simétricas centradas en un punto de control (CP) particular [19, 41, 45].

Las funciones de base radial proporcionan deformaciones suaves con un comportamiento fácilmente controlable. En 2D, una función de base radial consta típicamente de una componente global (habitualmente, una transformación afín) que corrige posibles traslaciones, rotaciones, cambios de escala, etc. (primer término de la expresión (3.7)) y una componente local capaz de modelar distorsiones no lineales. Matemáticamente:

$$\begin{aligned} x_i &= \sum_{j=0}^m \sum_{k=0}^j a_{jk} (x'_i)^{j-k} (y'_i)^k + \sum_{j=1}^n A_j g(r_j) \\ y_i &= \sum_{j=0}^m \sum_{k=0}^j b_{jk} (x'_i)^{j-k} (y'_i)^k + \sum_{j=1}^n B_j g(r_j) \end{aligned} \quad (3.7)$$

donde

$$r_j = \|(x'_i, y'_i) - (x'_j, y'_j)\| \quad (3.8)$$

y g es una función de base radial, habitualmente:

$$\begin{array}{ll} \textit{Thin-plate-spline} & g(r_j) = r_j^2 \log r_j^2 \\ \textit{Multi-quadric} & g(r_j) = (r_j^2 + \delta)^{\pm\mu}, \delta > 0, \mu \neq 0 \\ \textit{Gaussiana} & g(r_j) = e^{(-r_j^2/\sigma)}, \delta > 0 \end{array}$$

$$\begin{array}{ll} \textit{Shifted-LOG} & g(r_j) = \log (r_j^2 + \delta)^{\frac{3}{2}}, \delta \geq 0 \\ \textit{Cubic-spline} & g(r_j) = \|r_j\|^3 \end{array}$$

La función base determina la influencia de cada CP sobre cada píxel de la imagen registrada (segundo término de la expresión (3.7)), esto es, el alcance de cada CP. Algunas funciones base tienen una influencia más global (por ejemplo, *thin-plate-spline*) mientras otras tienen una influencia mucho más local (por ejemplo, *Gaussiana*).

El lector puede dirigirse a los trabajos de Goshtasby [42, 51] para un estudio completo de este tipo de transformaciones en términos de precisión y complejidad computacional.

3.3 Comparativa experimental

En la práctica, debido a una amplia variedad de factores como el tamaño y resolución espacial de las imágenes, la geometría de la escena, las posiciones de adquisición, etc., una transformación que ofrece buenos resultados con ciertas imágenes puede no producir resultados aceptables con otras, requiriéndose técnicas capaces de corregir distorsiones más complejas (típicamente, no lineales). Por ejemplo, en el campo de la teledetección, las funciones globales polinomiales muestran habitualmente un buen comportamiento con imágenes de baja- y media-resolución (por ejemplo, Landsat, IRS, Spot, etc.) adquiridas en el nadir, pero pueden no ser suficientemente efectivas para registrar imágenes de alta-resolución (por ejemplo, QuickBird, Ikonos o OrbView) ya

que son capturadas desde posiciones muy dispares¹, lo que da lugar a imágenes con diferencias geométricas muy importantes. (ver fig. 3.4).



Figura 3.4: Imágenes QuickBird de la misma escena adquiridas desde distintos ángulos de observación. Obsérvese las diferencias geométricas (y radiométricas) entre ambas imágenes.

En la literatura se han propuesto numerosos métodos elásticos que abordan el registro de imágenes afectadas de importantes diferencias geométricas, incluyendo: funciones lineales [39] o cúbicas [40] por trozos, funciones *multi-quadric* [29], funciones *thin-plate-spline* [19], funciones *B-spline* [59], etc. La gran mayoría de estos métodos estiman sus coeficientes a partir de un conjunto de puntos de control identificados en ambas imágenes, ya que, en términos generales, son menos costosas computacionalmente que los procedimientos basados en intensidad, que se formulan como procesos iterativos de optimización. Esta tesis contribuye con una evaluación experimental de tres de estos métodos y lo hace respecto a las siguientes cuestiones fundamentales:

- a) cómo modelan las diferencias relativas no-lineales inducidas por la observación de una escena no-plana desde distintos ángulos y
- b) cómo influye en la precisión del registro el número y distribución de los CPs utilizados en la estimación.

Está claro que, si un método de registro simple (por ejemplo, funciones globales polinomiales) logra la exactitud requerida para una aplicación particular,

¹Para ofrecer periodos de re-visita más cortos, estos satélites puede observar la escena desde órbitas y ángulos muy dispares (observación fuera del nadir).

no es necesario perder tiempo seleccionando los CPs extras requeridos por una técnica mucho más potente. El objetivo de esta evaluación es, por tanto, discernir en que circunstancias se aconseja una u otra de las transformaciones analizadas y el número de correspondencias idóneo para obtener resultados aceptables.

Las funciones de transformación evaluadas en esta sección son las siguientes:

- una técnica global basada en funciones polinomiales (de diversos órdenes) (expresión (3.5)),
- un método local basado en funciones lineales por trozos (expresión (3.6)), y
- una técnica híbrida basada en la combinación de una afinidad global y funciones *thin-plate-spline* locales (expresión (3.7)).

En esta evaluación experimental se emplean imágenes de alta-resolución del satélite QuickBird (0.6 meters/píxel) de una región con una orografía variada y adquiridas desde distintos ángulos de observación, lo que da lugar a importantes distorsiones, especialmente, en las regiones montañosas. Para cuantificar la precisión de cada método se utilizan dos medidas de consistencia: el error cuadrático medio (RMSE) (expresión (2.1)) y el error circular con 90 % de confianza (CE90) (expresión (2.2)).

3.3.1 Conjuntos de datos

En esta sección se describe brevemente las imágenes de prueba, así como los conjuntos de correspondencias utilizados en esta comparativa.

Imágenes de prueba

Se han considerado tres imágenes pancromáticas (etiquetadas como i_1 , i_2 e i_3) de la ciudad del Rincón de la Victoria (Málaga) adquiridas en diferentes fechas y desde diferentes puntos de vista. La figura 3.5(a) muestra dos de estas imágenes de prueba, en las que aparece marcadas las regiones de interés:

terreno casi plano (área urbana) y accidentado (área montañosa). Para definir dichas regiones se han utilizado las curvas de nivel mostradas en la figura 3.5(b). La gráfica mostrada en la parte superior de la figura 3.5(a) muestra el perfil del terreno² a lo largo de la flecha de la imagen de la derecha (desde la montaña hasta la costa) donde los niveles de altitud van desde los 300 m. hasta los 0 m.

Los pares de imágenes considerados para el registro son <i1-i2>, <i1-i3> y <i2-i3>. Estos pares presentan ángulos de observación relativos muy dispares que, en combinación con el relieve del terreno (el mismo para todas ellas), dan lugar a distorsiones geométricas como las observadas en la figura 3.4.

Conjuntos de puntos de control

Para cada par de imágenes se extraen, de manera automática mediante el procedimiento descrito en la sección 4.5.1, dos conjuntos diferentes de correspondencias: puntos de control (CP) para estimar los coeficientes de la función de transformación, y puntos independientes de control (ICP) para evaluar la precisión del registro. Para garantizar su distribución uniforme sobre las imágenes, se selecciona un punto por cada celda de una retícula rectangular dispuesta sobre las imágenes (ver fig. 3.6). Actuando sobre el tamaño de la celda se obtienen diferentes densidades, en particular, se han definido los siguientes: 50 píxeles de ancho para los conjuntos de ICPs y 100, 200 y 400 píxeles para los conjuntos de CPs, que para el caso concreto de las imágenes utilizadas, de 0.6 m./píxel de resolución, se corresponden con 30, 60, 120 y 240 m., respectivamente.

De esta manera se dispone de 900, 225 y 64 CPs y un conjunto de 3600 ICPs por cada par de imágenes. Aparte de estos conjuntos de datos, que nos permiten estudiar la precisión de los métodos en regiones en la que se combinan zonas planas con otras con una geometría más compleja (geometría mixta), se consideran las regiones de interés marcadas en la figura 3.5(a),

²La información de elevación ha sido obtenida de un DEM con una resolución espacial de 20×20 m. proporcionado por la “Consejería de Medio Ambiente” de la “Junta de Andalucía”. Este es el DEM más preciso de la zona disponible actualmente.

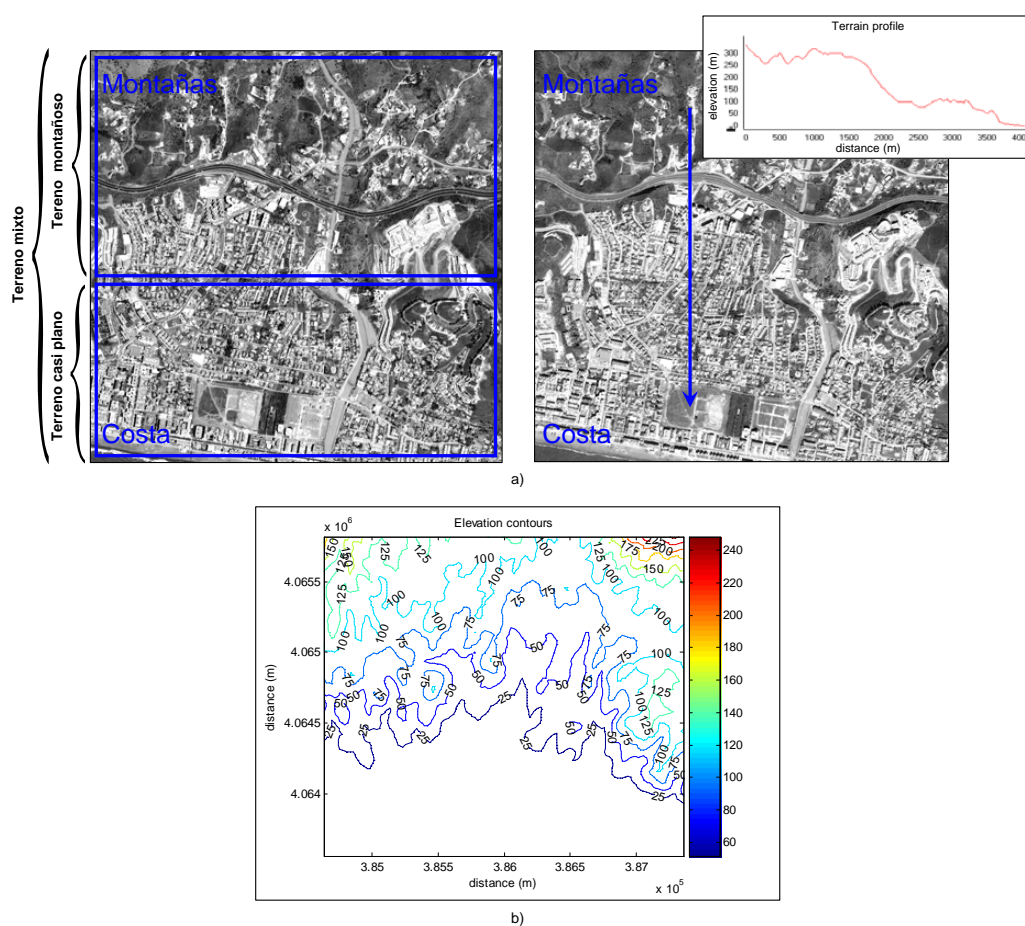


Figura 3.5: Imágenes de prueba utilizadas en el estudio comparativo. (a) Dos imágenes de la ciudad del Rincón de la Victoria (Málaga). Estas imágenes cubren aproximadamente un área de 4 km^2 e incluyen zonas con distintos perfiles de terreno (ver el perfil del terreno a lo largo de la flecha). (b) Información de elevación utilizada para definir apropiadamente las regiones de interés: áreas casi planas y de relieve accidentado.

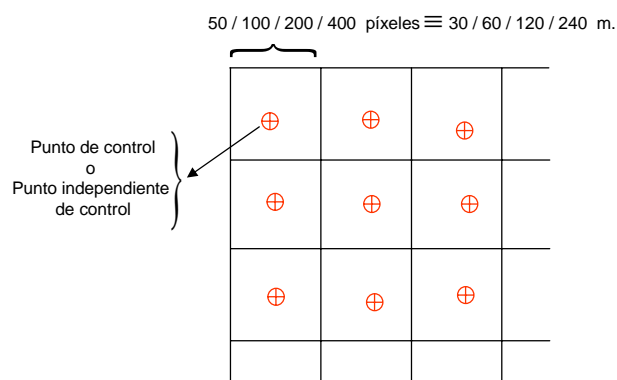


Figura 3.6: Procedimiento utilizado para distribuir uniformemente los CPs e ICPs sobre las imágenes. Concretamente, las imágenes se dividen de acuerdo a una rejilla rectangular de tamaño conocido y se elige un punto por cada celda. El tamaño de las celdas utilizados en este trabajo son 50 píxeles para el conjunto de ICPs y 100, 200 y 400 píxeles para los conjuntos de CPs.

obteniendo dos subconjuntos: uno para la zona urbana (casi plana) y otro para la zona montañosa (mucho más accidentada).

Aunque el procedimiento automático de identificación de los pares de CPs e ICPs (ver sec. 4.5.1), garantiza precisión y un número importante de puntos consistentemente emparejados, no garantiza que éstos estén situados adecuadamente para capturar las diferencias relativas de ambas imágenes (en el caso de los CPs) o la consistencia del registro (en el caso de los ICPs). Por ejemplo, un ICP perfectamente emparejado, pero situado en el techo de un edificio de 10 plantas, no es apropiado para medir cómo de bien se ha registrado el suelo o una urbanización de viviendas unifamiliares de la misma zona. Este problema queda patente en alguno de los experimentos de la siguiente sección.

3.3.2 Experimentos y resultados

En esta sección se analiza el rendimiento de las funciones polinomiales (desde 2º grado hasta 6º), lineales por trozos y *thin-plate-spline* para diferentes conjuntos de CPs, ángulos de observación y geometrías de la escena. Puesto que la influencia de cualquiera de estos parámetros en la precisión del re-

gistro depende de los valores fijados para el resto, un análisis exhaustivo de ellos requeriría pruebas sobre todas las posibles combinaciones. Aunque se han analizado todas las combinaciones posibles para los datos disponibles (imágenes y CPs), sólo se incluyen las comparativas correspondientes a las más representativas.

Respecto a la estimación de los parámetros de las funciones, se ha empleado un software desarrollado en C++ que utiliza las primitivas de *The OpenCV Library*³ para ganar en eficiencia. Resaltar también los siguientes puntos:

- Se ha verificado experimentalmente que los polinomios de grado superior al 4º apenas mejoran la precisión, aunque requieren una mayor número de CPs para ser estimados. Consecuentemente, sólo se muestran los resultados correspondientes al 4º grado, que requiere un mínimo de 15 CPs.
- Como se detalla en la literatura [66], cuando el número de CPs es alto (por ejemplo, más de 1000) el sistema de ecuaciones que se ha de resolver para estimar los coeficientes las funciones *thin-plate-splines* está mal condicionado. Para solventar este problema se ha utilizado el método iterativo propuesto en Beatson *et al.* [18].

La figura 3.7(a) muestra la influencia del número de CPs en la precisión de cada método. En estas gráficas se fijan el ángulo de observación (28.7°, correspondiente al par <i1-i2>) y la geometría de la escena (geometría mixta). A la vista de estos resultados se puede concluir que: los métodos locales (PWL y TPS) aprovechan el número de CPs, logrando resultados muy precisos para el conjunto más denso (sobre 1.75 m. RMSE); mientras que, por el contrario, las funciones globales (POL4) no mejoran su rendimiento con conjuntos más numerosos: empleando más de 15 CPs sólo se gana robustez en la estimación. Nótese también que, cuando el número de CPs decrece

³La librería *open source* de visión artificial es una colección de algoritmos y código de ejemplo para abordar varios problemas típicos en visión por computador. La librería OpenCV es compatible con Intel® Image processing Library (IPL) y utiliza las Intel® Integrated Performance Primitives (IPP) para un mejor rendimiento. Visitar [100] para más información.

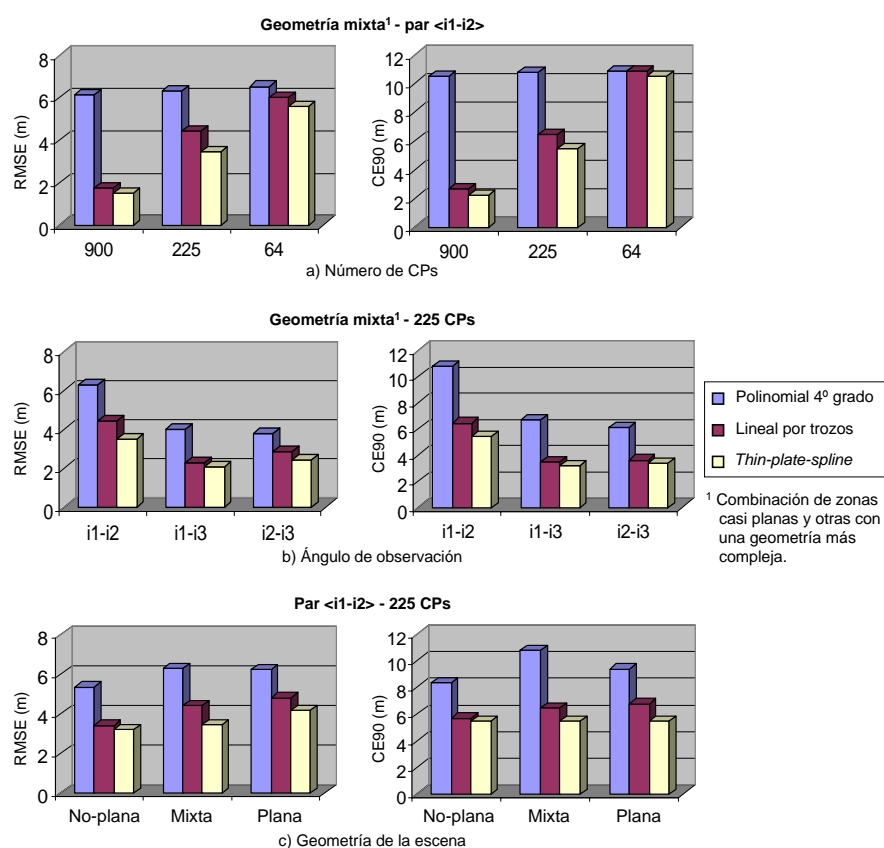


Figura 3.7: Valores RMSE y CE90 para las técnicas elásticas analizadas en esta comparativa atendiendo a: (a) número de CPs, (b) ángulo de observación (par de imágenes) y (c) geometría de la escena, respectivamente. Sobre cada gráfica se indica los valores fijados para los otros parámetros.

(conjuntos de CPs dispersos) las transformaciones locales exhiben un comportamiento global (esto es, una influencia local débil) y no pueden adaptarse adecuadamente a las diferencias locales entre ambas imágenes.

La figura 3.7(b) muestra la influencia del ángulo de observación para una geometría mixta y conjuntos de 225 CPs (tomados con celdas de 120 m. de ancho). En esta comparativa se observa que todos los métodos son sensibles a ángulo de observación, aunque esta influencia es mayor para el método global (POL4). Aunque los resultados no se incluyen aquí, para un conjunto de CPs más denso, POL4 ofrece unos resultados similares (no se aprecia mejora alguna con respecto a los ofrecidos en la figura 3.7(b)), mientras que

los métodos locales proporcionan resultados más precisos, especialmente para el par $\langle i1-i2 \rangle$, que exhibe las diferencias más importantes en sus ángulos.

Finalmente, en la figura 3.7(c) se muestran los resultados obtenidos dependiendo de la complejidad de la geometría de la escena. En esta gráfica se puede subrayar lo siguiente: los métodos locales trabajan de forma similar para todas las regiones (plana, accidentada y mixta), mientras que la función polinomial ofrece peores resultados con una geometría mixta. Este comportamiento tiene sentido, ya que este tipo de transformaciones no se puede ajustar adecuadamente y al mismo tiempo, a modelos de deformación diferentes. La figura 3.8 (que también contiene a la figura 3.7(c)) ilustra los efectos de observar una escena con una geometría no-plana desde puntos de observación muy dispares (las diferencias de los ángulos en los pares $\langle i1-i3 \rangle$ y $\langle i2-i3 \rangle$ son muy similares). Obsérvese que los errores de registro obtenidos para la zona casi plana son ligeramente mayores que para la zona accidentada. Este resultado se atribuye a la existencia, como se comentó en la sección 3.3.1, de un importante número de ICPs identificados en la parte de superior de los edificios.

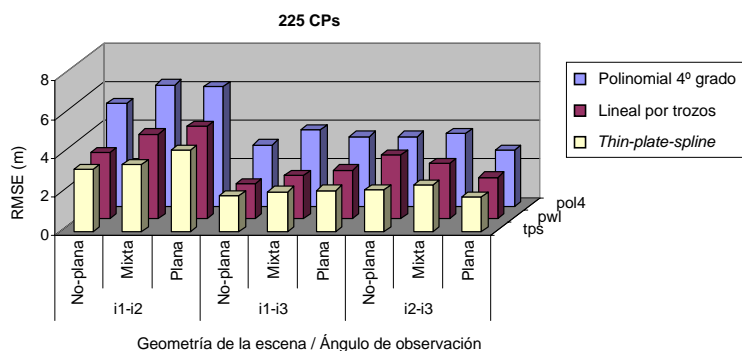


Figura 3.8: Valores RMSE para las técnicas analizadas agrupados por ángulo de observación y geometría de la escena.

Como se detalla en la sección 3.2, las funciones lineales por trozos pueden causar en la imagen registrada un efecto de “líneas quebradas” en las transiciones de los triángulos, que se manifiesta en los elementos rectos de la escena tales como carreteras, edificios, etc. Se ha comprobado visualmente que este problema sólo es significativo cuando la geometría de la escena varía

sustancialmente (esto es, en el área montañosa) y se utiliza la red triangular generada a partir del conjunto de CPs menos denso (sólo 64 CPs para la imagen completa). En estos casos, para evitar estos efectos indeseados, deberían considerarse otros métodos de registro tales como: registro cúbico por trozos [40] o *thin-plate-spline* [19].

Resumiendo, de estos resultados se puede extraer las siguientes conclusiones:

1. Los métodos locales (particularmente, *thin-plate-spline*) ofrecen mejores resultados que las funciones polinomiales en todos los experimentos. Cuando se emplea un método local, es deseable tener tantos CPs como sea posible. No ocurre así con los métodos globales, en los que no se gana en precisión con el empleo de muchos CPs (salvo robustez).
2. Cuando la escena observada contiene elementos a alturas muy dispares (esto es, regiones montañosas, objetos de distinto tamaño, etc.) los métodos globales sólo son adecuados cuando las posiciones de adquisición son relativamente próximas.

De modo que, y corroborando lo que en principio se podría suponer por lógica, excepto para aquellos casos donde se conoce con total certeza que la escena observada es prácticamente plana, sin elementos de diferente altura, o las dos imágenes han sido capturadas desde prácticamente la misma posición, se aconseja el empleo de métodos locales con tantos CPs como sea posible (cuanto más grande sea el número de CPs, mejor será el registro).

3.4 Conclusiones

Las diferencias geométricas relativas de un par de imágenes conjugadas dependen básicamente de la geometría de la escena y las posiciones desde las que éstas son adquiridas. La complejidad de estas diferencias determinará el tipo y alcance de la función de transformación a utilizar. Así, por ejemplo, se puede emplear una homografía para registrar imágenes de una escena plana o adquiridas después de una rotación pura (sin traslación) del sensor.

Sin embargo, si la escena no es plana (como ocurre normalmente) y se ha producido un desplazamiento del sensor debería considerarse algún tipo de transformación elástica capaz de modelar diferencias no lineales.

Este capítulo ha revisado las funciones de transformación (rígidas y elásticas) más representativas y ha desarrollado un análisis experimental en el que se ha evaluado el rendimiento de tres métodos de registro elástico ampliamente conocidos: funciones globales polinomiales, lineales por trozos y *thin-plate-spline*. El interés de esta comparativa se centra en valorar cómo afectan la geometría de la escena y la posición desde la que ésta es observada en su precisión. Otro de los aspectos analizados en esta comparativa es el efecto sobre la precisión del registro del tamaño del conjunto de correspondencias utilizado en la estimación.

Los resultados obtenidos vienen a confirmar algunos aspectos ya mencionados en secciones previas, pero lo más importante, cuantifica la influencia de estos elementos en el rendimiento de cada uno de los métodos analizados. En general, los métodos locales (TPS o PWL) exhiben un comportamiento superior al global (POL4) ya que pueden explotar la información proporcionada por muchos CPs, lo que obliga, por otro lado, a localizar un gran número de ellos. Mientras que, si la geometría de la escena no es demasiado compleja, basta una transformación polinomial y unos pocos pares para obtener resultados bastante buenos.

Capítulo 4

Registro lineal por trozos

La observación de una escena no plana desde puntos de vista distintos produce típicamente imágenes con diferencias geométricas difíciles de modelar. En el capítulo 3 se describen diferentes técnicas elásticas que tratan de abordar su registro y se realiza un análisis experimental de sus comportamientos. Este análisis concluye que, de las técnicas analizadas, sólo las transformaciones lineales por trozos (PWL) y *thin-plate-splines* (TPS) producen unos resultados aceptables, en términos de precisión, cuando estas diferencias locales existen.

Cuando la escena observada es poliédrica (típica en escenarios de interiores y urbanos), el rendimiento de las TPS cae significativamente en favor de las PWL, que continúan ofreciendo un comportamiento adecuado. Obsérvese en la figura 4.1 cómo la transformación lineal por trozos registra con precisión la imagen de entrada, mientras que la transformación *thin-plate-spline* introduce distorsiones adicionales en la imagen registrada.

Recuérdese que las transformaciones lineales por trozos dividen las imágenes a registrar en triángulos conjugados que son transformados individualmente mediante una transformación afín. Sin embargo, para que este método trabaje adecuadamente, cada par de triángulos correspondientes deben caer sobre las proyecciones de una superficie 3D plana, en otro caso, el registro puede generar artefactos indeseados, tales como “líneas quebradas” que reducen la calidad del registro (ver fig. 4.2).

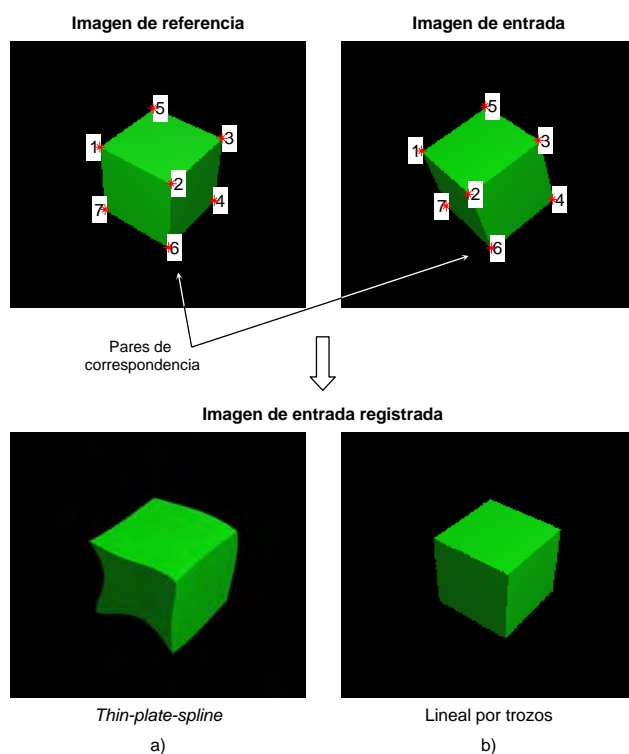


Figura 4.1: Registro de un par de imágenes de una escena poliédrica mediante una transformación: (a) *thin-plate-spline* y (b) lineal por trozos. Obsérvese la precisión de la transformación lineal por trozos, mientras que la *thin-plate-spline* produce una imagen registrada con importantes distorsiones.

Las implementaciones actuales del registro lineal por trozos incluidas en los paquetes científicos de procesamiento de imágenes tales como Matlab [99], Insight Segmentation and Registration Toolkit (ITK) [79], el software de Image Fusion Systems Research [51] o paquetes de teledetección tales como: ENVI/IDL [54], ERDAS [61], etc. generan las redes triangulares conjugadas a partir de un conjunto de correspondencias, utilizando para ello alguna técnica de triangulación, típicamente el método de Delaunay [93]. La triangulación de Delaunay de un conjunto de vértices maximiza el ángulo entre todas las posibles de triangulaciones de ese conjunto de vértices, pero no asegura que éstos se distribuyan sobre las imágenes de forma óptima, esto es, cayendo sobre proyecciones de trozos planos de la escena.

En este capítulo y el siguiente se aborda la generación de redes triangu-

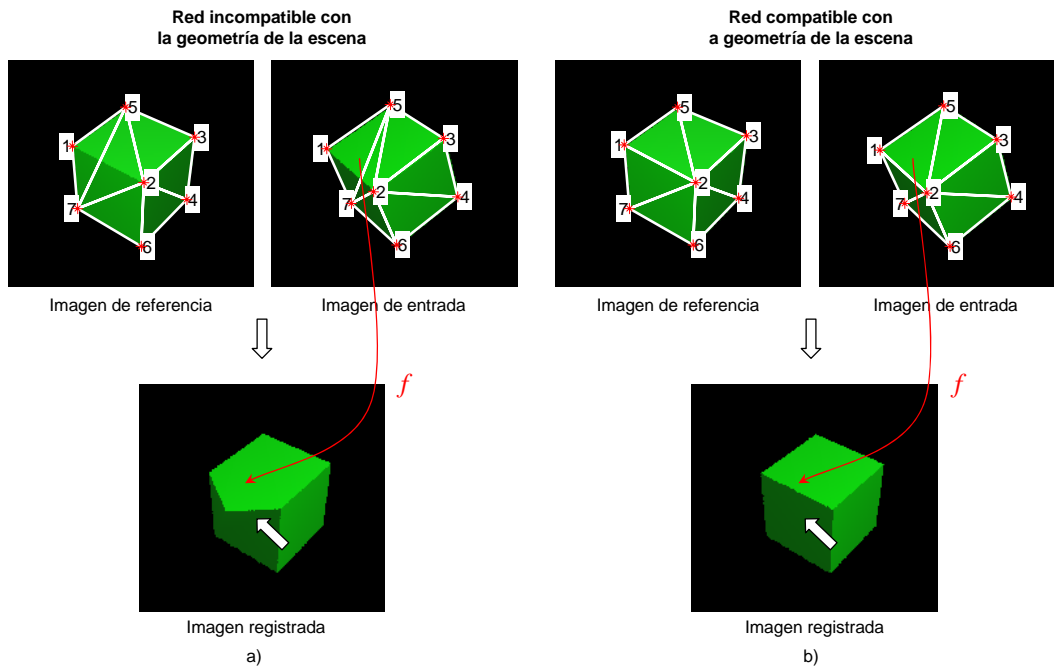


Figura 4.2: Para que un registro lineal por trozos trabaje adecuadamente, los triángulos correspondientes deben ser proyecciones de una única superficie plana de la escena, como el triángulo $\{1, 2, 5\}$ en (b); en otro caso, se produce el “quebramiento” de líneas y la consistencia del registro decrece, como se aprecia en (a).

lares óptimas para mejorar la consistencia del registro lineal por trozos. En este primero, se realiza una revisión de las distintas técnicas de optimización de redes triangulares propuestas en la literatura, poniendo especial énfasis en aquellas propuestas en el campo del registro de imágenes; también formaliza el registro lineal por trozos. En el capítulo 5, se describe el procedimiento de registro propuesto, se analizan diversas estrategias de optimización y se presentan algunos resultados experimentales.

4.1 Optimización de redes

La generación de redes triangulares óptimas es crucial en una amplia variedad de campos, como el modelado de objetos, aproximación de superficies, compresión y transmisión de imágenes, etc. Aunque el objetivo final es aproximar, tanto como sea posible, algún conjunto de datos (ya sea 2D o 3D)

mediante superficies triangulares por trozos, el objetivo concreto de las técnicas de optimización de redes varia con el tipo de problema.

Así, en el modelado de objetos la optimización está dirigida a generar redes triangulares que representen apropiadamente la forma 3D del cuerpo o escena, utilizando para ello el menor número de triángulos posible. En procesamiento de imágenes, por el contrario, se pueden encontrar enfoques interesantes tales como la triangulación basada en datos (DDT) [28] que se utiliza, entre otros, para aproximar el contenido de una imagen (por ejemplo, reconstrucción de imágenes [112]), para reducir la cantidad de información redundante (por ejemplo, compresión de imágenes [60]) o para aproximar un conjunto discreto de muestras de una imagen mediante una superficie lineal por trozos (problema conocido como interpolación de imágenes [103]).

Las técnicas de optimización de redes se pueden clasificar de acuerdo con los siguientes aspectos:

- el mecanismo utilizado para modificar la red (esto es, el tipo y alcance de las acciones),
- la métrica empleada para evaluar la bondad de una configuración de red dada (esto es, la función de energía o coste) y
- el procedimiento para realizar el refinamiento de la red (es decir, el método de optimización).

Atendiendo al tipo y alcance las acciones aplicadas para modificar la red (ver fig. 4.3), en la literatura se pueden encontrar técnicas donde:

1. se modifica la realización topológica mediante el intercambio de aristas (*edge swapping*) [77, 78, 82, 112],
2. se modifica la realización geométrica mediante el refinamiento de las coordenadas de los vértices [69, 89, 92, 111] y
3. se modifican simultáneamente la realización topológica y geométrica mediante la división/eliminación/intercambio de aristas y el refinamiento de las coordenadas de los vértices [27, 49, 50, 105].

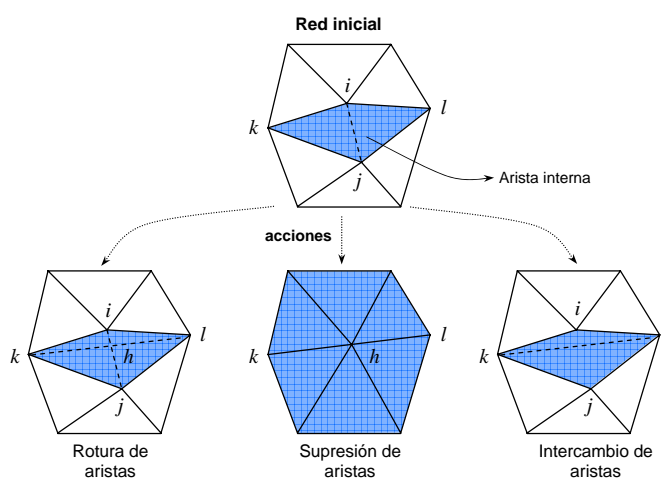


Figura 4.3: Acciones basadas en aristas empleadas para modificar la topología/geometría de una red dada.

Muchos de estos métodos se han desarrollado en el campo del modelado geométrico para simplificar y/o refinar una red 3D inicialmente obtenida a partir de un conjunto muy denso de vértices proporcionado por un sensor 3D, típicamente, un escáner láser [49, 50]. Métodos similares han sido utilizados más recientemente para reconstruir escenas 3D [77, 78, 105] y en diferentes aplicaciones del enfoque DDT para procesamiento de imágenes [15, 60, 112].

A diferencia de estos métodos, que trabajan sobre redes 3D, en el registro lineal por trozos de imágenes se parte de dos redes 2D conjugadas que se modifican en un intento de maximizar la consistencia del registro. Un ejemplo de esto es el trabajo de Servais [92], que refina la localización de los vértices (la topología de la red permanece fija) para compensar el movimiento afín en secuencias de video. Aunque permite modelar distorsiones suaves, el refinamiento de las coordenadas de los vértices no proporciona la flexibilidad suficiente para modelar las importantes diferencias geométricas que se producen cuando las imágenes son adquiridas desde ángulos muy dispares.

Típicamente, las técnicas de optimización se formulan como procesos de minimización (o maximización) que van desde búsquedas aleatorias [78], a procedimientos más complejos basados en el “recocido simulado” [91], mo-

delos estocásticos bayesianos [105], enfoques variacionales¹ [79], etc. Independientemente de la técnica de optimización utilizada, uno de los puntos clave en este proceso es la definición de la función de energía o coste idónea para evaluar la mejora obtenida al aplicar cierta acción sobre una configuración de red dada. Mientras que en otros campos (por ejemplo, modelado de objetos, ajuste de superficies, interpolación de imágenes, etc.) medir la calidad de una red puede realizarse sobre los puntos 3D o las muestras de la imagen disponibles, en registro de imágenes se ha de confiar en la similitud radiométrica de la imagen de referencia y la registrada. Hasta ahora, diversas métricas de similitud han sido utilizadas para este propósito: SSD [77], NCC [82], plantillas basados en diferencias de imágenes [78], etc. Como se detalló en la sección 2.3.1, ninguna de estas medidas son invariantes a las diferencias radiométricas no-lineales en las imágenes, como puede ocurrir en el caso de imágenes capturadas con diferentes cámaras, o en imágenes adquiridas con el mismo sensor, pero en condiciones de iluminación muy diferentes que pueden acarrear sombras, saturaciones, reflexiones, etc. En este capítulo se propone un proceso dirigido por una función de coste basada en la información mutua (MI). Aunque la MI ha sido utilizada como medida de consistencia del registro en numerosos trabajos [23, 24, 52, 57, 109], esta es la primera vez que se integra en un procedimiento de optimización de redes para el registro lineal por trozos.

4.2 Planteamiento del problema

En numerosas aplicaciones, la proyección de la escena sobre el sensor puede aproximarse mediante una transformación para-perspectiva [47, 70, 110], también denominada afín o paralela. Este modelo de transformación (que se ilustra en la fig. 4.4) asume que los puntos de un objeto se proyectan primero sobre un *plano de profundidad promedio* paralelo al plano imagen, para seguidamente, ser transferidos al plano imagen. Los puntos se proyectan sobre el plano promedio usando rayos paralelos al eje imaginario OG , que une el

¹En www.itk.org, el lector puede encontrar una amplia variedad de código abierto en el que se implementa muchas de estas técnicas: campos potenciales, cuerpos elásticos, etc.

centro óptico de la cámara O y el centro del objeto G , preservándose de este modo el paralelismo. Esta simplificación es válida en aquellas configuraciones donde los efectos de la distorsión perspectiva son pequeños, esto es, las líneas paralelas en el espacio continúan siéndolo en las imágenes.

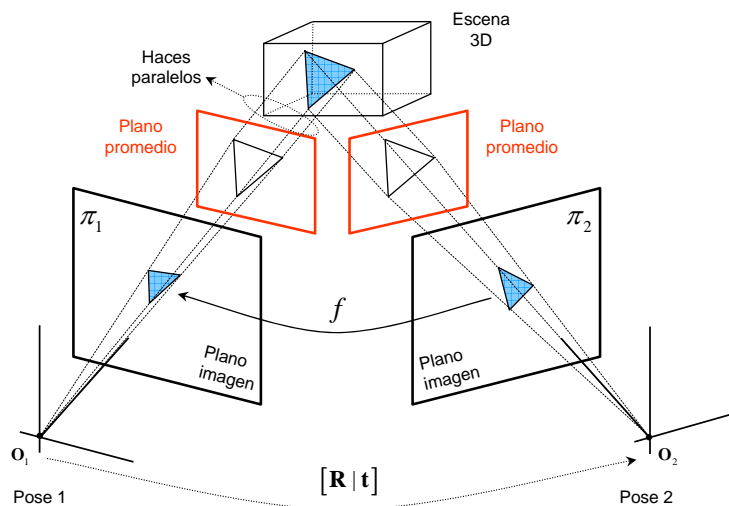


Figura 4.4: Proyección para-perspectiva de una cámara. Cuando el campo de visión de la cámara es pequeño y el tamaño del objeto respecto a la distancia que lo separa del sensor también lo es, la proyección perspectiva (una transformación no lineal) puede ser aproximada mediante una proyección para-perspectiva, esto es, una transformación lineal.

Las cámaras afines dan lugar a una importante reducción en la complejidad de numerosos problemas de visión. En particular, para el problema de registro, esta simplificación significa que tres correspondencias (no-colineales), en lugar de las cuatro necesarias en su forma general, son suficientes para estimar la transformación afín (homografía) que transfiere puntos de un trozo de imagen a otro [47]. Así, cuando se realiza un mapeo afín entre dos regiones triangulares conjugadas de ambas imágenes, estas deben alinearse perfectamente (esto es, la consistencia del registro es máxima); en caso contrario, las regiones triangulares son proyecciones de una superficie no plana.

En la figura 4.5 se muestra el resultado de registrar, mediante una transformación lineal por trozos (PWL), imágenes de una escena plana afectadas de deformación perspectiva y afín, respectivamente. Obsérvese como, mien-

tras la primera presenta diversos artefactos (indicados con flechas en la imagen) que deterioran la calidad del registro, la segunda se registra con absoluta precisión. Como se indica arriba, suponer que la transformación que alinea con precisión dos regiones triangulares conjugadas es una afinidad sólo es válido en aquellas configuraciones cámara-escena que dan lugar a imágenes en las que la distorsión perspectiva es inapreciable.

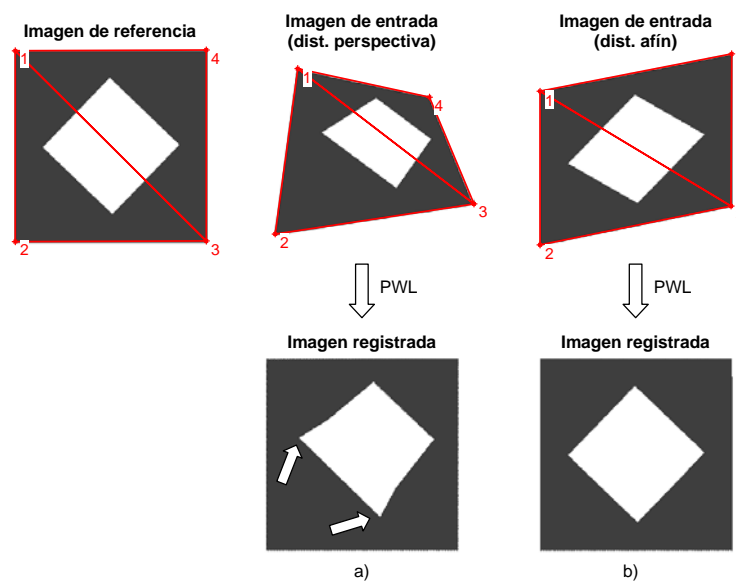


Figura 4.5: Registro lineal por trozos (PWL) de imágenes afectadas de distorsión (a) proyectiva y (b) afín. Los triángulos empleados para la transformación PWL son el $\{1, 2, 3\}$ y el $\{1, 3, 4\}$. Obsérvese en (a) los artefactos (indicados con flechas) que se producen al aplicar este tipo de transformación a imágenes con distorsiones perspectivas.

Resumiendo, si se logra modificar la topología y/o geometría de la red triangular utilizada en un proceso genérico de registro lineal por trozos mediante acciones dirigidas a maximizar el número de elementos triangulares que se registran con precisión, esto es, que caen sobre proyecciones de superficies 3D planas: a) se eliminan configuraciones indeseadas como las ilustradas en la figura 4.2 y b) se mejora la precisión global del registro.

En esta tesis se explota la relación entre las redes triangulares 2D que se utilizan en el registro PWL y su correspondiente malla triangular 3D que modela la escena observada. En base a esto se formaliza, no sólo la generación

de redes triangulares 2D óptimas (desde el punto de vista de la consistencia del registro), sino también la reconstrucción a partir éstas de la superficie 3D observada, yendo de este modo, más allá de un puro enfoque 2D del registro.

A continuación se introducen los conceptos y definiciones utilizados en la formulación del registro lineal por trozos y del método de optimización propuesto.

4.3 Formalización del problema

Supóngase una escena cualquiera cuya geometría es aproximada mediante una superficie lineal por trozos, esto es, un conjunto de triángulos conectados unos con otros conformando una red triangular en \mathbb{R}^3 (ver fig. 4.6). Aunque existen diversas maneras de formalizar mallas triangulares [60, 77, 78, 105], en esta tesis se emplea el *complejo simplicial*, una estructura matemática empleada en topología algebraica para definir espacios topológicos, sus invariantes, las relaciones entre símlices, etc.

El complejo simplicial aporta ciertas ventajas respecto a otras formalizaciones, entre otras:

- Modela con una misma estructura puntos, segmentos, triángulos y conjuntos de estos, así como sus relaciones topológicas (conexión, adyacencia, etc.) y sus invariantes topológicos (huecos, túneles, etc.).
- Separa las realizaciones topológica y geométrica de la estructura, de modo que se puede actuar sobre la realización topológica con independencia su realización geométrica, y viceversa.
- Facilita el almacenamiento, la actualización y el acceso eficiente a los diferentes elementos de la estructura del complejo mediante matrices de adyacencia y conexión.

Gracias a estas características, los complejos simpliciales pueden modelar estructuras topológicas con independencia del espacio en el que se encuentren definidas: el espacio o el plano. Aunque, a primera vista, esto puede parecer de un interés marginal en el contexto del registro de imágenes, mantener la

dualidad espacio-plano es crucial en aplicaciones como la reconstrucción 3D a partir de imágenes. Concretamente, si se resuelve con precisión el registro lineal por trozos, se dispondrá de una buena aproximación triangular de la escena.

En esta sección se formaliza el registro de imágenes por trozos, así como también, la interpretación tridimensional de éste. Esta formalización se explota en el siguiente capítulo, no sólo para registrar las imágenes captadas por un par de cámaras, sino para reconstruir tridimensionalmente el trozo de escena captada por éstas, e ilustrar así los beneficios de emplear divisiones triangulares de las imágenes que sean topológica y geoméricamente consistentes.

Antes de proceder a la lectura de la siguiente sección es aconsejable que el lector se familiarice con algunos de los términos y conceptos relacionados con los Complejos Simpliciales que se describen en el apéndice C.

4.3.1 Definiciones previas

Una red triangular es una tupla $M = (K, V)$ donde K es un complejo simplicial abstracto que especifica la estructura de la red, esto es, los elementos (triángulos, aristas y vértices) y la conectividad de éstos y V es un conjunto de posiciones de los vértices que definen la geometría de K en \mathbb{R}^2 (plano imagen) o \mathbb{R}^3 (escena).

Para referirse a cualquier punto dentro de M se emplea la notación $p \in \phi_V(|K|)$, donde $\phi_V(|K|)$ es la realización geométrica de M (ver def. C.5). Por ejemplo, $p \in \phi_V(|e|)$ con $e = \{i, j\} \in K$ se refiere a un punto de la arista e ; $p \in \phi_V(|t|)$ con $t = \{i, j, k\} \in K$ se refiere a un punto del triángulo t ; $p \in \phi_V(|q|)$ con $q = \{\{i, j, k\}, \{i, j, l\}\} \in K$ se refiere a un punto del cuadrilátero q y así sucesivamente.

4.3. FORMALIZACIÓN DEL PROBLEMA

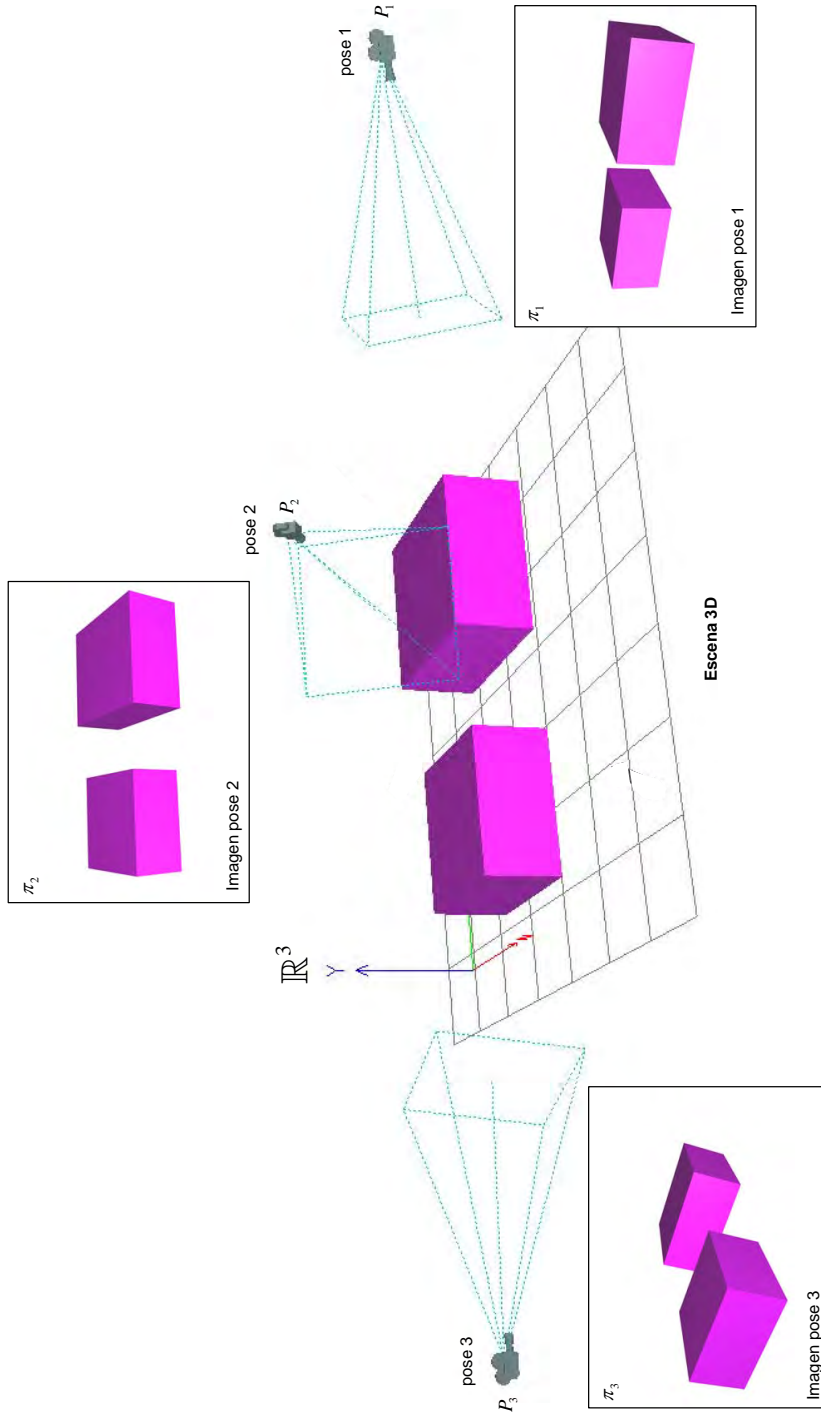


Figura 4.6: Una configuración escena-cámara sintética donde objetos poliédricos son observados desde diferentes posiciones.

Considérese una configuración escena-cámara como la ilustrada en la figura 4.6, donde se tiene una escena poliédrica y un conjunto de imágenes de ésta capturadas desde diferentes posiciones:

Definición 4.1 Se define la red triangular $M^3 = (K, V^3)$ que aproxima la escena observada, donde K consta de un conjunto de m vértices junto con un conjunto de símlices de K y $V^3 = \{X_j = (x_j, y_j, z_j)^T, j = 1, \dots, m\}$ son las posiciones de los m vértices en \mathbb{R}^3 (ver fig. 4.7).

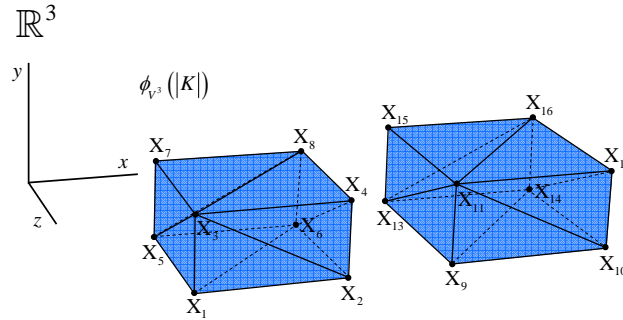


Figura 4.7: Una red triangular en \mathbb{R}^3 que aproxima la escena de la figura 4.6.

Definición 4.2 Se definen las redes triangulares $M_i^2 = (s_i, V_i^2)$, con $i = 1, \dots, c$ (siendo c el número de cámaras), que aproximan las proyecciones de la escena sobre los planos imagen π_i , donde $s_i \subseteq K$ consta de n_i vértices, $n_i \leq m$, junto con un conjunto de símlices de s_i y $V_i^2 = \{x_{i,j} = (x_{i,j}, y_{i,j})^T, j = 1, \dots, n_i\}$ son las coordenadas de los n_i vértices en \mathbb{R}^2 (ver fig. 4.8).

Se dice que las realizaciones geométricas $\phi_{V_i^2}(|r|)$ y $\phi_{V_j^2}(|r|)$ se pueden superponer, sí y sólo sí, $r = (s_i \cap s_j) \neq \emptyset$ y existe un homeomorfismo² $f : \mathbb{R}^2 \mapsto \mathbb{R}^2$ que mapea $\phi_{V_i^2}(|r|)$ a $\phi_{V_j^2}(|r|)$ (ver fig. 4.9).

Definición 4.3 (Registro lineal por trozos) Dadas dos realizaciones geo-

²Dados dos espacios topológicos X e Y , $f : X \mapsto Y$ es un homeomorfismo, sí y sólo sí, f es biyectiva y continua; y f^{-1} es continua.

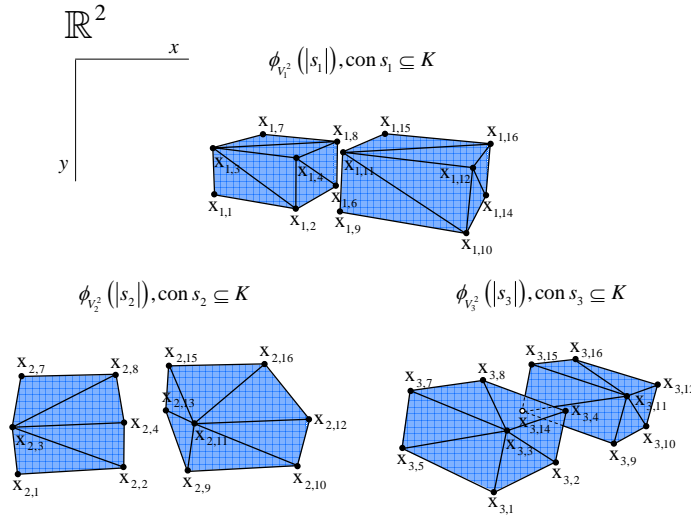


Figura 4.8: Redes triangulares en \mathbb{R}^2 que aproximan las proyecciones de la escena de la figura 4.6 sobre los planos imagen de las cámaras.

métricas $\phi_{V_i^2}(|s|)$ y $\phi_{V_j^2}(|s|)$, se define la transformación lineal por trozos $\mathbf{f} : \mathbb{R}^2 \mapsto \mathbb{R}^2$ que mapea un punto $\mathbf{x}_i = (x_i, y_i)^\top \in \phi_{V_i^2}(|s|)$ a otro $\mathbf{x}_j = (x_j, y_j)^\top \in \phi_{V_j^2}(|s|)$ como:

$$\mathbf{x}_j = \mathbf{f}_s(\mathbf{x}_i) = \begin{cases} f_{\sigma^0}(\mathbf{x}_i) & \text{si } \mathbf{x}_i \in \phi_{V_i^2}(|\sigma^0|), \forall \sigma^0 \in s \\ f_{\sigma^1}(\mathbf{x}_i) & \text{si } \mathbf{x}_i \in \phi_{V_i^2}(|\sigma^1|), \forall \sigma^1 \in s \\ f_{\sigma^2}(\mathbf{x}_i) & \text{si } \mathbf{x}_i \in \phi_{V_i^2}(|\sigma^2|), \forall \sigma^2 \in s \end{cases} \quad (4.1)$$

donde f_{σ^k} son transformaciones lineales entre símplices estimadas a partir de las realizaciones geométricas de sus vértices: 1 para σ^0 , 2 para las aristas (σ^1) y 3 para los triángulos (σ^2). Matemáticamente:

$$\mathbf{x}_j = f_{\sigma^0}(\mathbf{x}_i) \equiv \begin{cases} x_j = x_i + t_x \\ y_j = y_i + t_y \end{cases}$$

$$\mathbf{x}_j = f_{\sigma^1}(\mathbf{x}_i) \equiv \begin{cases} x_j = a_{11} x_i - a_{12} y_i + t_x \\ y_j = a_{12} x_i + a_{11} y_i + t_y \end{cases}$$

$$x_j = f_{\sigma^2}(x_i) \equiv \begin{cases} x_j = a_{11} x_i + a_{12} y_i + t_x \\ y_j = a_{21} x_i + a_{22} y_i + t_y \end{cases}$$

esto es, una traslación, una similaridad y una afinidad, respectivamente (ver sec. 3.1).

Recuérdese que σ^0 , σ^1 y σ^2 son símlices de orden 0, 1 y 2, esto es, vértices, aristas y triángulos, respectivamente (ver apéndice C).

Definición 4.4 Se definen las cámaras afines $P_i : \mathbb{R}^3 \mapsto \mathbb{R}^2$, con $i = 1, \dots, c$, que proyectan la realización geométrica de s_i en \mathbb{R}^3 , $\phi_{V^3}(|s_i|)$, sobre los planos imagen π_i , $\phi_{V_i^2}(|s_i|)$.

Se dice que $\phi_{V^3}(|r|)$ se puede reconstruir a partir de dos proyecciones $\phi_{V_i^2}(|r|)$ y $\phi_{V_j^2}(|r|)$, sí y sólo sí, $r = (s_i \cap s_j) \neq \emptyset$ y las cámaras afines P_i y P_j son ambas homeomorfismos (ver fig. 4.9).

Definición 4.5 (Reconstrucción lineal por trozos) Dadas dos realizaciones geométricas $\phi_{V_i^2}(|s|)$ y $\phi_{V_j^2}(|s|)$, se define la reconstrucción que aproxima la superficie lineal por trozos proyectada sobre los planos imagen π_i e π_j como la realización geométrica $\phi_{V^3}(|s|)$ que satisface la expresión:

$$P_i^{-1}(\phi_{V_i^2}(|s|)) - P_j^{-1}(\phi_{V_j^2}(|s|)) = 0 \quad (4.2)$$

Estrictamente hablando, sólo se pueden registrar aquellas regiones de la imagen (realizaciones geométricas) cuya estructura topológica (complejo simplicial) es común. Análogamente, sólo se puede obtener una reconstrucción a partir de dos realizaciones geométricas con la misma estructura topológica. Así, en el ejemplo de la figura 4.8, las realizaciones geométricas de los conjuntos simpliciales $\{\{2, 6, 4\}, \{4, 6, 8\}\}$, sólo captado por la cámara 1, y $\{\{11, 15, 13\}, \{9, 11, 13\}\}$, sólo captado por la cámara 2 ni se pueden registrar ni se puede generar una reconstrucción a partir de ellos.

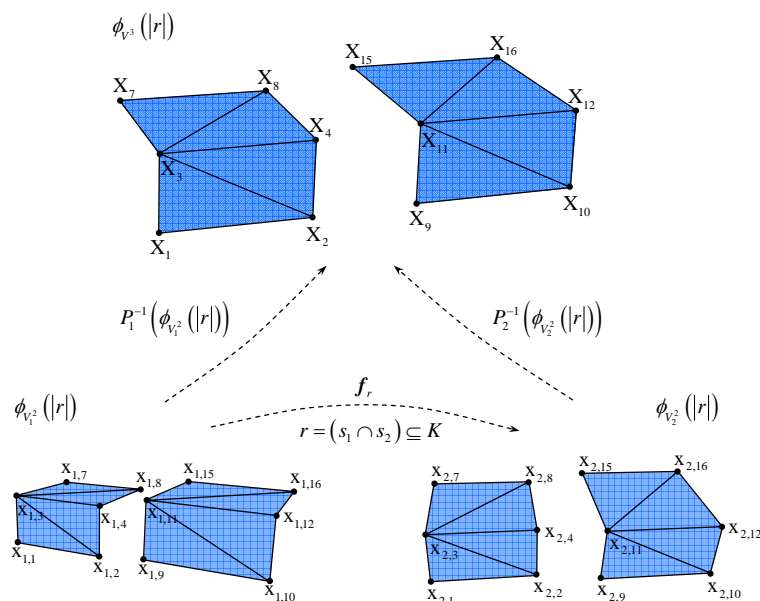


Figura 4.9: Registro lineal por trozos de dos realizaciones geométricas con la misma estructura topológica. Reconstrucción 3D que aproxima la superficie lineal por trozos proyectada sobre ambas imágenes.

4.4 Generación de las redes

En la práctica, rara vez se dispone de información *a priori* sobre la escena y, consecuentemente, no se puede saber cual es la red triangular que mejor aproxima la geometría de ésta a efecto de llevar a cabo un registro óptimo de las imágenes capturadas. Así pues, la generación de las redes triangulares debe abordarse en el dominio de la imagen con las limitaciones que esto conlleva.

Para generar las redes $M_i^2 = (K, V_i^2)$ y $M_j^2 = (K, V_j^2)$ utilizadas en el registro lineal por trozos de dos imágenes I_i e I_j de la misma escena, se suele proceder del siguiente modo:

1. se selecciona un conjunto de correspondencias en ambas imágenes $\{(x_{i,k}, x_{j,k}), k = 1, \dots, n | x_{i,k} \in V_i^2, x_{j,k} \in V_j^2\}$ (la selección robusta de correspondencias se aborda en detalle en la sección 4.5);
2. se genera la red $M_i^2 = (K, V_i^2)$, mediante la triangulación (utilizando Delaunay) del conjunto V_i^2 de puntos identificados en la imagen de

referencia; finalmente,

3. se toma el complejo simplicial K de M_i^2 y se genera la red $M_j^2 = (K, V_j^2)$, mediante la sustitución de los puntos en V_i^2 por sus correspondientes en V_j^2 .

Aunque la realización geométrica $\phi_{V_i^2}$ generada mediante el procedimiento de triangulación siempre es un *encaje*³, su correspondiente $\phi_{V_j^2}$ puede no serlo, lo que se traduce en la aparición de ciertas inconsistencias denominadas *patch reversals*. Los *patch reversals* se producen típicamente cuando la escena es observada desde puntos muy diferentes (ver fig. 4.10(a)), y tienen graves implicaciones:

1. $\phi_{V_i^2}(|K|)$ y $\phi_{V_j^2}(|K|)$ no se pueden registrar, ya que, la transformación f no es un homeomorfismo, y
2. $\phi_{V_j^2}(|K|)$ no se puede reconstruir, ya que, aunque P_i es un homeomorfismo, P_j no lo es, luego no se verifica la expresión (4.2).

Para eliminar estas inconsistencias, se revisa la topología inicial aplicando las siguientes modificaciones:

1. si la arista que produce el *patch reversal* es frontera, se elimina el triángulo que la contiene, como se muestra en la figura 4.10(b).
2. si la arista que lo produce es compartida, se intercambia la arista, como se ilustra en la figura 4.10(c).

Este proceso se repite iterativamente hasta que se eliminan la totalidad de las inconsistencias.

El procedimiento de generación de redes triangulares descrito tiene dos inconvenientes (ver fig. 4.11):

1. no garantiza que los triángulos conjugados caigan sobre proyecciones de superficies planas. Por ejemplo, los triángulos $\{1, 7, 3\}$ y $\{10, 12, 15\}$ de K .

³Dados dos espacios topológicos X e Y , $f : X \mapsto Y$ es un encaje (*embedding*), sí y sólo sí, f es un homeomorfismo entre X y su imagen $f(X)$.

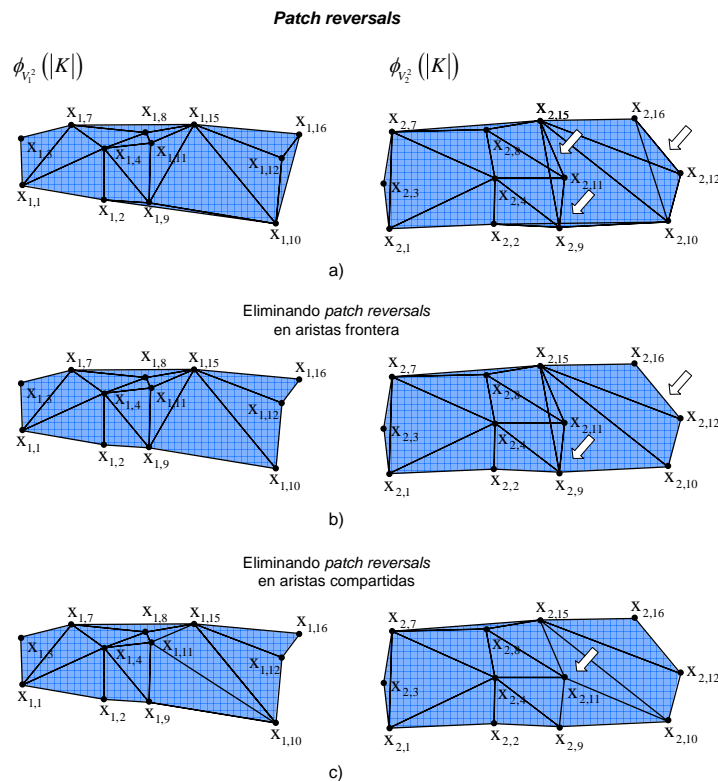


Figura 4.10: (a) *Patch reversals* (PR) producidos por la existencia de caras ocultas cuando la escena (descrita con una red triangular) es observada desde posiciones diferentes. Para eliminar estas inconsistencias se proponen los siguientes cambios topológicos en la redes iniciales: (b) Si la arista es frontera: se elimina el triángulo que la contiene, esto es, $\{10, 12, 16\}$, $\{2, 9, 10\}$. (c) Si la arista es compartida: se intercambia, esto es, $\{9, 15\}$ por $\{10, 11\}$.

2. generan estructuras topológicas regulares 1-conectadas, esto es, todos los símlices máximos de K son triángulos y todos los triángulos tienen al menos un adyacente, dando lugar a nuevos triángulos sin correspondencia real con la escena observada. Por ejemplo, los triángulos $\{2, 9, 4\}$ y $\{8, 11, 15\}$ de K .

En relación a la optimización de la red, en esta tesis se propone maximizar el número de triángulos que caen sobre proyecciones de superficies planas de la imagen, para lo cual se modifica la topología y/o geometría de las redes $M_i^2 = (K, V_i^2)$ y $M_j^2 = (K, V_j^2)$. Las acciones implementadas no

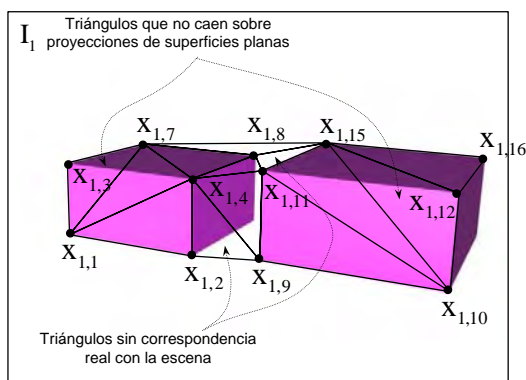


Figura 4.11: Triángulos no consistentes con las proyecciones de la escena observada: (a) triángulos que no caen sobre proyecciones de superficies planas y (b) triángulos sin correspondencia real con la escena observada.

contemplan la posibilidad de “desconectar” triángulos para generar símplices de dimensión menor (aristas y puntos), así como tampoco, eliminar triángulos sin correspondencia real con la escena. Este tipo de acciones se dejan como trabajo futuro.

4.5 Selección del conjunto de correspondencias

Son muchas las aplicaciones de visión por computador que requieren de técnicas robustas de búsqueda de correspondencias para extraer información relevante de dos o más imágenes, ya sea para registrarlas con precisión, para recuperar la disposición relativa del sensor o sensores en el momento de la captura, para reconstruir tridimensionalmente una escena, etc.

Una *característica* es un elemento distintivo o representativo detectado manual, o preferiblemente, de forma automática en la imagen, por ejemplo: regiones, bordes, contornos, segmentos, intersecciones de rectas, esquinas, etc. Las características poseen propiedades que las describen (esto es, descriptores), en forma vectorial $d = (d_1, \dots, d_m)^T \in \mathbb{R}^m$, y vectores de coordenadas, $x = (x, y)^T \in \mathbb{R}^2$, que las localizan en la imagen.

Una *correspondencia* (o características conjugadas) es el conjunto formado por los vectores de coordenadas que localizan una misma característica

en las imágenes que intervienen en el análisis. Así, dadas dos imágenes I_i e I_j , la correspondencia k -ésima seleccionada en ambas imágenes vendría dada por la tupla $(x_{i,k}, x_{j,k})$.

Esta sección se centra exclusivamente en las técnicas de detección y emparejamiento robusto de puntos de interés (esquinas). Para una revisión exhaustiva de los diferentes métodos propuestos en la literatura, el lector puede dirigirse a los siguientes trabajos [20, 42, 113].

4.5.1 Detección y emparejamiento de esquinas

Una *esquina* es un punto o localización de la imagen que tiene, para una escala determinada, grandes gradientes en todas las direcciones. La búsqueda de esquinas ha recibido, y continúa recibiendo, gran atención en la literatura. A continuación se revisan brevemente algunos de los trabajos más representativos.

En 1982, Kitchen y Rosenfeld [58] propusieron explotar las derivadas parciales de segundo orden para detectar esquinas. A pesar de los prometedores resultados, este detector de esquinas demostró ser excesivamente sensible al ruido. Con objeto de minimizar esta limitación, Förstner y Gulch proponen en [35] un detector sensiblemente más robusto basado en la primera derivada, aunque computacionalmente muy costoso. En 1980, Moravec presenta un método basado en las derivadas parciales de primer orden e introduce el concepto de la matriz de *auto-correlación*. En 1988, Harris y Stephens extienden el trabajo de Moravec en [46], proponiendo una función detección de esquinas muy eficiente. En [94], Shi y Tomasi introducen una nueva función de detección que explota igualmente la matriz de auto-correlación para extraer esquinas fáciles de seguir en secuencias de imágenes. Merece especial atención el trabajo Lowe [63], donde se propone una novedosa técnica para detección de esquinas basada en la búsqueda de extremos (máximos y mínimos locales) en un *espacio de escalas* construido a partir de diferencias de Gaussinas (DoG). Este enfoque produce, a diferencia de otros detectores como los de Harris o Shi y Tomasi, esquinas en diferentes escalas.

Descriptores de esquinas

Los vectores de propiedades (o descriptores) describen las esquinas o, más formalmente, el entorno de éstas con un único objetivo: lograr que dicho vector sea lo suficientemente distintivo como poder diferenciarlo inequívocamente del resto de esquinas.

El procedimiento más sencillo consiste en utilizar los niveles de gris de un entorno de la esquina como descriptor. Esta representación, aunque muy utilizada tiene sus limitaciones: es sensible a los cambios en el brillo y/o contraste, la presencia de ruido y las distorsiones geométricas. A esto hay que sumar problemas asociados específicamente al tamaño del descriptor, un entorno grande es mucho más distintivo pero más sensible a los problemas descritos con anterioridad, y viceversa.

Aunque en la literatura se pueden encontrar técnicas que tratan de compensar estas limitaciones mediante procedimientos de normalización radiométrica, o que consideran que los vecindarios pueden verse afectados por un conjunto limitado de posibles diferencias geométricas, la utilización de este tipo de técnicas suele restringirse a aplicaciones en los que las diferencias entre imágenes no son excesivamente importantes, como por ejemplo, para el seguimiento de características en secuencias de imágenes [65].

Numerosos autores han tratado de suplir estas limitaciones dotando a las esquinas extraídas con detectores tradicionales como los de Harris, Shi y Tomasi o Lowe, con descriptores adicionales. Este es el caso de Schmid y Mohr [90], que emplea una versión extendida del detector de Harris para la búsqueda de contenidos en imágenes. Otros autores como Baumberg [16], Mikolajczyk y Schmid [71], Schaffalitzky y Zisserman [88] o Brown y Lowe [21], con su famoso SIFT, mejoran la estabilidad de los descriptores al dotarlos de invarianza a un mayor número de diferencias geométricas, incluidas las transformaciones afines.

En la literatura también se pueden encontrar variantes de estos procedimientos, como el denominado PCA-SIFT, propuesto por Ke y Sukthankar [56], que realiza un análisis PCA para generar descriptores equivalentes a los SIFT pero de menor dimensión, o técnicas relativamente recientes como

el SURF [17], un descriptor invariante a la rotación y la escala, basado en el uso de imágenes integrales [100] y en la matriz Hessiana de la imagen. En [74], el lector puede encontrar una comparativa de distintos detectores de esquinas y descriptores evaluada sobre imágenes de objetos tridimensionales.

Emparejamiento de esquinas

Una vez localizadas las esquinas en ambas imágenes se procede a la generación de correspondencias mediante su emparejamiento. En general, dada una esquina localizada en una imagen, su emparejamiento consiste en localizar, en el conjunto de esquinas detectadas en la otra imagen, aquella con el descriptor más “parecido”, esto es, el par de esquinas conjugadas cuyos descriptores asociados minimizan, por ejemplo, la distancia Euclídea.

Ésta búsqueda se puede abordar de forma exhaustiva (esto es, todos con todos) o restringirla al conjunto de esquinas detectadas en un determinado área de la segunda imagen, con el consiguiente ahorro computacional. Por ejemplo, en [110], Xu and Zhang restringen la búsqueda de correspondencias dentro de un determinado área denominada ventana de búsqueda, otros explotan la restricción epipolar, para restringir la comparación a aquellas esquinas que caen sobre sus correspondientes líneas epipolares [75, 76]. Naturalmente, este enfoque requiere un conocimiento *a priori* de las posibles diferencias geométricas que presentan las imágenes o del desplazamiento relativo de los sensores en el momento de la adquisición (por ejemplo, en estéreo).

4.5.2 Selección robusta de correspondencias

En la práctica, el empleo de una u otra técnica de detección de esquinas dependerá fundamentalmente de las diferencias geométricas que presenten las imágenes que intervienen en el análisis y del tipo de aplicación. Así, el detector de Harris es sensible a los cambios de escala, pero tiene un coste computacional pequeño, lo que lo hace especialmente útil en aplicaciones que requieran de rapidez de computo y las imágenes no presenten importantes cambios de escala. Por otro lado, el detector de Lowe tiene un elevado

coste computacional (especialmente, la etapa de búsqueda de extremos) pero proporciona características altamente estables a los cambios de escala y otras deformaciones geométricas afines, lo que lo hace especialmente apropiado en aplicaciones en las que las imágenes han sido adquiridas desde puntos de vista muy dispares, y que presentan obviamente, importantes diferencias geométricas.

En esta tesis, al igual que otros trabajos encontrados en la literatura [32, 76], se explotan los beneficios de ambas técnicas en pos de obtener un mejor rendimiento computacional y aumentar el porcentaje de correspondencias válidas en la etapa de emparejamiento. Más concretamente, el proceso de búsqueda y emparejamiento se aborda del siguiente modo:

1. se localizan las esquinas con el detector Harris⁴,
2. se calculan sus correspondientes descriptores SIFT, y finalmente,
3. se establecen los emparejamientos utilizando la distancia Euclídea.

Además, para dotar de robustez al procedimiento, se recupera la geometría epipolar afín intrínseca a ambas imágenes utilizando para su estimación el conjunto de correspondencias establecidas anteriormente. Esto proporciona tres beneficios: 1) detectar los posibles pares espurios (*outliers*), es decir, pares no consistentes con la restricción epipolar; 2) refinar la localización de los pares válidos (*inliers*) y 3) obtener una estimación robusta de la matriz fundamental afín, F_A , que será utilizada en el siguiente capítulo para dirigir el proceso de división de aristas en la red triangular.

Para estimar la matriz fundamental afín se emplea un procedimiento basado en el algoritmo RANdom SAmple Consensus (RANSAC) [31]. Esta técnica explota la redundancia de información de un conjunto de muestras para proporcionar una estimación robusta del modelo que se ajusta con precisión a la mayoría de ellas. En el apéndice B se discute en detalle la estimación robusta de parámetros y se describen éste y otros métodos propuestos en la literatura.

⁴El lector puede encontrar en el apéndice D una descripción detallada de los detectores de Harris y Lowe.

En este caso particular, el modelo es la matriz fundamental afín, F_A , y la función de error que mide la consistencia de una correspondencia dada, (x_i, x_j) , es el error geométrico de primer orden (o distancia de Sampson), esto es, la distancia de una esquina a su correspondiente línea epipolar, que se expresa como:

$$d(x_j, F_A x_i) = \frac{(x_j^\top F_A x_i)^2}{(F_A x_i)_1^2 + (F_A x_i)_2^2 + (F_A^\top x_j)_1^2 + (F_A^\top x_j)_2^2} \quad (4.3)$$

donde los términos de la forma $(v)_k^2$ representan el cuadrado de la coordenada k -ésima del vector v .

Para ganar en precisión, en su lugar, el procedimiento utilizado en esta tesis emplea el error epipolar simétrico, que considera la distancia del par (x_i, x_j) a sus respectivas líneas epipolares. El error epipolar simétrico se deriva de (4.3) y se define como:

$$\begin{aligned} d(x_j, F_A x_i)^2 + d(x_i, F_A^\top x_j)^2 &= \\ &= \left(\frac{(x_j^\top F_A x_i)^2}{(F_A x_i)_1^2 + (F_A x_i)_2^2} + \frac{(x_i^\top F_A^\top x_j)^2}{(F_A^\top x_j)_1^2 + (F_A^\top x_j)_2^2} \right) \end{aligned} \quad (4.4)$$

El paso final de algoritmo RANSAC es la re-estimación del modelo considerando sólo las correspondencias válidas (*inliers*). Esta re-estimación se realiza mediante un proceso de minimización del que se deriva la estimación de máxima verosimilitud (MLE) de la matriz fundamental afín y una estimación de las coordenadas óptimas de las correspondencias. La MLE asume que los errores cometidos en la identificación de las esquinas siguen una distribución Gaussiana. En tal caso, la MLE de la F_A es aquella que minimiza

el error de re-proyección. Formalmente, se formula como:

$$\min_{\{F_A, \hat{x}_{i,k}, \hat{x}_{j,k}\}} \sum_{k=1}^n d(x_{i,k}, \hat{x}_{i,k})^2 + d(x_{j,k}, \hat{x}_{j,k})^2 \quad (4.5)$$

donde $(x_{i,k}, x_{j,k})$ son las correspondencias iniciales y $(\hat{x}_{i,k}, \hat{x}_{j,k})$ las refinadas, siendo n el número de correspondencias válidas.

El lector puede encontrar en el texto de Hartley y Zisserman [47] (pag. 349) el desarrollo de teórico de este problema de minimización.

La figura 4.12 ilustra el proceso de selección robusta de correspondencias descrito en esta sección.

4.6 Conclusiones

En este capítulo se ilustran las ventajas de emplear funciones lineales por trozos para el registro de imágenes conjugadas de una escena poliédrica. Este procedimiento divide las imágenes en regiones triangulares conjugadas que se registran individualmente mediante afinidades. Sin embargo, para que este método trabaje adecuadamente, las regiones triangulares deben caer sobre proyecciones de superficies planas de la escena, aspecto que no abordan las diferentes implementaciones de esta técnica que se pueden encontrar en el software disponible en el mercado.

En este capítulo se ha analizado esta problemática y se propone una formalización del registro lineal por trozos basado en complejos simpliciales que abarca, no sólo la generación de redes triangulares 2D óptimas (desde el punto de vista de la consistencia del registro), sino también la reconstrucción a partir éstas de la superficie 3D observada. También se ha propuesto un procedimiento automatizado para la generación de redes triangulares conjugadas libres de *patch reversals*. La descripción del procedimiento de optimización propuesto se aborda en el siguiente capítulo.

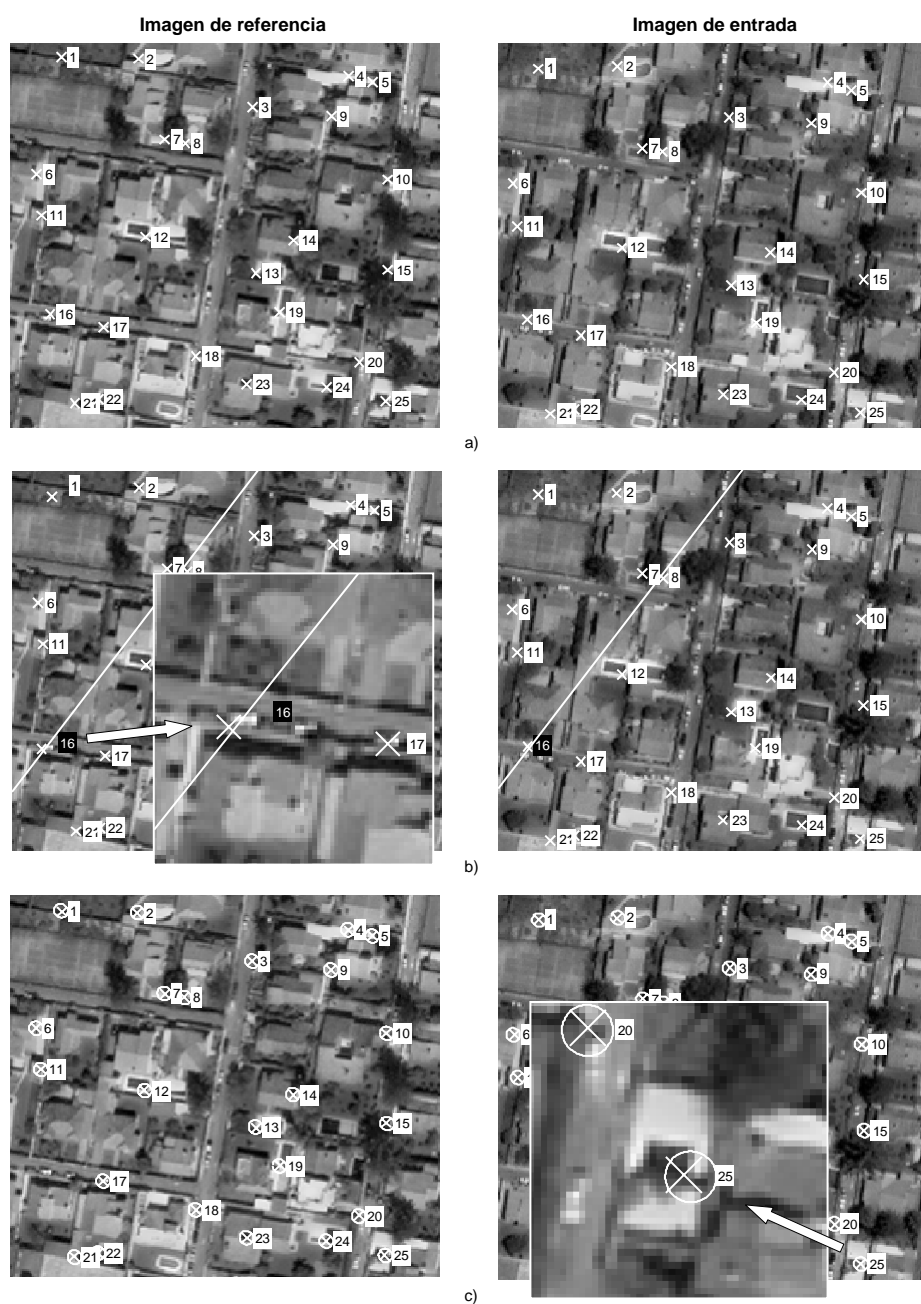


Figura 4.12: Selección robusta de correspondencias: a) detección y emparejamiento, b) detección de los *outliers* y c) refinamiento de las coordenadas de los *inliers* (los círculos denotan las coordenadas refinadas). Los fragmentos ampliados de las imágenes muestran, en detalle, las actuaciones (b) y (c) sobre las correspondencias 16 y 25, respectivamente.

Capítulo 5

Optimización de redes triangulares

En este capítulo se propone un procedimiento de optimización de redes triangulares para mejorar la precisión del registro lineal por trozos. Este procedimiento contempla un conjunto de acciones que modifican topológica y geoméricamente la red inicial, utilizando una función de energía o coste basada en la información mutua (MI) que cuantifica la mejora obtenida con cada acción.

Más concretamente, se propone un método que actúa sobre la red, eliminando, intercambiando y dividiendo aristas (ver fig. 4.3). Este proceso se formula como una búsqueda *greedy* [25] que, en cada iteración, toma una arista concreta, verifica el número y tipo de las acciones aplicables, y comprueba si éstas producen una mejora de la consistencia. Aunque la búsqueda *greedy* no garantiza que se alcance un máximo global, su simplicidad y eficiencia lo hacen, como se verá más adelante, especialmente adecuado para la aplicación que nos ocupa.

El método propuesto se ha probado satisfactoriamente con diferentes conjuntos de imágenes, sintéticas y reales, adquiridas desde distintos puntos de observación y condiciones de iluminación. En los test se ha evaluado la mejora en la precisión global del registro en comparación con otros métodos, así como la exactitud de las reconstrucciones 3D derivadas de las redes optimizadas.

5.1 Definiciones

En esta sección se definen diversos elementos de la red, así como las posibles acciones topológicas sobre ella. Es conveniente que el lector esté familiarizado con términos tales como complejos simpliciales, cierre (*closure*) y estrella (*star*) de un simplex, etc. para lo cual puede dirigirse al apéndice C.

5.1.1 Elementos de la red

Dada una red triangular $M = (K, V)$ y una arista compartida $e = \{i, j\} \in K$, se introducen las siguientes definiciones (ver fig. 5.1):

Definición 5.1 (quad) El *cuadrilátero* de la arista e en K es el complejo simplicial que se obtiene de unir sus 2-símplices padre (esto es, los triángulos que la comparten, de ahí que se denomine *arista compartida*) y todas las caras de estos:

$$\text{quad}(e, K) = \text{closure}(\text{star}(e, K), K) \quad (5.1)$$

donde $\text{star}(e, K)$ es el conjunto de símplexes de K de los que e es cara y $\text{closure}(s, K)$ es el complejo simplicial más pequeño de K que contiene todos los símplexes de e .

Si $e = \{i, j\}$, entonces el $\text{quad}(e, K)$ está formado por los siguientes símplexes:

$$\begin{cases} 0\text{-símplexes } (\sigma^0) = \{\{i\}, \{j\}, \{k\}, \{l\}\} \\ 1\text{-símplexes } (\sigma^1) = \{\{i, j\}, \{i, k\}, \{k, j\}, \{j, l\}, \{l, i\}\} \\ 2\text{-símplexes } (\sigma^2) = \{\{i, k, l\}, \{j, l, i\}\} \end{cases}$$

Los 0-símplexes $\{k\}, \{l\} \in K$ se denominan *vértices opuestos* de e en K y la hipotética arista $\bar{e} = \{k, l\} \notin K$ que se obtendría al conectarlos, *arista opuesta* de e .

Definición 5.2 (bound) El *contorno* de la arista e en K es el conjunto de

d -símplices, con $d = 0, 1$, (esto es, vértices y aristas) que forman la envoltura convexa del quad (e, K) :

$$\text{bound}(e, K) = \text{quad}(e, K) \cap \text{star}(e, K) \quad (5.2)$$

Si $e = \{i, j\}$ entonces, el $\text{bound}(e, K)$ está formado por los símplices:

$$\begin{cases} 0\text{-símplices } (\sigma^0) = \{\{i\}, \{j\}, \{k\}, \{l\}\} \\ 1\text{-símplices } (\sigma^1) = \{\{i, k\}, \{k, j\}, \{j, l\}, \{l, i\}\} \end{cases}$$

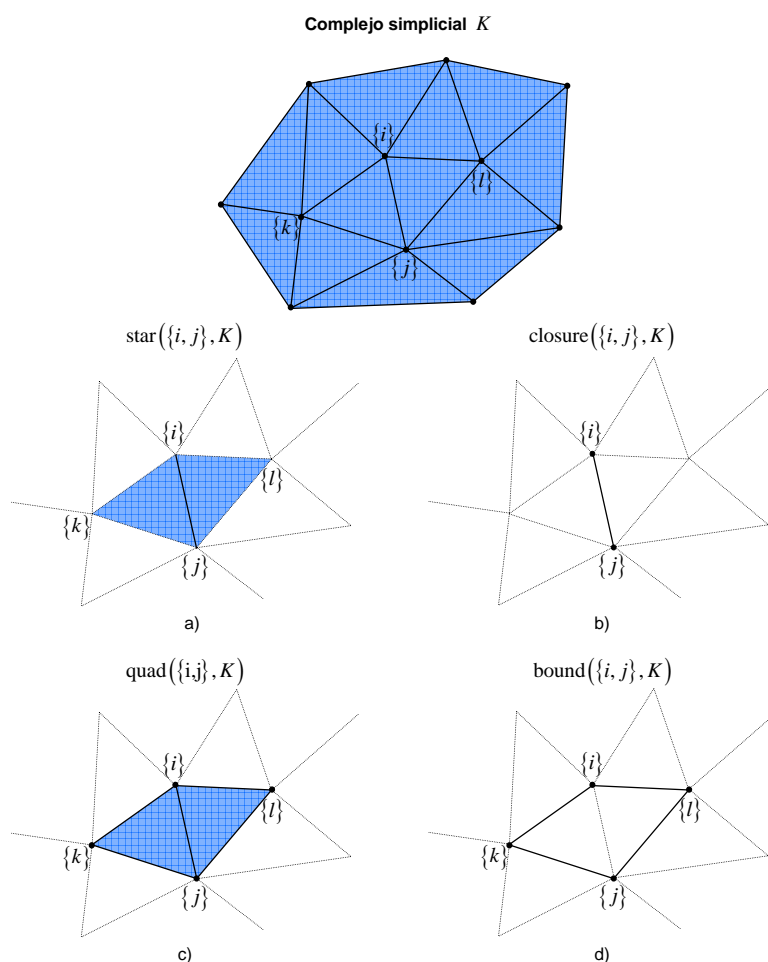


Figura 5.1: Conjuntos de símplices característicos: (a) star, (b) closure, (c) quad y (d) bound.

5.1.2 Acciones topológicas

En esta sección se definen las siguientes acciones topológicas de interés para el método de optimización propuesto:

Definición 5.3 (swap) Dada la arista $e = \{i, j\}$ del complejo simplicial K de la figura 5.1, se define la acción *intercambiar* (swap) *arista*

$$\hat{K} = \text{swap}(e, K) \quad (5.3)$$

como el conjunto de operaciones sobre K (ver fig. 5.2):

- eliminar el conjunto de símplexes $\text{star}(e, K)$,
- insertar la arista opuesta $\bar{e} = \{k, l\}$, e
- insertar los triángulos $\{i, k, l\}$ y $\{j, l, k\}$.

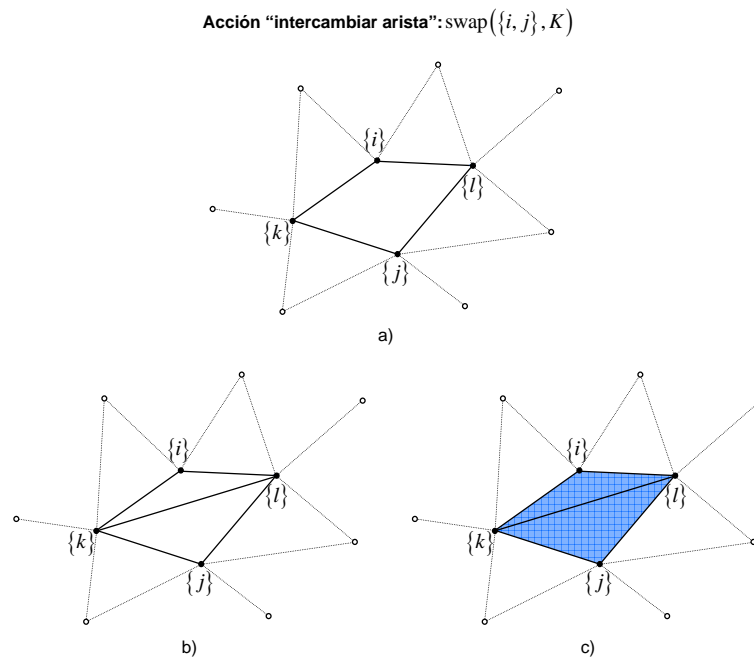


Figura 5.2: Pasos de la acción $\text{swap}(e, K)$: (a) eliminar el conjunto de símplexes $\text{star}(e, K)$; (b) insertar la arista opuesta $\bar{e} = \{k, l\}$ y (c) insertar los triángulos $\{i, k, l\}$ y $\{j, l, k\}$.

Definición 5.4 (split) Dado un vértice $v \notin K$, se define la acción *dividir* (split) *arista*

$$\hat{K} = \text{split}(e, v, K) \quad (5.4)$$

como el conjunto de operaciones sobre K (ver fig. 5.3):

- eliminar el conjunto de símplexes $\text{star}(e, K)$,
- insertar el vértice v ,
- insertar una arista $\{p, v\}$ por cada vértice $\{p\} \in \text{bound}(e, K)$ e
- insertar un triángulo $\{p, q, v\}$ por cada arista $\{p, q\} \in \text{bound}(e, K)$.

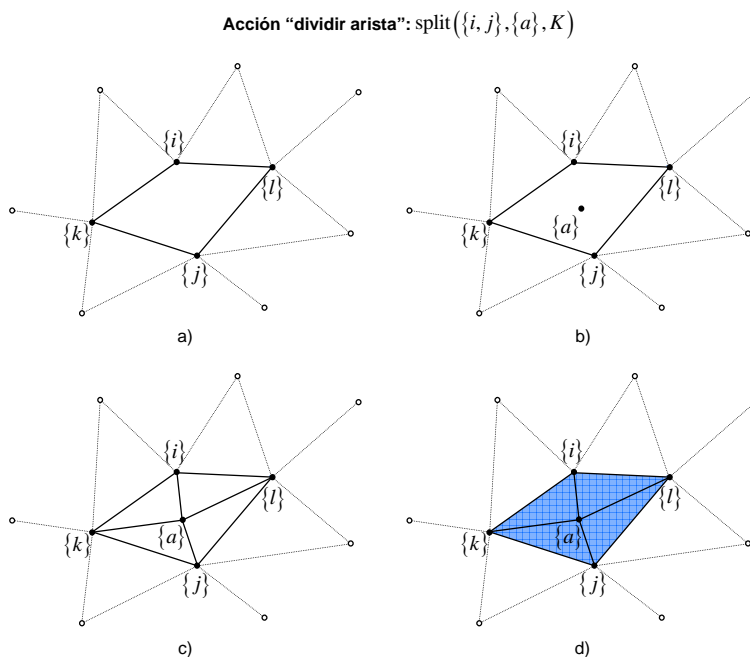


Figura 5.3: Pasos de la acción $\text{split}(e, \{a\}, K)$: (a) eliminar el conjunto de símplexes $\text{star}(e, K)$; (b) insertar el vértice $\{a\}$; (c) insertar una arista $\{p, a\}$ por cada vértice $\{p\} \in \text{bound}(e, K)$ y (d) insertar un triángulo $\{p, q, a\}$ por cada arista $\{p, q\} \in \text{bound}(e, K)$.

Definición 5.5 (collapse) Dado un vértice $v \in e$, se define la acción *eliminar (collapse) arista*

$$\hat{K} = \text{collapse}(e, v, K) \quad (5.5)$$

como el conjunto de operaciones sobre K (ver fig. 5.4):

- eliminar el conjunto de símplexes $\text{star}(\{i\}, K) \cup \text{star}(\{j\}, K)$,
- insertar el vértice v ,
- insertar una arista $\{p, v\}$ por cada vértice $\{p\} \in \text{bound}(e, K)$ e
- insertar un triángulo $\{p, q, v\}$ por cada arista $\{p, q\} \in \text{bound}(e, K)$.

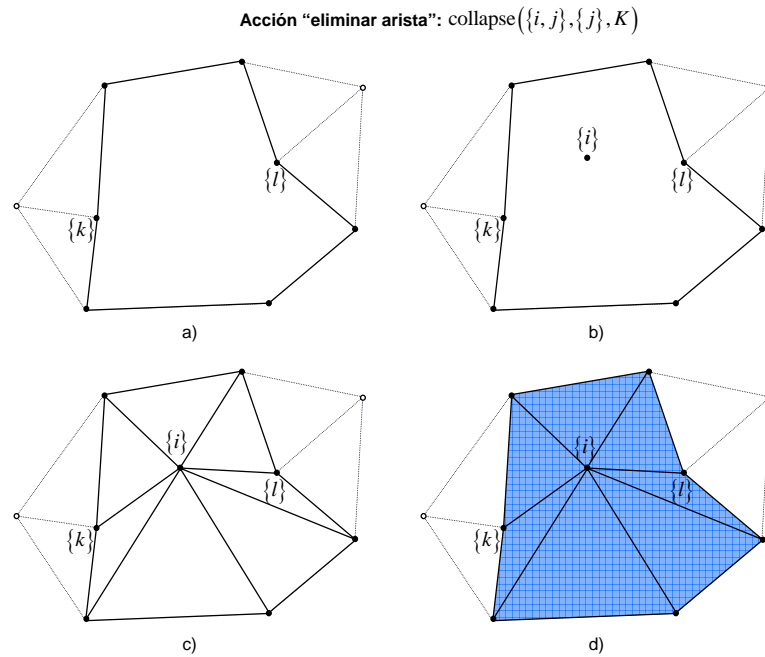


Figura 5.4: Pasos de la acción $\text{collapse}(e, \{j\}, K)$: (a) eliminar $\text{star}(\{i\}, K) \cup \text{star}(\{j\}, K)$; (b) insertar el vértice $\{i\}$; (c) insertar una arista $\{p, i\}$ por cada vértice $\{p\} \in \text{bound}(e, K)$ y (d) insertar un triángulo $\{p, q, i\}$ por cada arista $\{p, q\} \in \text{bound}(e, K)$.

Nótese que para la arista e se pueden generar dos acciones, una por cada vértice de e . Se podría haber optado, sin embargo, por una única acción que

eliminase simultáneamente los dos vértices, lo que, aunque con menor coste computacional, puede obviar interesantes configuraciones de la red.

5.2 Aplicación de las acciones

Es importante subrayar que, al igual que ocurre en la generación automática de las redes conjugadas (ver sec. 4.4), la aplicación de una acción puede producir un *patch reversal*. En este caso, la función de transformación f ya no es un homeomorfismo y, por tanto, las realizaciones geométricas de K ni se pueden registrar ni reconstruir. Para evitar este tipo de circunstancias, antes de aplicar cualquier acción, conviene evaluar si ésta produce este tipo de inconsistencias.

Dadas dos redes $M_1^2 = (K, V_1^2)$ y $M_2^2 = (K, V_2^2)$, una arista compartida $e = \{i, j\} \in K$ y el complejo simplicial $s = \text{star}(e, K)$, se dice que:

1. el $\text{swap}(e, K)$ es aplicable, si y sólo si, la arista opuesta \bar{e} no produce un *patch reversal*;
2. el $\text{split}(e, v, K)$ es aplicable, si y sólo si, $\phi_{V_1^2}(|v|) \in \phi_{V_1^2}(|s|)$ y $\phi_{V_2^2}(|v|) \in \phi_{V_2^2}(|s|)$, es decir, la realización geométrica del nuevo vértice $\phi_{V_1^2}(|v|)$ está dentro de la región cuadrangular definida por la realización geométrica $\phi_{V_1^2}(|s|)$ (análogamente para $\phi_{V_2^2}(|v|)$);
3. el $\text{collapse}(e, v, K)$ es aplicable, si y sólo si, v no es frontera en K .

A diferencia de la acción *intercambiar arista*, donde la generación de un *patch reversal* impide su aplicación, las inconsistencias que se producen en $\phi_{V_1^2}(|\hat{K}|)$ o $\phi_{V_2^2}(|\hat{K}|)$ al eliminar aristas se pueden resolver de forma análoga a la descrita en la sección 4.4. La figura 5.5 ilustra este proceso.

Especial interés recibe la división de aristas. A diferencia de las acciones *intercambiar* y *eliminar arista* que se aplican directamente, la acción *dividir arista* necesita determinar el punto de división idóneo, de modo que, dependiendo de la localización de éste, la acción puede conllevar una mejora de la consistencia o nó. A esto hay que sumar su mayor coste computacional,

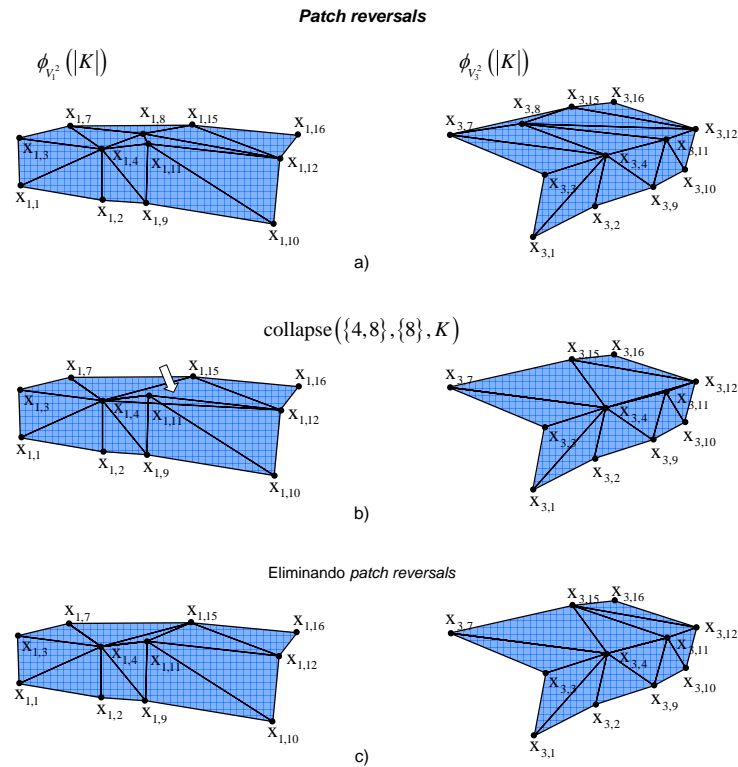


Figura 5.5: (a) *Patch reversal* (PR) producido por la eliminación de la arista $\{4,8\} \in K$. Para eliminar estas inconsistencias se procede forma análoga a lo descrito en la sección 4.4, esto es, se intercambia la arista compartida por los triángulos afectados, esto es, $\{4,12\}$ por su opuesta $\{11,15\}$.

ya que al no disponerse de información sobre la geometría de la escena, la división debe abordarse en el dominio de la imagen.

Por ejemplo, la figura 5.6 muestra una realización geométrica típica: diversas aristas que conectan puntos situados en proyecciones de superficies planas diferentes, por ejemplo la arista $e = \{4,10\}$. Observando la configuración de la red se aprecia claramente que si se intercambiara o eliminara e la consistencia no mejoraría mucho. Por el contrario, si se lograra dividir e en el borde del poliedro (ver flecha en la figura), las realizaciones geométricas de los nuevos triángulos, $\{9,10,p\}$ y $\{10,12,p\}$, caerían sobre proyecciones de superficies planas, obteniéndose una clara mejora de la consistencia en la región afectada y, por ende, en la consistencia global.

El procedimiento de división propuesto en este trabajo explota la infor-

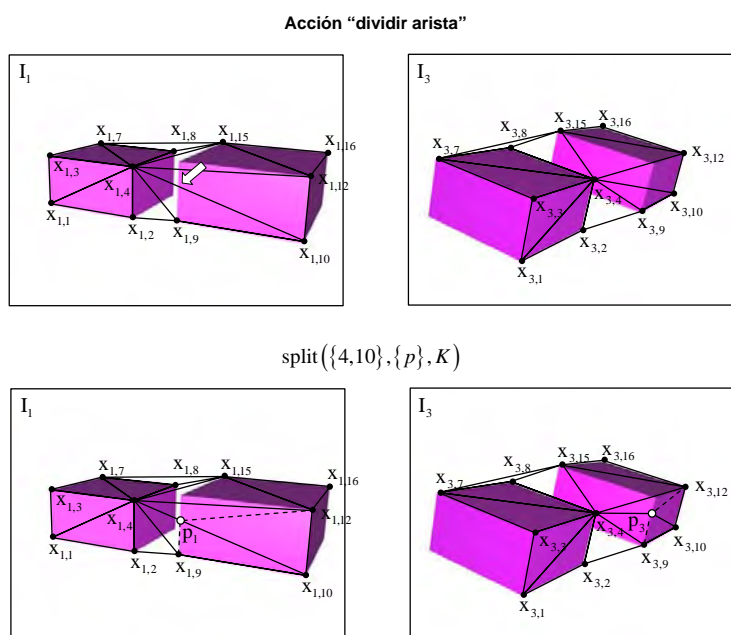


Figura 5.6: Inserción óptima de nuevos vértices en la red triangular. Supóngase, por ejemplo, que se desea dividir la arista $e = \{4, 10\}$. Una posición idónea para dividirla sería la intersección de ésta con el borde del poliedro, ya que las realizaciones geométricas de dos de los nuevos triángulos, en particular $\{9, 10, p\}$ y $\{10, 12, p\}$, caerían sobre una de las caras del poliedro, obteniéndose una clara mejora de la consistencia en la región de influencia de e .

mación de bordes de las imágenes (a partir de la cual se extraen segmentos) y la geometría epipolar afín intrínseca a ambas vistas para determinar la localización idónea del punto de división. En aras de una mayor claridad, la descripción del procedimiento se apoya en el caso ilustrado en las figuras 5.6 y 5.7.

Considérense las imágenes I_1 y I_3 , las redes $M_1^2 = (K, V_1^2)$ y $M_3^2 = (K, V_3^2)$ y la arista compartida $e = \{4, 10\} \in K$ de la figura 5.6. Para la división de e se procede del siguiente modo:

1. *Extracción de segmentos en ambas imágenes:* se extraen bordes en ambas imágenes y se aproximan segmentos. Los segmentos de longitud inferior a un umbral dado se descartan. En la implementación de esta etapa se pueden utilizar el detector de Canny [22] y el método de aproximación propuesto por Kovessi en [83].

2. *Extracción de los puntos conjugados de división* (ver fig. 5.7(b)): se determinan los puntos de intersección de $\phi_{V_1^2}(|e|)$ con los segmentos de I_1 :

$$\{p_{1,i} = (x_{1,i}, y_{1,i})^\top, i = 1, \dots, n\}$$

A continuación, se determinan los puntos de intersección de las líneas epipolares $l_{3,i} = F_A \times p_{1,i}$ con los segmentos de I_3 :

$$\{p_{3,j} = (x_{3,j}, y_{3,j})^\top, j = 1, \dots, m\}$$

descartándose los $p_{3,j} \notin \phi_{V_3^2}(|s|)$, con $s = \text{star}(e, K)$, es decir, los puntos que no están en el interior de la región cuadrangular definida por la realización geométrica de s . Para el resto se calculan sus descriptores SIFT y se emparejan utilizando la distancia Euclídea. Finalmente, los pares cuya distancia entre descriptores esté por encima de un umbral dado también se descartan.

El procedimiento descrito puede generar varias correspondencias por arista, en cuyo caso, se selecciona la que presenta una mayor disparidad, ya que se ha comprobado experimentalmente que es la correspondencia con mayor probabilidad de estar situada en el borde que “conecta” proyecciones de diferentes superficies (ver fig. 5.7(c)). Se pueden considerar otras estrategias de selección como, por ejemplo, generar tantas acciones de división como pares hayan sido localizados (con el consiguiente incremento en el coste computacional). Si no se obtiene ninguna correspondencia válida, la acción se marca como no aplicable.

La estimación de la matriz fundamental afín, la extracción de segmentos, así como la identificación de los puntos de división se realiza antes de iniciar el proceso de optimización. Una vez iniciado éste, cada vez que se aplica una acción se extraen puntos de división en las nuevas aristas y se eliminan los pares correspondientes a las aristas que ya no existen. El coste computacional de esta acción es superior a las acciones *eliminar* e *intercambiar arista* debido, fundamentalmente, al cálculo de los SIFTs. Para ganar en eficiencia,

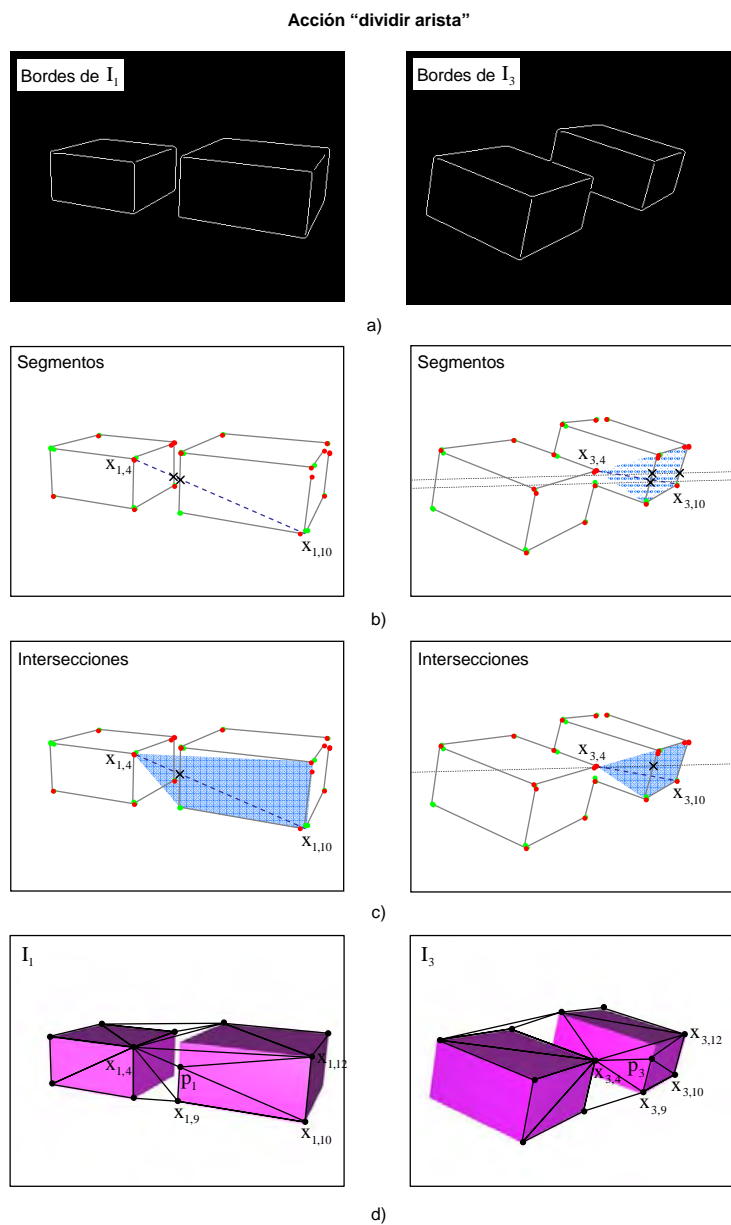


Figura 5.7: Proceso de inserción de nuevos vértices en las redes triangulares: (a) detección de bordes y extracción de segmentos, (b) extracción de puntos candidatos, (c) emparejamiento y selección de la “mejor” correspondencia (aquella con la disparidad máxima) y (d) configuración de la red una vez finalizada la división.

los descriptores SIFT se pueden obtener con la librería SiftGPU [95], una implementación que explota las prestaciones de las modernas unidades de procesamiento gráfico (GPU).

En ocasiones, la mejora obtenida al dividir una arista es idéntica¹ a la obtenida con una operación de intercambio (ver fig. 5.8), en cuyo caso, siempre se aplica la operación de intercambio, evitándose la inserción de vértices superfluos y que ralentizarían el proceso de optimización.

5.3 Descripción del método

El objetivo del método presentado en este capítulo es mejorar la precisión del registro lineal por trozos. Para este propósito se modifica iterativamente la topología/geometría de las redes triangulares (conjugadas) iniciales mediante la aplicación de distintas acciones, que se han descrito en la sección 5.1.2. Este proceso se formula como una búsqueda *greedy* [25] que, en cada iteración, toma una arista concreta, verifica el conjunto de acciones aplicables y mide la mejora en la consistencia de cada una ellas, seleccionando finalmente, la acción que produce el mayor incremento. Para cuantificar la mejora introducida se emplea una función de coste local que mide cuanto mejora el registro en la zona afectada, antes y después de aplicar cada acción.

5.3.1 La función de coste local

En este trabajo se explota la robustez de la MI, más concretamente del ECC, para cuantificar la mejora del registro introducida por una acción.

Dadas dos imágenes I_1 y I_2 ; dos redes $M_1^2 = (K, V_1^2)$ y $M_2^2 = (K, V_2^2)$; una arista compartida $e = \{i, j\} \in K$ y una acción act sobre e , se definen los siguientes elementos:

Definición 5.6 La *región de influencia* r de la acción $act(e, K)$ en M_1^2 (análogamente para M_2^2) es el conjunto de coordenadas $r \subset \mathbb{N}^2$ que se ven

¹Puede haber pequeñas diferencias debidas a la interpolación de los niveles de intensidad.

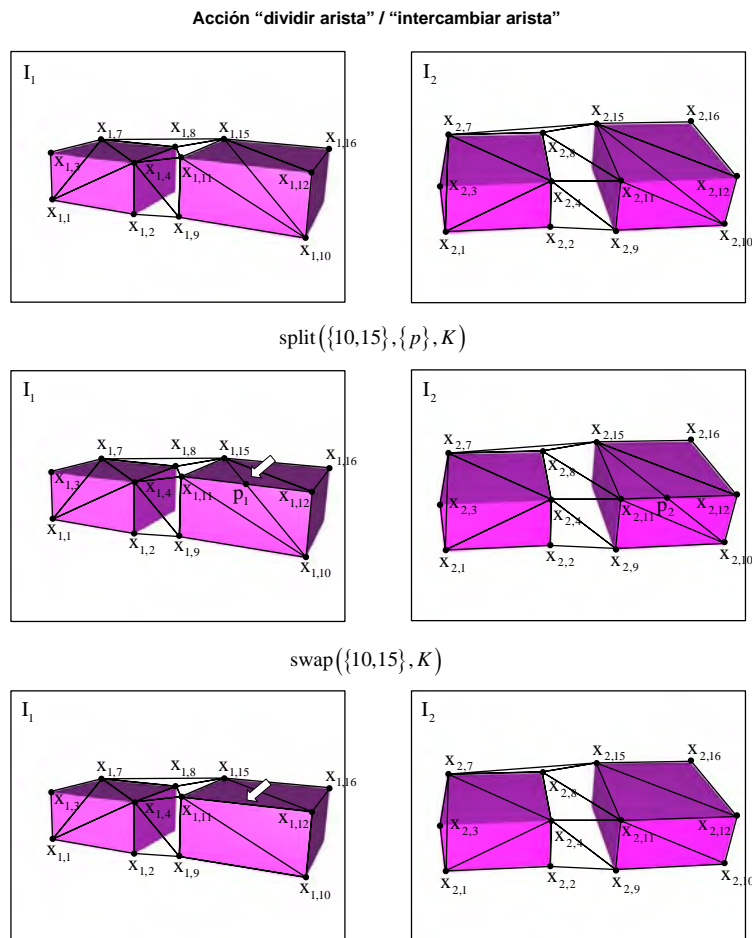


Figura 5.8: Comparación entre “intercambiar arista” y “dividir arista”. En ocasiones, una operación de división e intercambio proporcionan la misma mejora en la consistencia, obsérvese por ejemplo la arista $e = \{10, 15\}$, en cuyo caso la acción de división se marca como no aplicable.

afectadas por su aplicación, tal que $r \subset \phi_{V^2}(|s|)$. La región de influencia viene definida por el complejo simplicial s y depende del tipo de acción, en particular:

- $s = \text{star}(e, K)$, para las acciones swap y split, y
- $s = (\text{star}(\{i\}, K) \cup \text{star}(\{j\}, K))$, para la acción collapse.

Definición 5.7 La mejora $\Delta\omega$ introducida por la acción $act(e, K)$ es la variación en la consistencia del registro de su región de influencia antes (ω_1) y después (ω_2) de aplicarla, esto es:

$$\Delta\omega(e, act) = \omega_2(\hat{s}) - \omega_1(s) \quad (5.6)$$

donde

$$\begin{aligned} \omega_1(s) &= ECC(I_1(r), I_2(\mathbf{f}_s(r))), \\ \omega_2(\hat{s}) &= ECC(I_1(r), I_2(\mathbf{f}_{\hat{s}}(r))), \end{aligned} \quad (5.7)$$

$I_1(r)$ representa los píxeles de I_1 dados por la región de influencia r , $I_2(\mathbf{f}_s(r))$ y $I_2(\mathbf{f}_{\hat{s}}(r))$ los píxeles de I_2 dados por la región de influencia transformada de acuerdo con los dos posibles complejos s y $\hat{s} = act(e, s)$, respectivamente.

La acción se acepta, si y sólo si, $\Delta\omega(e, act) \geq \delta > 0$, esto es, cuando la acción da lugar a una mejora por encima de un umbral δ dado. Este umbral debe ser pequeño, pero suficiente, para impedir que se apliquen acciones cuyas regiones de influencia caen sobre proyecciones de superficies planas y que, debido a la interpolación de los niveles de intensidad, dan lugar a pequeñas variaciones positivas de ω . Se ha determinado experimentalmente que un $\delta \approx 0,01$ resuelve esta problemática.

Como se discutió en la sección 2.4, la estimación de la MI y de sus variantes normalizadas (como el ECC), depende de la estimación de la distribución de probabilidad conjunta de las intensidades de las imágenes. Puesto que la fiabilidad de esta estimación se ve comprometida cuando el número de píxeles es pequeño, aquellas acciones cuyas zonas de influencia tienen un tamaño (en

píxeles) inferior a un umbral dado son descartadas (en nuestros experimentos, 1000 píxeles). Por cuestiones de eficiencia, se ha utilizado el método de estimación paramétrico con 32 acumuladores.

5.3.2 Método de optimización

El método de optimización se formula como una búsqueda *greedy* [96], que comienza con dos redes triangulares $M_1^2 = (K, V_1^2)$ y $M_2^2 = (K, V_2^2)$ generadas como se detalla en la sección 4.4, y finaliza con sus correspondientes \hat{M}_1^2 y \hat{M}_2^2 optimizadas que maximizan la consistencia global del registro, formalmente:

$$\{\hat{K}, \hat{V}_1^2, \hat{V}_2^2\} = \arg \max_{\{K, V_1^2, V_2^2\}} \omega(K) \quad (5.8)$$

donde

$$\omega(K) = ECC(I_1(m), I_2(f_K(m))) \quad (5.9)$$

siendo $m \subset \phi_{V_1^2}(|K|)$ las coordenadas de los píxeles localizados en el interior de la red triangular dada por K .

La búsqueda *greedy*, detallada en el pseudocódigo de la figura 5.9, empieza creando una lista ordenada (en orden descendente) de $\Delta\omega$ indexada por (e, accion) . Generar esta lista conlleva un importante costo computacional, pero sólo se realiza una vez, al inicio del procedimiento. En cada iteración del proceso de búsqueda se toma el primer elemento de la tabla, esto es, la acción con el mayor $\Delta\omega$, y se aplica si $\Delta\omega > \delta$. Seguidamente se actualiza la lista ordenada con el conjunto de acciones aplicables sobre las aristas de la región de influencia de la acción y se evalúan sus correspondientes condiciones, así como la posibilidad de que su aplicación de lugar a la aparición una configuración de red previa (ciclo). El algoritmo finaliza cuando la lista de acciones aplicables está vacía o ninguna de las acciones aplicables produce una mejora por encima del δ dado.

Por último, para ilustrar el método de optimización propuesto se presenta un experimento consistente en registrar dos imágenes de una escena de inte-

OBJETIVO
 Dadas dos imágenes I_1 y I_2 a registrar y dos redes $M_1^2 = (K, V_1^2)$ y $M_2^2 = (K, V_2^2)$ sobre ellas, generadas como se detalla en la sección 4.4, determinar, mediante la optimización topológica y geométrica de las primeras, dos nuevas redes $\hat{M}_1^2 = (\hat{K}, \hat{V}_1^2)$ y $\hat{M}_2^2 = (\hat{K}, \hat{V}_2^2)$ que mejoran la consistencia del registro lineal por trozos.

ALGORITMO

1. % Construir una lista ordenada de $\Delta\omega$ para cada par {arista, acción}
 $\Delta\omega_list = []$
 FOR-EACH tupla $\{e, action\}$, $e = \{i, j\} \in K$
 IF $action(e, K)$ verifica las pre-condiciones THEN
 estimar $\Delta\omega(e, action)$ con la expresión (5.6)
 almacenar $\Delta\omega(e, action)$ en $\Delta\omega_list$
 END IF
 END FOR-EACH
 ordenar $\Delta\omega_list$ en orden descendente
2. % Iterar mientras exista una acción sin aplicar que mejore la
 % consistencia
 WHILE el primer elemento de $\Delta\omega_list > \delta$
 aplicar $\hat{K} = action(\{a, b\}, K)$, generar, si procede, \hat{V}_1^2 y \hat{V}_2^2
 eliminarlo de $\Delta\omega_list$

 % Actualizar $\Delta\omega_list$ con el $\Delta\omega$ de todas las aristas afectadas
 % por la aplicación de la acción
 FOR-EACH tupla $\{e, action\}$, $e = \{m, n\} \in closure(s, \hat{K})$, donde s es el
 complejo simplicial que lugara a la zona de influencia
 IF $action(e, K)$ verifica las pre-condiciones THEN
 estimar $\Delta\omega(e, action)$ con la expresión (5.6)
 almacenar $\Delta\omega(\{m, n\}, action)$ en $\Delta\omega_list$
 END IF
 END FOR-EACH

 actualizar las redes triangulares: $K = \hat{K}$, $V_1^2 = \hat{V}_1^2$ y $V_2^2 = \hat{V}_2^2$
 ordenar $\Delta\omega_list$ en orden descendente
 END WHILE

Figura 5.9: Algoritmo *greedy* propuesto para mejorar el registro lineal por trozos de dos imágenes.

rior que, debido a las diferencias en las posiciones de adquisición, presentan importantes diferencias geométricas, radiométricas y de contenido (ver fig. 5.10). La figura 5.10 también recoge las reconstrucciones 3D generadas a partir de las redes inicial y optimizada, así como la evolución de la consistencia global del registro y del número de acciones evaluadas en cada iteración. A partir de estos resultados se pueden subrayar fundamentalmente dos aspectos:

- a) la evaluación del grueso de las acciones (unas 600 aprox.) tiene lugar en la primera iteración, mientras que en el resto de iteraciones rara vez se superan las 25, lo que viene a corroborar lo expuesto arriba (ver fig. 5.10(a)), y
- b) la aplicación iterativa de acciones da lugar a mejores configuraciones de red, como ilustra la evolución de la consistencia global del registro. Fruto de esta mejora son una imagen registrada y una reconstrucción 3D sin apenas distorsiones (ver fig. 5.10(c)).

Pese a la mejora observada en la precisión, la imagen registrada muestra una región sin correspondencia alguna con la imagen de referencia, que aparece marcada en la figura 5.10(c). Este efecto indeseado se debe a la oclusión producida por la importante separación de las cámaras. Para evitarlo sería conveniente “desconectar” o eliminar subconjuntos del complejo simplicial inicial con objeto de que la red se ajuste con precisión a la escena observada, sin inconsistencias.

Convergencia

El procedimiento descrito garantiza la evaluación de todas las posibles acciones aplicables a cada arista de la red, terminando cuando todas han sido revisadas y ya no se produce mejora alguna. Para mejorar la eficiencia, las aristas que han sido divididas, intercambiadas y/o colapsadas pero que no mejoran la consistencia no son consideradas en futuras iteraciones y sólo serán revisadas de nuevo si cualquiera de sus caras se ve afectada por la aplicación de una acción previa. Esta consideración acelera la convergencia del algoritmo.

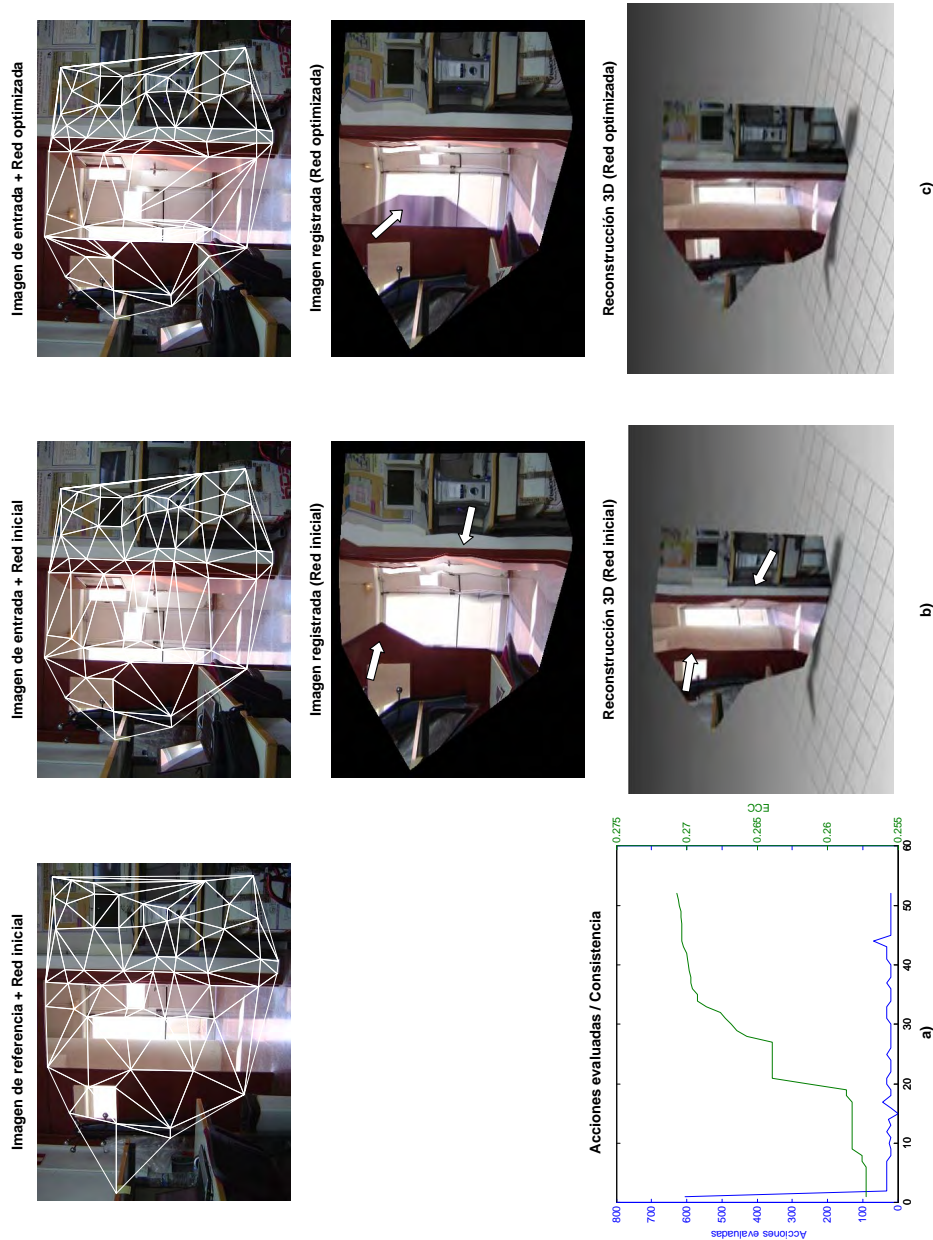


Figura 5.10: Ilustración del proceso de optimización propuesto. (a) Evolución de la consistencia global del registro y del número de acciones evaluadas en cada iteración. (b) Imagen registrada y reconstrucción 3D generada con la red inicial. (c) Imagen registrada y reconstrucción 3D realizada a partir de la red optimizada. Las flechas indican distintas distorsiones producidas por inconsistencias en las redes.

Un aspecto clave en este procedimiento y que, como se verá en la siguiente sección, redundará en una reducción significativa del coste computacional es el enfoque eminentemente local utilizado en el proceso de optimización. A diferencia de otros métodos que utilizan funciones de coste global (esto es, miden la consistencia global del registro) [77, 105], el método descrito aquí garantiza la convergencia del registro global mediante sucesivas mejoras locales.

Proposición 5.1 La mejora iterativa de la consistencia local da lugar a una mejora de la precisión global del registro.

Demostración: Supóngase que se realizan n particiones binarias de K , tales que $\bar{r}_i = K - r_i$, con $i = 1, \dots, n$, siendo n es el número de acciones aplicables, r_i el conjunto simplicial que determina la región de influencia de la acción i -ésima y \bar{r}_i el conjunto simplicial restante (ver fig. 5.11). Si la aplicación de la acción j -ésima da lugar a una mejora de la consistencia de las realizaciones geométricas de r_j , al permanecer el conjunto simplicial \bar{r}_j sin modificar, la consistencia de las realizaciones geométricas de $K = r_j \cup \bar{r}_j$, esto es, la consistencia global, también lo hace. Esto se verifica para las n particiones binarias. Por otro lado, puesto que siempre se selecciona la acción que produce la mayor mejora local (esto es, el mayor $\Delta\omega \geq \delta > 0$), la mejora global siempre es la máxima obtenible para las realizaciones geométricas disponibles en cada iteración.

Coste computacional

El orden computacional del método propuesto es proporcional al número de aristas compartidas de la red. Por cada arista se generan cuatro posibles acciones: un intercambio, una división, y dos acciones de eliminación (una por cada vértice no frontera). Por ejemplo, un conjunto de 258 pares de correspondencias da lugar a una triangulación de Delaunay con 759 aristas, de las cuales, 747 son compartidas, esto daría lugar a $747 \times 4 = 2988$ posibles acciones, aunque no todas ellas aplicables. Estos valores ilustran el tamaño del espacio de búsqueda, y que puede crecer a la vez que se insertan nuevos

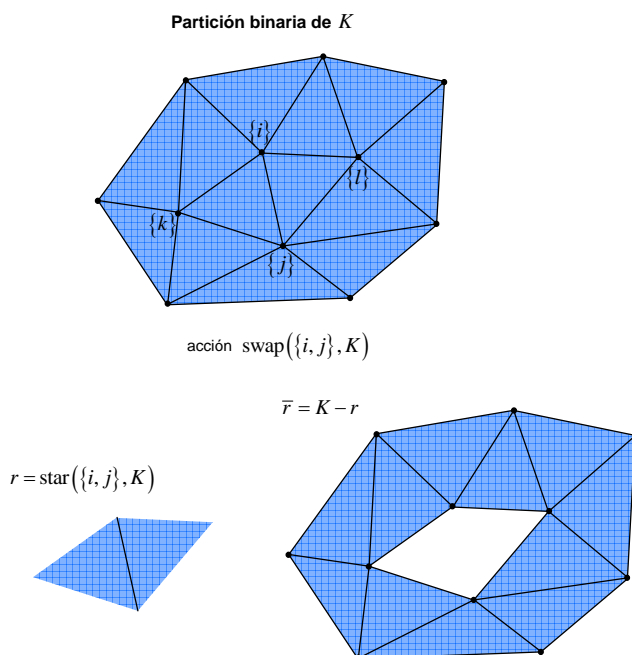


Figura 5.11: Partición binaria de una red triangular.

vértices. Aunque se ha comprobado experimentalmente que muchas acciones son descartadas, ya sea por no producir mejora o por no ser aplicables, la ejecución de una acción genera a su vez nuevas configuraciones que deben explorarse y, por tanto, más posibles acciones. En el peor de los casos una acción de intercambio puede generar 16 y una acción de división 32, la eliminación de una arista, sin embargo, dispara este número.

El enfoque local propuesto en este trabajo presenta diversas ventajas, en términos de eficiencia, con respecto a otros enfoques propuestos en la literatura, especialmente cuando el espacio de búsqueda crece sustancialmente. Estas ventajas residen en la reducción de:

1. El coste de evaluación de la función de energía. En [77, 105] Morris y Kanade y Vogiatzis *et al.* presentan algoritmos de optimización de redes triangulares 3D a partir de sus proyecciones sobre imágenes de la escena. Estos métodos utilizan funciones de costo globales para evaluar las posibles inconsistencias, lo que requiere la evaluación de todos los píxeles de las imágenes. La función propuesta en este trabajo sólo evalúa

la región de influencia de la acción.

2. El coste de ejecución del método de optimización. La exploración exhaustiva (en amplitud o profundidad) o mediante otros procedimientos, como el *back-tracking*, de un espacio de búsqueda como el descrito son computacionalmente inabordables, ya que obliga a generar, mantener y explorar, en cada iteración, todas las posibles configuraciones de red. La búsqueda *greedy*, por el contrario, sólo mantiene una configuración en memoria, la mejor, y en cada iteración sólo aplica una acción, la que produce la mayor mejora, descartando el resto.

En ocasiones, y dependiendo del tipo de aplicación, es posible reducir el conjunto de acciones aplicables en el proceso de optimización. Así, si la escena observada es prácticamente plana (como ocurre en muchas ocasiones, por ejemplo, en teledetección), la inserción de nuevos puntos ralentizaría el proceso de optimización sin producir una mejora sustancial en la consistencia del registro; igualmente, si los puntos iniciales para la construcción de la red son situados (manualmente o con un procedimiento automático) en bordes y esquinas de la escena, la división y eliminación de aristas podrían omitirse sin pérdida de precisión. Incluso en ocasiones, cuando el número de aristas es muy importante, se aconseja simplificar las redes mediante la aplicación iterativa de acciones de eliminación de aristas, para, a continuación, lanzar un proceso con todas las acciones. En todos estos casos se ganaría sustancialmente en eficiencia, al reducir significativamente el espacio de búsqueda. Todas estas estrategias se ilustran la sección 5.4.

5.4 Pruebas experimentales y comparativas

Esta sección muestra diversas comparativas y resultados experimentales que ilustran el comportamiento del proceso de optimización propuesto. Estos tests incluyen pruebas con imágenes reales de diferente naturaleza e imágenes sintéticas, así como comparativas con otras técnicas propuestas en los campos de los gráficos por ordenador, teledetección y medicina. Más concretamente, se compara con:

- procedimientos de optimización similares que utilizan como funciones de costo medidas de similitud basadas en intensidad: SSD [77] y diferencias de imágenes [78].
- métodos de registro no-rígidos tradicionales: polinomial [37], lineal por trozos (sin optimización) [39], cúbico por trozos [41] y *thin-plate-splines* [19].

5.4.1 Datos y metodología

En la realización de estos experimentos se han empleado imágenes del repositorio ALOI [36] (que incluye una amplia variedad de objetos), escenas urbanas (por ejemplo, fachadas de edificios) e imágenes de satélite de alta resolución. El propósito de elegir esta amplia variedad de imágenes es validar el comportamiento del método con distintos tipos de iluminación, contenidos de la imagen y ángulos de observación.

El conjunto de puntos de control (CP) utilizado para generar las redes triangulares iniciales se identifica automáticamente mediante el procedimiento descrito en la sección 4.4. A pesar de que este método proporciona un importante número de puntos correctamente emparejados, muchos de ellos son identificados en diferentes planos de la escena, los cuales, al ser triangulados mediante Delaunay [93], dan lugar a gran cantidad de triángulos no consistentes.

El objetivo es, por tanto, detectar y corregir estas configuraciones indeseadas y generar otra red que maximice el número de triángulos que caen sobre proyecciones de planos 3D de la escena, y por ende, mejore la consistencia del registro. Para cuantificar esta mejora se emplean los siguientes procedimientos:

1. Midiendo la consistencia global del registro. Para ello se utilizan las siguientes métricas:
 - (a) el ECC de las imágenes de referencia y registrada, aunque para ser preciso, sólo se evalúa la región de la imagen delimitada por la envolvente convexa de la red (expresión (5.9)), y

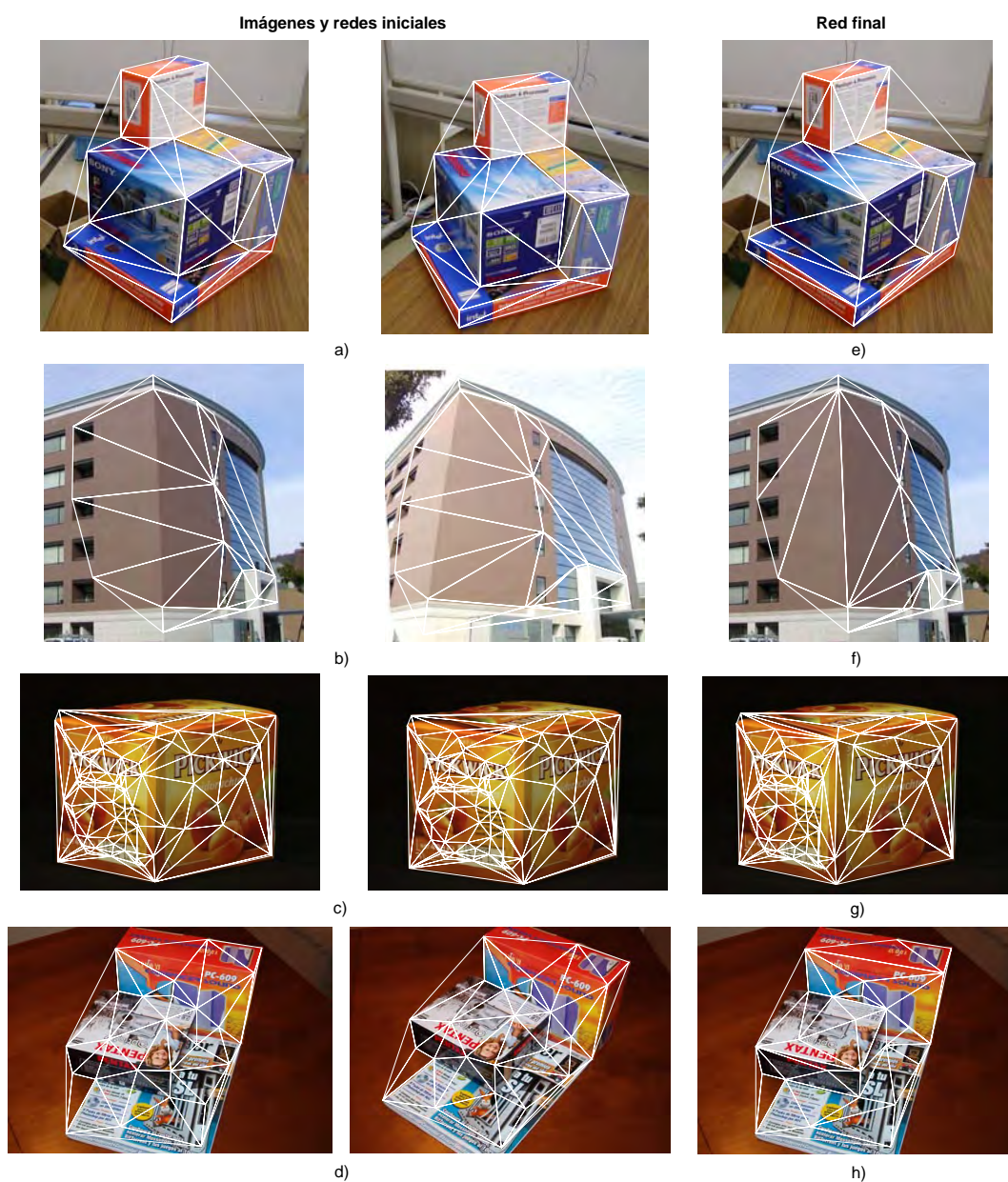


Figura 5.12: (a-d) Imágenes reales de escenas poliédricas y sus correspondientes redes triangulares de Delaunay. (e-h) Redes triangulares optimizadas por el método propuesto (sólo acciones de intercambio).

Tabla 5.1: Datos de las redes triangulares, número de acciones de intercambio aplicadas, consistencia global y tiempo empleado en el proceso de optimización.

Test	Morris & Kanade		Nakatsuji <i>et al.</i>		Nuestro método					
	# aristas	# acciones	Tiempo ²	ECC	# acciones	Tiempo	ECC			
a	56	13	11.42	0.239	14	9.21	0.235	13	1.42	0.239
b	47	13	7.39	0.226	10	8.11	0.321	8	1.65	0.332
c	275	102	133.78	0.423	65	72.82	0.398	27	3.93	0.425
d	65	18	38.43	0.263	11	9.75	0.211	16	3.15	0.264

²En los experimentos se ha empleado Matlab sobre un Pentium Core 2 Duo 6400. Para ganar en velocidad, algunas de las rutinas más costosas computacionalmente, se han desarrollado en C como extensiones de Matlab, así como el software de búsqueda y emparejamiento de correspondencias. El tiempo se expresa en segundos.

- (b) el RMSE de un conjunto de ICPs seleccionados en ambas imágenes (sólo en imágenes de satélite).
2. Evaluando las inconsistencias de la reconstrucción 3D (sin escala) generada a partir de las redes triangulares, antes y después de ser optimizadas. Puesto que no se dispone de un DTM de la zona en el caso de las imágenes de satélite ni se conoce la geometría exacta de la escena observada en el resto, la evaluación se realiza mediante inspección visual.

5.4.2 Comparativa con métodos similares

En esta sección se compara el comportamiento del método propuesto con los métodos de Morris y Kanade [77] y Nakatuji *et al.* [78]. Estos dos métodos detectan inconsistencias topológicas en redes triangulares 3D que aproximan la geometría de una escena y tratan de eliminarlas mediante el intercambio de aristas. Más concretamente, el primero emplea una búsqueda *greedy* y una función de coste global basada en SSD, mientras que el segundo propone una búsqueda aleatoria y una función de detección de inconsistencias basada en la diferencia de imágenes. Esta función emplea una plantilla de tamaño fijo que se aplica a la diferencia de los cuadriláteros afectados por la acción, los cuales han sido mapeados mediante transformaciones afines.

Como ambos métodos solamente emplean acciones de intercambio de aristas, con objeto de que éstos se puedan comparar con el método propuesto, el proceso de optimización sólo considera esta acción. La figura 5.12(a-d) muestra algunas de las imágenes empleadas en este trabajo, así como las redes iniciales generadas a partir de los conjuntos de CPs identificados en ellas.

La tabla 5.1 reúne los datos de las redes triangulares iniciales, así como el número de acciones aplicadas, el tiempo de procesamiento y la consistencia del registro correspondientes a cada uno de los métodos comparados. De estos resultados se pueden destacar diversos aspectos: el bajo coste computacional del método propuesto, su efectividad (consistencia *vs.* número de acciones) y su robustez a los cambios en la iluminación, como se aprecia en los resultados obtenidos para el par de imágenes de la figura 5.12(b) (dos imágenes

adquiridas en condiciones de iluminación muy diferentes).

Obsérvese la consistencia del registro del método de Morris y Kanade para el mismo par de imágenes, a lo que hay que sumar su coste computacional, prohibitivo cuando el número de aristas crece significativamente (ver los resultados para el par de imágenes de la figura 5.12(c)). El método de Nakatsuji *et al.* tiene un comportamiento similar a éste último a pesar de que emplea una función de detección local, que es costosa ya que requiere la estimación de 4 afinidades por arista.

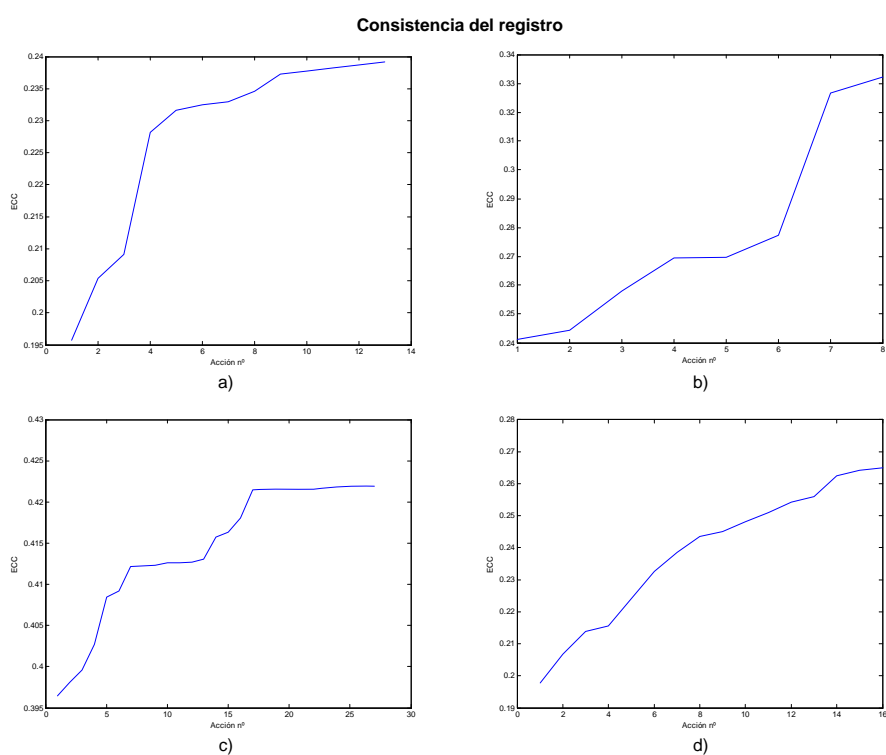


Figura 5.13: Evolución de la consistencia global (medida con el ECC) durante el proceso de optimización de los pares de imágenes mostradas en las figuras 5.12(a-d)). En general, se observa una mejora significativa de la consistencia con todos los pares analizados.

Las figuras 5.12(e-h) muestran las redes optimizadas por el método propuesto correspondientes a los pares de imágenes de las figuras 5.12(a-d). Como se aprecia en todos los ejemplos, el proceso de optimización finaliza con redes triangulares consistentes con la geometría de la escena y en todos

los casos, en menos de 25 iteraciones, como recogen las gráficas de la figura 5.13(a-d). Nótese que durante el proceso de optimización hay acciones que, aparentemente, no mejoran la consistencia global del registro (como revelan los pequeños trozos planos en las curvas de la figura 5.13(b,c)). Obsérvese que se dice aparentemente, ya que, como se menciona arriba, las gráficas muestran la consistencia global del registro, no la consistencia de la región afectada por la acción, que siempre mejora³. Estas acciones, a pesar de que apenas mejoran la consistencia global, dan lugar a configuraciones topológicas que son explotadas en iteraciones posteriores para generar redes consistentes con las proyecciones de la escena.

Los métodos analizados proponen el intercambio de aristas como única acción para la búsqueda de nuevas configuraciones topológicas. A continuación se muestra la red optimizada para las imágenes de la figura 5.12(d) (la más densa de las utilizadas) obtenida con el método propuesto, pero incluyendo el resto de las acciones descritas en este capítulo. Como se detalla en la sección previa, en primer lugar se simplifica la red mediante la aplicación iterativa de acciones de eliminación (ver fig. 5.14(b)), una vez simplificada, se lanza el proceso optimización con la totalidad de las acciones (ver fig. 5.14(c)). Comparando este resultado con el obtenido con la sola aplicación de acciones de intercambio (ver fig. 5.12(g)), la red continúa siendo consistente pero mucho menos densa. Se ha comprobado experimentalmente que esta estrategia ofrece mejores resultados (en términos de precisión y eficiencia) que un único proceso de optimización, sin simplificación previa.

5.4.3 Comparativa con métodos tradicionales

En esta sección se compara el comportamiento del método propuesto con diversas técnicas de registro no-rígidas utilizadas en la comunidad científica e incluidas en una amplia variedad de software: Matlab [99], Insight Segmentation and Registration Toolkit (ITK) [79], el software de Image Fusion Systems Research [51] o paquetes de procesamiento de imágenes para tele-

³Como se detalla en la sección 5.3.2, una acción se aplica si y sólo si, $\Delta\omega \geq \delta > 0$ lo que asegura un incremento monótono de la consistencia global, aunque cuanto menor sea la variación menor relevancia tendrá sobre ésta.

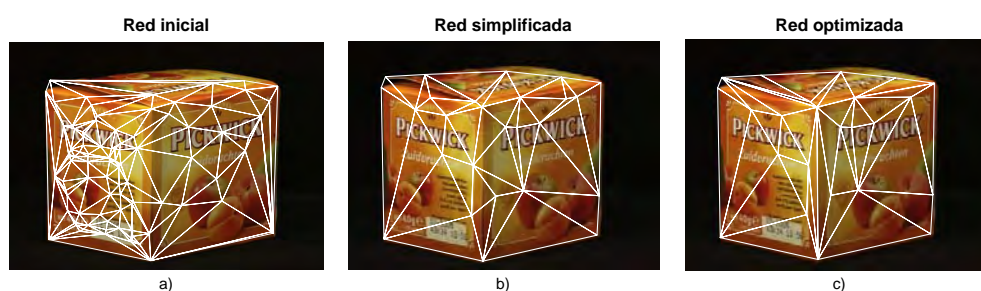


Figura 5.14: Red triangular optimizada con el método propuesto: (a) red inicial, (b) red simplificada (sólo acciones de eliminación) y (c) red optimizada (todas las acciones). Obsérvese como, dada la disposiciones de las correspondencias, no ha sido necesaria la inclusión de nuevos vértices.

detección como: ENVI/IDL [54], ERDAS [61], etc. Más concretamente, las técnicas analizadas aquí son funciones polinomiales (de 2º hasta 4º grado) (POL2 - POL4), lineales por trozos (PWL), cúbicas por trozos (PWC) y *thin-plate-spline* (TPS).

Para realizar los experimentos se han utilizado dos imágenes de alta-resolución del satélite QuickBird de aprox. 3000×1500 píxeles $\equiv 1800 \times 900$ m², de la ciudad costera del Rincón de la Victoria (Málaga) (ver fig. 5.15). Las imágenes, adquiridas en invierno y primavera, presentan diferencias radiométricas muy significativas (brillo y contraste, sombras, cambios en la cobertura terrestre, etc.) e importantes distorsiones inducidas por la adquisición fuera del nadir (con diferencias de 32.1 grados). A parte de las imágenes completas, en esta comparativa se analizan tres áreas de estudio extraídas de éstas (de aprox. of 900×750 píxeles):

- *área urbana*, que contiene edificios de diferentes alturas sobre un terreno casi plano,
- *área rural*, una región prácticamente plana pero, sin elementos de importante altura (excepto algunos árboles) y
- *área residencial*, una región accidentada que contiene edificios de baja altura (viviendas unifamiliares de una o dos plantas).

Al seleccionar estas regiones de estudio se pretende validar el rendimiento

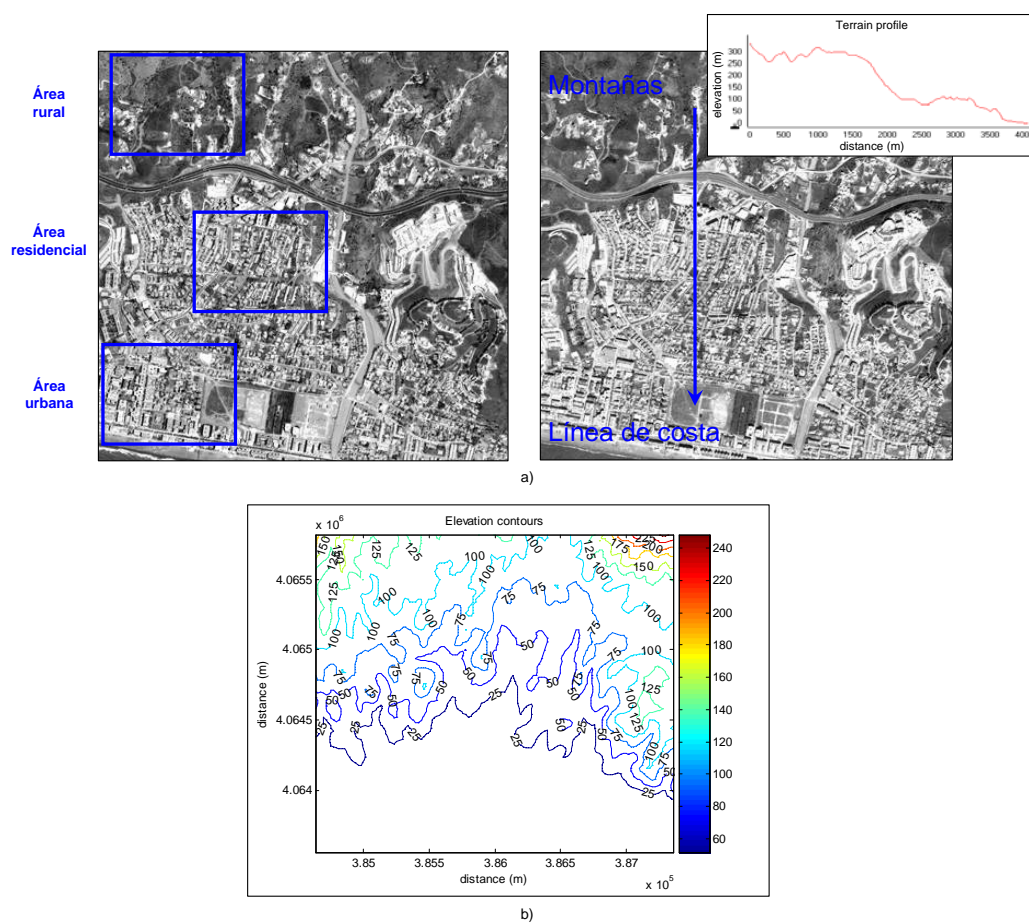


Figura 5.15: Imágenes de prueba utilizadas en el estudio comparativo. (a) Dos imágenes de la ciudad del Rincón de la Victoria (Málaga). Las regiones de interés utilizadas en los tests aparecen marcadas en la imagen de referencia. (b) Información de elevación utilizada para definir apropiadamente las regiones de estudio.

del método propuesto con escenas de distinta geometría y contenidos. Los resultados de esta comparativa aparecen resumidos en la tabla 5.2.

A la vista de estos resultados se pueden subrayar diversos aspectos:

- a) las funciones globales polinomiales, como era esperado, ofrecen peores resultados que las técnicas locales e híbridas, ya que las imágenes presentan distorsiones relativas de importancia, especialmente, en las zonas con relieve accidentado;

Tabla 5.2: Comparación del método propuesto con otros métodos de registro bien conocidos, concretamente: un método global polinomial (de 2° a 4° grado); dos métodos locales (lineal y cúbico por trozos); y un método híbrido (*thin-plate-spline*).

Función	RMSE (m.)				ECC			
	Urbana	Residencial	Rural	Img. completa	Urbana	Residencial	Rural	Img. completa
POL 2° orden	3.20	2.71	2.18	8.12	0.330	0.362	0.356	0.217
POL 3 ^{er} orden	3.27	2.05	1.82	7.67	0.310	0.404	0.383	0.219
POL 4° orden	2.86	2.07	1.75	6.00	0.338	0.406	0.393	0.243
PWL	2.00	1.26	1.06	1.82	0.447	0.527	0.472	0.449
PWC	2.24	1.33	1.27	2.15	0.425	0.494	0.444	0.435
TPS	2.02	0.92	1.03	1.66	0.437	0.556	0.487	0.464
Nuestro método	1.80	0.93	0.96	1.45	0.465	0.549	0.494	0.467

- b) entre los tres métodos tradicionales con alcance local, TPS es el más preciso de los analizados, lo que viene a confirmar lo observado en el capítulo 3, y
- c) el método propuesto en este trabajo iguala e incluso mejora los resultados del TPS.

La figura 5.16 muestra la evolución de la consistencia global al aplicarlo sobre las diferentes áreas de estudio, incluida la imagen completa. El lector puede observar que la consistencia mejora en todos los casos analizados, independientemente de la medida utilizada (RMSE o ECC). Estas mejoras se acentúan en las regiones urbana y residencial ya que, aunque la zona rural es casi plana, las redes de Delaunay generadas inicialmente son razonablemente precisas y no requieren de la aplicación de muchas acciones. Para las tres primeras áreas (a, b y c) el algoritmo requiere entre 20 y 25 iteraciones, mientras que para la imágenes completas requiere en torno a 71. El número de iteraciones, y por tanto, el coste computacional, depende fuertemente del número de aristas compartidas, como se observa en los tiempos de ejecución mostrados en la tabla 5.3.

Tabla 5.3: Datos de la red, número de acciones aplicadas y tiempo empleado en el proceso de optimización.

Región	# triángulos	# aristas	# acciones	Tiempo
Urbana	168	258	23	7.30
Residencial	155	239	17	6.80
Rural	162	248	24	7.73
Toda la imagen	502	759	71	39.04

Nótese que durante el proceso de optimización hay acciones que mejoran la consistencia del registro, según se observa en la curva del ECC (incrementando su valor), mientras que el RMSE permanece fijo (como revelan los pequeños trozos planos en las curvas). Esta discrepancia entre ambas métricas se debe al hecho que los triángulos implicados en las acciones no contienen ICPs en su interior y, consecuentemente, el RMSE no se ve alterado por la aplicación de la acción. Es importante recordar que, aunque los ICPs fuesen

identificados manualmente en el interior de todos y cada uno de los triángulos de la red es muy probable que, durante el proceso de optimización, algunos de los nuevos triángulos generados no tengan ICPs en su interior.

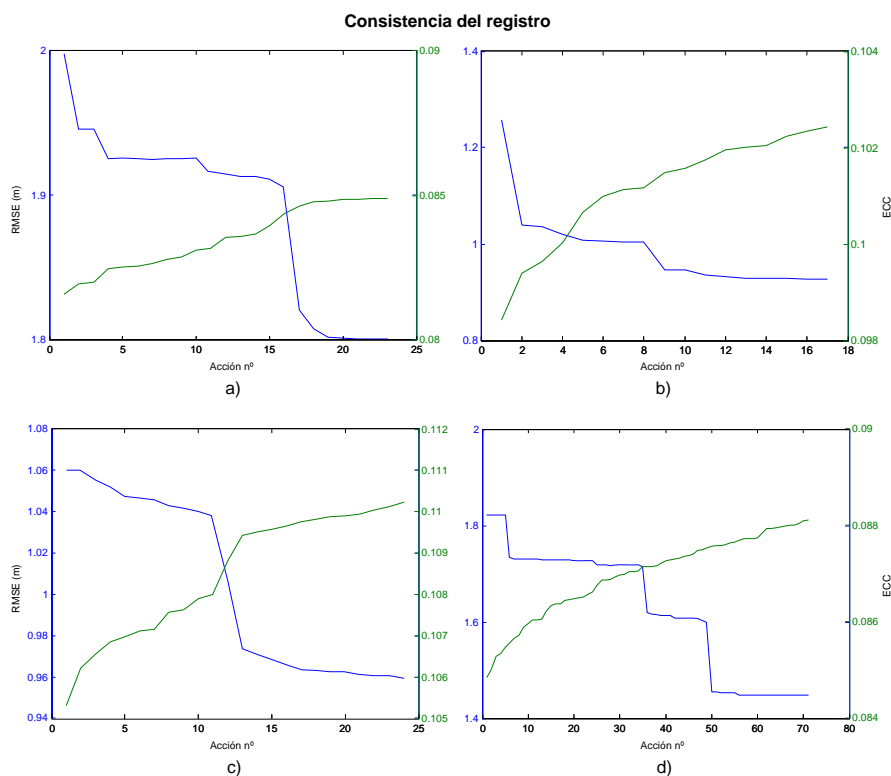


Figura 5.16: Evolución de la consistencia global (RMSE y ECC) durante el proceso de optimización de las áreas de estudio consideradas: (a) urbana, (b) residencial, (c) rural y (d) la imagen completa. Las gráficas muestra una mejora significativa en las áreas urbana y residencial (a,b), mientras que en el área rural (c), que presenta muchas menos distorsiones, la mejora no es tan acentuada.

Esta claro que el método propuesto requiere un coste computacional más alto que las funciones locales PWL (sin optimización), PWC y TPS, siendo éste su principal handicap. Aunque este inconveniente puede ser irrelevante en muchas aplicaciones, en aplicaciones robóticas o médicas, donde el factor tiempo es crítico, se puede convertir en un importante obstáculo. A pesar de esto, se pueden obtener mejores tiempos utilizando lenguajes de programación mucho más eficientes, librerías optimizadas como RapidMind® [84] u OpenVidia [81], y arquitecturas hardware capaces de paralelizar operaciones,

tales como GPUs de última generación (nVidia® y ATI®), procesadores CELL®, etc. que implementan primitivas para el manejo de redes y operaciones sobre éstas.

5.4.4 Reconstrucción 3D

Finalmente, para ilustrar la dualidad espacio-plano descrita en el capítulo 4, las redes generadas son utilizadas para generar una reconstrucción 3D por trozos de la escena observada (expresión 4.2). En los ejemplos incluidos en esta sección se ha utilizado el método de factorización para reconstrucción afín propuesto por Hartley y Zisserman en [47] (pag. 437).

A parte del indudable interés de la reconstrucción 3D, los modelos generados permiten verificar la efectividad del procedimiento propuesto mediante la comparación de la reconstrucción generada a partir de las redes iniciales y las optimizadas. Como no se dispone de información precisa sobre la geometría de la escena u objetos observados, este análisis se realiza mediante inspección visual. Las figuras 5.17 y 5.18 muestran las reconstrucciones 3D por trozos de las imágenes de prueba consideradas en este trabajo.

Cuando la reconstrucción se realiza a partir de las redes iniciales se observan distorsiones de líneas “quebradas” en los elementos rectos y en los bordes de los objetos poliédricos. Distorsiones que no aparecen cuando la red utilizada para generarla están convenientemente refinadas.

5.5 Conclusiones

En este capítulo se ha propuesto un procedimiento de optimización de redes triangulares conjugadas para mejorar la precisión del registro lineal por trozos. Para lograr esto se modifica iterativamente la topología y geometría de las redes mediante acciones de intercambio, división y eliminación de aristas. La función empleada para evaluar la mejora introducida por cada acción se basa en la información mutua, una medida notablemente más robusta que otras técnicas bien conocidas como son la NCC o la SSD, debido a su mayor robustez a los cambios en las condiciones de iluminación, cambios no

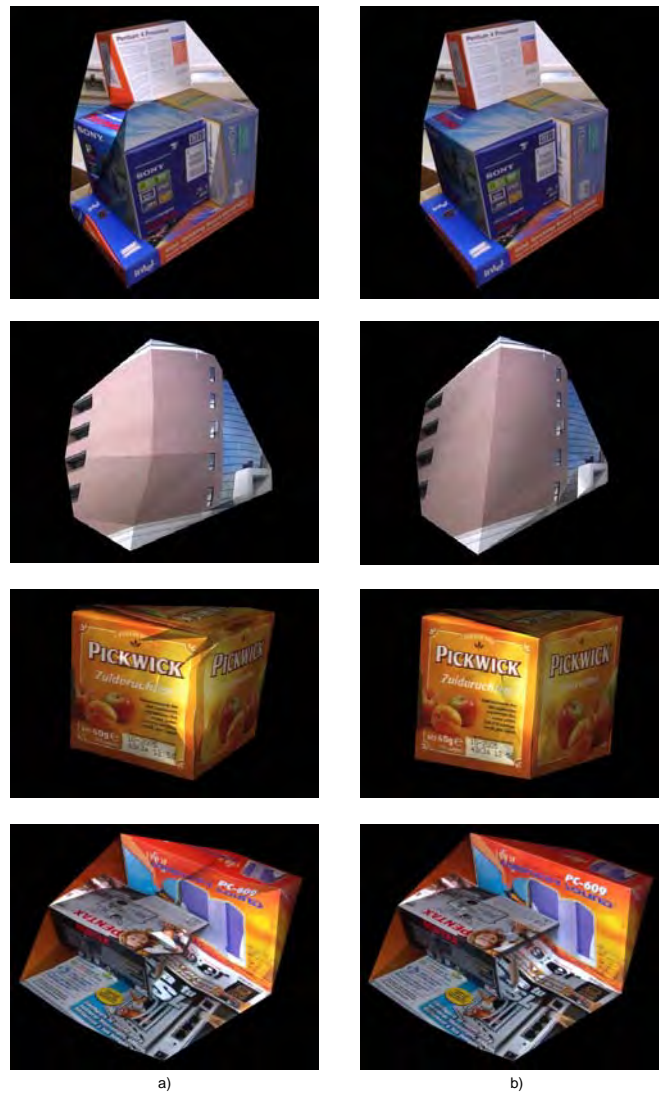


Figura 5.17: Reconstrucciones 3D generadas a partir de las redes: (a) iniciales y (b) refinadas de la figura 5.12.

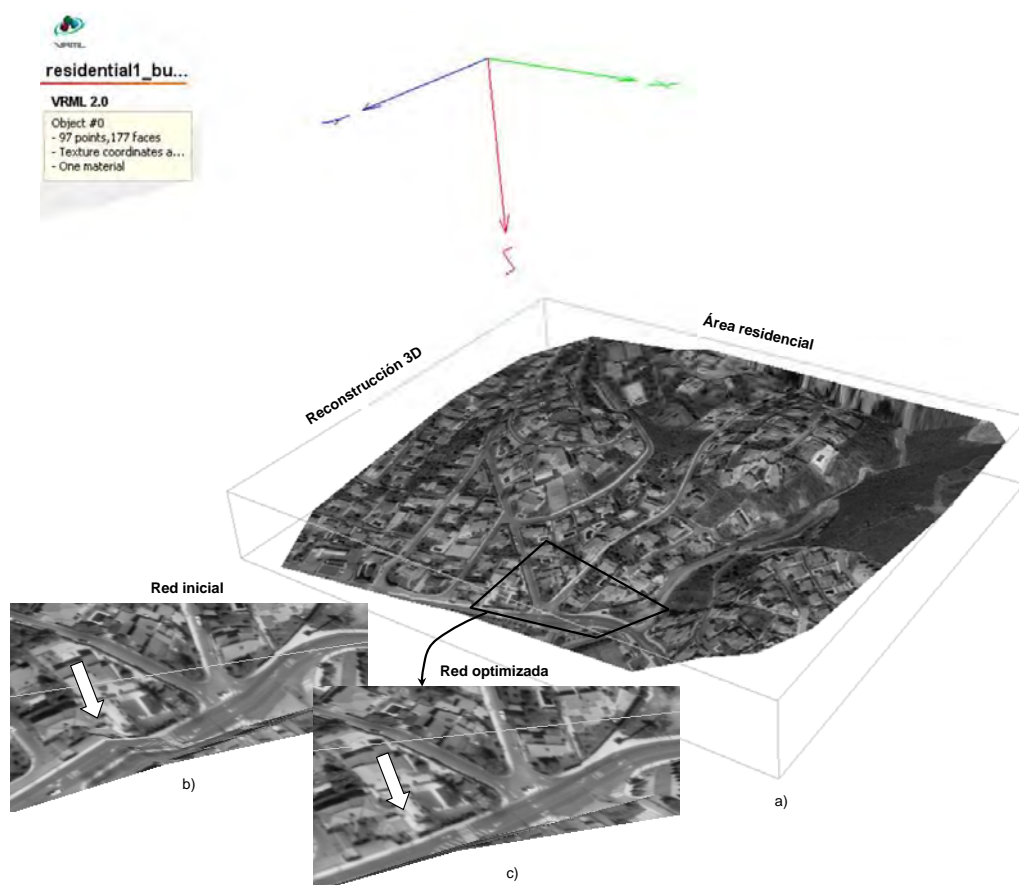


Figura 5.18: Reconstrucción 3D generada a partir de la red optimizada de la región residencial de la figura 5.15 (a). Las figuras (b) y (c) muestran fragmentos ampliados de la reconstrucción antes y después del proceso de optimización. Obsérvese la aparición de calles deformadas debida a una incorrecta triangulación (Delaunay).

lineales de la paleta de colores, contenido de las imágenes, etc. El procedimiento de optimización se formula como una búsqueda *greedy* que finaliza cuando la consistencia del registro no puede ser mejorada mediante los cambios topológicos y geométricos descritos. También se ilustra la posibilidad de generar aproximaciones 3D precisas de la escena observada mediante la reconstrucción lineal por trozos de las redes optimizadas. El método propuesto ha sido evaluado satisfactoriamente con diferentes tipos de imágenes: reales (de satélite, de interiores, de exteriores, etc.) y sintéticas (librería ALOI), adquiridas desde distintos ángulos de observación y presentando cambios en la iluminación.

Capítulo 6

Conclusiones y trabajos futuros

6.1 Conclusiones

La proliferación de sensores de elevadas prestaciones capaces de proporcionar imágenes de alta resolución que capturan diferentes propiedades de la escena (color, infrarrojos, etc.) obliga a un desarrollo constante de las técnicas utilizadas en su procesamiento, incluidas las técnicas de registro. En teledetección se tiene un claro ejemplo, los satélites de reciente lanzamiento capturan imágenes con una resolución de hasta 50 cm/píxel (los edificios, por ejemplo, aparecen ahora con gran detalle) desde diferentes órbitas y con distintas orientaciones. Estas diferencias en la captura dan lugar a imágenes con importantes diferencias geométricas relativas, de modo que, funciones de transformación (por ejemplo, las polinomiales) que proporcionaban resultados aceptables con imágenes de mucha menor resolución y captadas en el nadir no proporcionan los resultados esperados con éstas otras.

El registro de imágenes también es esencial en otras disciplinas, por ejemplo: en radiología se comparan y/o integran imágenes (a veces de distinta naturaleza) de un mismo sujeto para monitorizar la evolución de una masa tumoral o una lesión muscular; en robótica se combinan imágenes de un mismo cuerpo o escena para obtener una reconstrucción 3D que facilite la manipulación o la navegación del robot; en numerosos procesos industriales se comparan imágenes de las piezas fabricadas con objeto de detectar posi-

bles defectos, etc. Estos son sólo algunos ejemplos pero que hacen entrever la importancia de este proceso y que justifica el interés en desarrollar nuevos métodos y procedimientos en este campo. Fruto de este interés son este trabajo y los diferentes artículos científicos publicados durante su realización.

Esta tesis aborda la problemática asociada al proceso de registro, así como los distintos elementos que intervienen en él y que, de uno u otro modo, afectan a la precisión de los resultados obtenidos. Por su importancia, una parte significativa de este trabajo se centra en el análisis de dos aspectos claves de este proceso: la medida de la consistencia, crucial para determinar cómo de bueno ha sido el “encaje” de una imagen sobre la otra, y el tipo y alcance de la función de transformación utilizada y que determina la complejidad de las diferencias geométricas puede corregir. De esta evaluación se puede subrayar lo siguiente:

- La robustez de la información mutua (MI) (y sus diferentes variantes normalizadas) para medir la similitud radiométrica de dos imágenes, incluso cuando las imágenes presentan cambios radiométricos no funcionales. Su dependencia de la correcta estimación de las distribuciones de probabilidad de los niveles de intensidad de las imágenes. Estimaciones que depende, a su vez, de aspectos tales como el tamaño de las imágenes, la función de interpolación utilizada y la correcta estimación del histograma conjunto. Entre los procedimientos de estimación analizados destaca la ventana de Parzen, un método capaz de producir estimaciones poco ruidosas a partir de conjuntos pequeños de muestras.
- Las funciones de transformación locales (lineales por trozos) e híbridas (*thin-plate-spline*) exhiben un comportamiento superior a las funciones globales (polinomial) ya que pueden explotar la información, relativa a las diferencias geométricas de ambas imágenes, proporcionada por conjuntos de correspondencias muy densos, decayendo su rendimiento con conjuntos dispersos.

De todas las transformaciones evaluadas, sólo las funciones lineales por trozos proporcionan un rendimiento aceptable cuando la geometría de la escena es poliédrica, típica en entornos de interiores y urbanos. Para que este

método trabaje adecuadamente las regiones triangulares conjugadas deben caer sobre proyecciones de superficies planas de la imagen (triángulos consistentes con la escena). Sin embargo, los procedimientos utilizados para generar automáticamente las regiones conjugadas en el software disponible actualmente no garantiza este hecho, lo que introduce importantes errores en el registro.

Este trabajo analiza en detalle esta problemática proponiendo un enfoque novedoso para generar redes triangulares conjugadas que se disponen sobre las imágenes maximizando el número de triángulos consistentes. Para ello, plantea, mediante complejos simpliciales, una formalización del registro lineal por trozos que va más allá de un mero enfoque 2D, puesto que abarca no sólo la generación de redes triangulares 2D óptimas (desde el punto de vista de la consistencia del registro), sino también formaliza la reconstrucción 3D de la superficie de la escena observada. También se propone un procedimiento de optimización sustentado en esta formalización que modifica topológica y geoméricamente la red triangular (complejo simplicial) utilizada en el registro lineal por trozos para mejorar su precisión. Este método emplea el coeficiente de correlación de entropía como medida de consistencia.

El método propuesto se ha comparado experimentalmente con diversas funciones de transformación no-rígidas y procedimientos similares propuestos en los campos del registro de imágenes y de la reconstrucción 3D. Los resultados de estas comparativas reflejan la precisión del método propuesto, con errores en todos los casos inferiores a los otros métodos evaluados. Respecto a su coste computacional, su tiempo de ejecución es sensiblemente menor que los otros métodos iterativos analizados, aunque, lógicamente, continúa siendo excesivamente costoso comparado con los métodos de registro no-rígido basado en funciones polinomiales, lineales por trozos (sin optimización) y *thin-plate-splines*, lo que limita su uso en cierto tipo de aplicaciones.

6.2 Trabajos futuros

Una de las contribuciones de esta tesis es la formalización del proceso de registro lineal por trozos mediante complejos simpliciales. Esta formalización

asume que los complejos simpliciales utilizados en el registro son proyecciones de una superficie lineal por trozos que aproxima la escena observada. De este modo, si se determina la topología y/o geometría de red óptima, el proceso de registro no sólo mejora la precisión del registro, sino que, de generarse una reconstrucción 3D a partir de estas proyecciones, ésta estaría libre de inconsistencias y se ajustaría con precisión a la escena observada.

Pese a las posibilidades este novedoso enfoque, los complejos simpliciales manejados en nuestra implementación son redes 1-conectadas (esto es, todos los simplices máximos son triángulos y todos los triángulos tienen al menos un adyacente). Esta circunstancia da lugar a que, una vez finalizado el proceso de optimización, continúe habiendo triángulos que no se corresponden con proyecciones reales de la escena. Sería conveniente, por tanto, incluir en el proceso de optimización nuevas acciones capaces de conectar/desconectar coherentemente componentes de los complejos simpliciales ganando así en flexibilidad y potencia.

La búsqueda *greedy* asegura un coste computacional bajo, pero puede caer en mínimos locales. Existen mecanismos que pueden mejorar significativamente este aspecto aunque a costa de incrementar su coste computacional. Por ejemplo, una interesante opción consistiría en implementar un procedimiento de *lookahead* que permitiría anticipar como evolucionaría la consistencia del registro antes de aplicar una determinada acción, de forma que sería posible descartar una acción que, aunque produce la mayor mejora, da lugar a una configuración tal que no es posible seguir mejorando la consistencia.

Uno de los principales inconvenientes del método de registro propuesto, en comparación con otras técnicas de registro, es su elevado coste computacional. Quizás sea éste el mayor lastre para su incorporación en sistemas que requieran un tiempo de respuesta mucho más crítico como por ejemplo las aplicaciones robóticas. Pese a esto, la utilización de redes triangulares ofrece la posibilidad de paralelizar operaciones: sería relativamente fácil, por ejemplo, paralelizar el proceso de estimación de la función de coste local. Este hecho que adquiere mayor relevancia dada la reciente aparición de plataformas especialmente preparadas para realizar estas tareas. Este es el caso, por ejemplo, de librerías como RapidMind® [84] u OpenVidia [81] que pro-

porciona soporte para procesadores de múltiples núcleos (Intel® y AMD®), GPUs de última generación (nVidia® y ATI®), y procesadores Cell®.

Apéndice A

Entropía, entropía relativa e información mutua

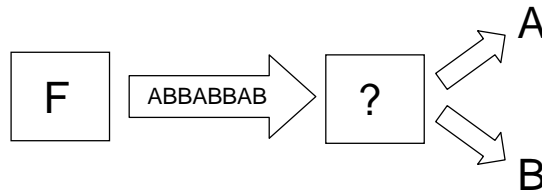
El concepto de información es muy amplio para ser recogido en una única definición. Sin embargo, para cualquier distribución de probabilidad es posible definir una cantidad que posee muchas de las propiedades que una medida de información debería tener, la *entropía*. La noción de entropía se extiende para definir la *información mutua*, que mide la cantidad de información que una variable aleatoria contiene sobre otra. La información mutua es un caso especial de una cantidad más general denominada *entropía relativa*, que mide la distancia entre dos distribuciones de probabilidad. La entropía, entropía relativa e información mutua están íntimamente relacionadas y comparten numerosas propiedades. Este apéndice define todas y cada una de estas cantidades, deriva algunas de sus propiedades e introduce la mayoría de las ideas básicas necesarias para entender algunos de los conceptos desarrollados en esta tesis. El lector puede referirse al libro de Cover y Tomas [26] para un estudio detallado sobre los elementos de la teoría de la información.

A.1 Entropía

La entropía es una medida de información, o más concretamente, de la incertidumbre de una fuente de información. Para poder entender el concepto

que encierra obsérvese el siguiente ejemplo:

Ejemplo A.1 Dada una fuente binaria de información, F, que proporciona el ganador de un partido de fútbol, esto es: el equipo A, con una probabilidad $p(A)$ de $3/4$, o el B, con probabilidad $p(B)$ de $1/4$, de tal manera que la situación que se tiene es la siguiente:



La transmisión de esta información a través del canal se puede abordar de distintas formas. Supóngase, por ejemplo, que se envían codificados en binario los resultados de tres partidos dando lugar a una codificación como la que sigue:

Cadena	Prob.	Código	Longitud
AAA	$27/64$	0	1
ABA	$9/64$	100	3
BAA	$9/64$	101	3
AAB	$9/64$	110	3
BBA	$3/64$	11100	5
BAB	$3/64$	11101	5
ABB	$3/64$	11110	5
BBB	$1/64$	11111	5

Observando el código se comprueba que se transfieren una mayor cantidad de bits en aquellas cadenas que tienen menor probabilidad de ocurrir. La longitud media del código es:

$$L = \frac{27}{64} \times 1 + \frac{9}{64} \times 3 + \frac{3}{64} \times 3 + \frac{1}{64} \times 1 = 3,47 \text{ bits}$$

De acuerdo con la nueva codificación, se tiene que $p(0) = 0,36$ y $p(1) =$

0,63. La *entropía* es una medida de la cantidad de información que se recibe al final del canal y se mide típicamente en bits.

Definición A.1 Dada una variable aleatoria discreta X que tiene una determinada distribución de probabilidades, $p(x)$, se define la *entropía* $H(X)$ como:

$$\begin{aligned} H_b(X) &= - \sum_{x \in X} p(x) \log_b p(x) \\ &= - \sum_{x \in X} p(x) \log_b p(x) \\ &= -E \log_b p(X) \end{aligned} \tag{A.1}$$

La unidad en que se mide depende de la base del logaritmo utilizada, estas son: BITS si es base 2, DITS si es base 10 o NATS si son logaritmos neperianos.

Ejemplo A.2 De acuerdo con el ejemplo anterior, la entropía de la fuente F sería:

$$H_2(X) = \frac{3}{4} \log_2 \frac{3}{4} + \frac{1}{4} \log_2 4 = 0,81 \text{ bits}$$

El resultado esperado más probable es el que, obviamente, menos contribuye al valor de la entropía; luego ésta puede entenderse como una ponderación de las contribuciones de cada suceso. A continuación se enumeran algunas propiedades de la entropía:

Lema A.0.1 $H_b(X) \geq 0$

Prueba: $0 \leq p(x) \leq 1$ implica $\log(1/p(x)) \geq 0$

Lema A.0.2 $H_b(X) = H_a(X) \log_b a$

Prueba: $\log_b p = \log_b a \log_a p$

Como se desprende del lema A.0.2, la entropía puede cambiarse de una base a otra (esto es, de una unidad a otra) multiplicando por el factor apropiado.

Ejemplo A.3 Si F es una fuente de información binaria tal que $p(0) = p$ y $p(1) = 1 - p = q$ la entropía (en base 2) vendría dada por la expresión:

$$H(X) = H(p, q) = -p \log_2 p - q \log_2 q$$

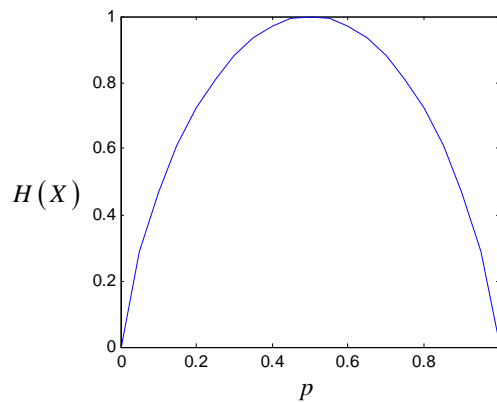


Figura A.1: Comparativa $H(X)$ versus p .

La figura A.1 muestra la representación gráfica de $H(X)$ en función de p . De la observación de esta curva se derivan algunas de las propiedades de la entropía: si $p(x) = 0$ ó $p(x) = 1$, esto es, no hay incertidumbre sobre el valor que tomará X , entonces $H(X) = 0$. Por el contrario, si la incertidumbre es máxima $p(x) = 1/2$, entonces el valor de $H(X)$ también lo es.

A.2 Entropía conjunta y entropía condicional

En esta sección se extiende el concepto de entropía a un par de variables aleatorias X e Y .

Definición A.2 Dado el par de variables aleatorias discretas (X, Y) con distribución conjunta de probabilidades, $p(x, y)$, se define la *entropía conjunta* $H(X, Y)$ como:

$$\begin{aligned} H(X, Y) &= - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(x, y) \\ &= -E \log_2 p(X, Y) \end{aligned} \quad (\text{A.2})$$

igualmente,

Definición A.3 Se define la *entropía condicional* $H(Y|X)$ de la variable aleatoria Y dada X como:

$$\begin{aligned} H(Y|X) &= \sum_{x \in X} p(x) H(Y|X = x) \\ &= - \sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2 p(y|x) \\ &= - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(y|x) \\ &= -E \log_2 p(Y|X) \end{aligned} \quad (\text{A.3})$$

La relación natural entre entropía conjunta y condicional queda patente en la denominada *regla de la cadena*, esta es: la entropía conjunta de un par de variables aleatorias es la entropía de una de ellas más la entropía condicional de la otra. Formalmente:

Teorema A.1 (Regla de la cadena)

$$H(X, Y) = H(X) + H(Y|X) \quad (\text{A.4})$$

Prueba:

$$\begin{aligned}
 H(X, Y) &= - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(x, y) \\
 &= - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 (p(x) p(y|x)) \\
 &= - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(x) - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(y|x) \\
 &= - \sum_{x \in X} p(x) \log_2 p(x) - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(y|x) \\
 &= H(X) + H(Y|X)
 \end{aligned}$$

Ejemplo A.4 Sean X e Y dos variables aleatorias discretas con la siguiente función de distribución:

Y \ X	X			
	1	2	3	4
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
4	$\frac{1}{4}$	0	0	0

Las distribuciones marginales de X e Y son $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})$ y $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, respectivamente, por lo tanto, $H(X) = 7/4$ bits y $H(Y) = 2$ bits. También,

$$\begin{aligned}
 H(X|Y) &= \sum_{i=1}^4 p(Y=i) H(X|Y=i) \\
 &= \frac{1}{4} H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right) + \frac{1}{4} H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{8}, \frac{1}{8}\right) + \\
 &+ \frac{1}{4} H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) + \frac{1}{4} H(1, 0, 0, 0) \\
 &= \frac{1}{4} \times \frac{7}{4} + \frac{1}{4} \times \frac{7}{4} + \frac{1}{4} \times 2 + \frac{1}{4} \times 0 \\
 &= \frac{11}{8} \text{ bits}
 \end{aligned}$$

(A.5)

Similarmente, $H(X|Y) = 13/8$ bits y $H(X, Y) = 27/8$ bits.

Nótese que $H(Y|X) \neq H(X|Y)$. Sin embargo, $H(X) - H(X|Y) = H(Y) - H(Y|X)$, una propiedad que será explotada más adelante.

A.3 Entropía relativa e Información mutua

La entropía de una variable aleatoria es una medida de la incertidumbre de dicha variable; es una medida de la cantidad media de información necesaria para describirla. En esta sección, se introducen dos conceptos relacionados: la entropía relativa y la información mutua.

La *entropía relativa* $D(p||q)$ es una medida de la distancia entre dos distribuciones de probabilidad p y q , o en otras palabras, de la ineficiencia de asumir la distribución p cuando la distribución verdadera es q . Por ejemplo, si se conoce la distribución verdadera q de una variable aleatoria, se podría construir un código con una longitud media $H(q)$. Si, en lugar de esto, se usa el código para una distribución p , se necesitarían $H(q) + D(p||q)$ bits de media para describir la variable aleatoria.

Definición A.4 La *entropía relativa* or distancia de Kullback Leibler entre dos funciones de masa de probabilidad $p(x)$ y $q(x)$ se define como:

$$\begin{aligned} D(p||q) &= \sum_{x \in X} p(x) \log_2 \frac{p(x)}{q(x)} \\ &= E_p \log_2 \frac{p(X)}{q(X)} \end{aligned} \quad (\text{A.6})$$

En esta definición, se emplea la siguiente convención (basada en la continuidad de los argumentos): $0 \log \frac{0}{q}$ y $p \log \frac{p}{0} = \infty$. A continuación se detallan algunas propiedades de la entropía relativa:

- $D(p||q) \geq 0$
- $D(p||q) = 0$ si y sólo si $p(x) = q(x)$

Ejemplo A.5 Sea $X = \{0, 1\}$ y dos distribuciones de probabilidad p y q

134A. ENTROPÍA, ENTROPÍA RELATIVA E INFORMACIÓN MUTUA

sobre X , tales que: $p(0) = 1 - r$, $p(1) = r$ y $q(0) = 1 - s$, $q(1) = s$. Entonces

$$D(p||q) = (1 - r) \log_2 \frac{1 - r}{1 - s} + r \log_2 \frac{r}{s}$$

y

$$D(q||p) = (1 - s) \log_2 \frac{1 - s}{1 - r} + s \log_2 \frac{s}{r}.$$

Si $r = s$, entonces $D(p||q) = D(q||p) = 0$. Por el contrario, si $r = 1/2$ y $s = 1/4$, se obtiene:

$$D(p||q) = \frac{1}{2} \log_2 \frac{1/2}{3/4} + \frac{1}{2} \log_2 \frac{1/2}{1/4} = 1 - \frac{1}{2} \log_2 3 = 0,2075$$

mientras que

$$D(q||p) = \frac{3}{4} \log_2 \frac{3/4}{1/2} + \frac{1}{4} \log_2 \frac{1/4}{1/2} = \frac{3}{4} \log_2 3 - 1 = 0,1887.$$

Nótese que excepto para $r = s$, $D(p||q) \neq D(q||p)$.

La información mutua $I(X; Y)$ es una medida de la cantidad de información que una variable aleatoria contiene sobre otra, esto es, la reducción en la incertidumbre de una variable aleatoria dado el conocimiento de la otra.

Definición A.5 Considérense dos variables aleatorias X e Y con una función de masa de probabilidad conjunta $p(x, y)$ y una funciones de masa de probabilidad marginales $p(x)$ y $p(y)$. La *información mutua* $I(X; Y)$ es la entropía relativa del distribución conjunta y el producto de las distribuciones marginales, esto es:

$$\begin{aligned} I(X; Y) &= \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 \frac{p(x, y)}{p(x) p(y)} \\ &= D(p(x, y) || p(x) p(y)) \\ &= E_{p(x, y)} \log_2 \frac{p(X, Y)}{p(X) p(Y)} \end{aligned} \tag{A.7}$$

A.4 Relación entre entropía e información mutua

Dado que $p(x, y) = p(y)p(x|y)$, la información mutua podría escribirse como:

$$\begin{aligned}
 I(X; Y) &= \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \\
 &= \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 \frac{p(x|y)}{p(x)} \\
 &= - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(x) + \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(x|y) \\
 &= - \sum_{x \in X} p(x) \log_2 p(x) - \left(- \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(x|y) \right) \\
 &= H(X) - H(X|Y) \tag{A.8}
 \end{aligned}$$

De lo que se deduce que la información mutua $I(X; Y)$ es la reducción en la incertidumbre de X dado el conocimiento de Y . Verificándose, por simetría, lo siguiente:

$$I(X; Y) = H(Y) - H(Y|X) \tag{A.9}$$

Es decir, X reduce la incertidumbre de Y tanto como Y reduce la incertidumbre X .

Puesto que $H(X, Y) = H(X) - H(Y|X)$, como se detalla en la sección A.2, se tiene:

$$I(X; Y) = H(X) + H(Y) - H(X, Y) \tag{A.10}$$

Finalmente, nótese que

$$I(X; X) = H(X) - H(X|X) = H(X) \tag{A.11}$$

Es decir, la información mutua de una variable aleatoria con ella misma es su entropía. Esta es la razón de que la entropía se denomine, en ocasiones,

como *información propia*¹.

Resumiendo los resultados de esta sección, se tiene el siguiente teorema.

Teorema A.2 (Información mutua y entropía)

$$\begin{aligned}
 I(X;Y) &= H(X) - H(X|Y), \\
 &= H(Y) - H(Y|X), \\
 &= H(X) + H(Y) - H(X,Y), \\
 &= I(Y;X) \\
 I(X;X) &= H(X) \\
 I(X;X) &\leq \min(H(X), H(Y)) \\
 &\leq (H(X) + H(Y))/2 \\
 &\leq H(X,Y)
 \end{aligned}$$

La figura A.2 ilustra, mediante un diagrama de Venn, la relación entre $H(X)$, $H(Y)$, $H(X,Y)$, $H(X|Y)$, $H(Y|X)$ y $I(X;Y)$. Nótese que la información mutua $I(X;Y)$ se corresponde con la intersección de la información de X con la información Y .

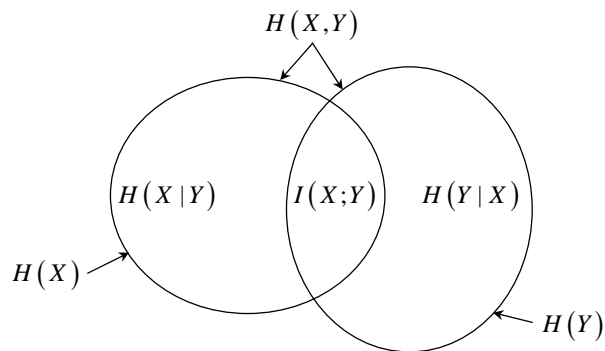


Figura A.2: Relación entre entropía e información mutua.

Ejemplo A.6 La información mutua para la distribución de probabilidad

¹Traducción de los términos ingleses Self Information.

conjunta del ejemplo A.4 sería $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = 0,375$ bits.

138A. ENTROPÍA, ENTROPÍA RELATIVA E INFORMACIÓN MUTUA

Apéndice B

Métodos de regresión robusta

La mayoría de los algoritmos utilizados en visión por computador generan interpretaciones de los datos observados. En numerosas ocasiones, el objetivo de éstos es la minimización de alguna función de coste, utilizándose en muchos casos, la suma de los errores de los residuos al cuadrado como estimador (esto es, un ajuste de mínimos cuadrados). Esto se debe, fundamentalmente, a dos razones:

1. es el estimador de probabilidad idóneo cuando los errores son Gaussianos y
2. existen numerosos algoritmos rápidos y estables que hacen uso de él para obtener la solución.

Cuando el conjunto de datos está contaminado por espurios, la primera justificación ya no es aplicable y la segunda se convierte en irrelevante, ya que la solución proporcionada queda muy lejos de la solución verdadera, convirtiendo la estimación en un proceso completamente arbitrario.

La figura B.1 ilustra este hecho mostrando los resultados obtenidos al realizar un ajuste lineal a un conjunto de datos contaminado con espurios mediante (ver tab. B.1): a) un procedimiento de mínimos cuadrados y b) un algoritmo de muestreo aleatorio, concretamente RANSAC.

Obsérvese como el estimador de mínimo error cuadrático proporciona una solución errónea (Fit 1) arrastrado por la muestra número 7, mientras que

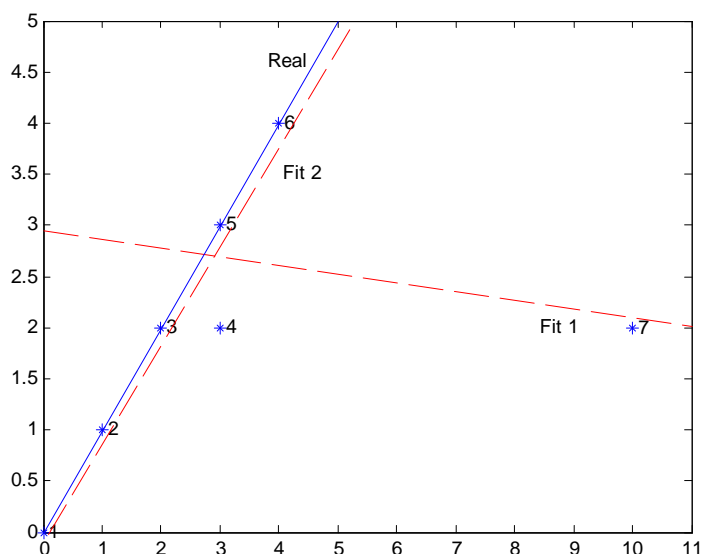


Figura B.1: Comparativa de un ajuste de mínimos cuadrados (Fit 1) y el procedimiento de estimación robusta RANSAC (Fit 2).

RANSAC etiqueta dicha muestra como espurio, lo descarta del conjunto final de muestras, proporcionando una solución que se ajusta mejor a la real (Fit 2).

El *punto de ruptura* de un estimador es la proporción más pequeña de espurios que puede forzar que el valor de un cálculo esté fuera de un rango determinado. Para un estimador de mínimo error cuadrático estándar basta con un espurio para alterar el resultado, por tanto, su punto de ruptura es $1/n$ donde n es el número de muestras del conjunto.

Tabla B.1: Datos utilizados en la comparativa de la figura B.1.

<i>Punto</i>	x	y	$x - \bar{x}$	$y - \bar{y}$
1	0	0	-3.29	-2
2	1	1	-2.29	-1
3	2	2	-1.29	0
4	3	2	-0.29	0
5	3	3	-0.29	1
6	4	4	0.71	2
7	10	2	6.71	0

A continuación se describen en detalle dos de los métodos de regresión utilizados en esta tesis: RANSAC y MAPSAC. Por completitud, también se incluye el estimador LMedS. El lector puede encontrar información detallada sobre regresión robusta y detección de espurios en [86, 102]. En este último trabajo, Torr analiza diversos métodos de estimación y su aplicación para recuperar robustamente la geometría epipolar.

B.1 LMedS

Dadas n muestras, el método de regresión robusta LMedS (Least Median of Squares) [86] resuelve el siguiente problema de minimización no lineal:

$$\min \{ \text{median} (r_i^2) \} \quad (\text{B.1})$$

donde r_i son los residuos de las n muestras utilizadas para ajustar el modelo.

No hay una solución analítica para este problema, y debe ser resuelto buscando en el espacio de todas las posibles soluciones. Una búsqueda exhaustiva obliga necesariamente a estimar todas las posibles combinaciones. Matemáticamente, este número viene dado por la expresión:

$$C_n^p = \frac{(n)!}{(n-p)!} \quad (\text{B.2})$$

donde p es el número mínimo de muestras requerido para determinar la solución.

Por ejemplo, si $n = 70$, y $p = 2$, se requerirían cerca de 5000 estimaciones. Si $p = 8$, serían necesario $3,8 \times 10^{14}$. Este es, muy a menudo, un problema inabordable, y sólo pueden analizarse subconjuntos de muestras. En este caso, el número mínimo de posibles ajustes viene dado por la expresión:

$$m = \frac{\log(1-P)}{\log(1-(1-\varepsilon)^p)} \quad (\text{B.3})$$

donde P es la probabilidad deseada de encontrar un subconjunto sin espurios y ε es la fracción estimada de espurios, respectivamente.

En este caso, si $P = 80\%$, y ε está entorno al 50% , utilizando $p = 8$ puntos, el número mínimo de subconjuntos aleatorios requerido es aproximadamente 600, mientras que si $p = 2$ sólo 9 son suficientes. Las m diferentes soluciones son comparadas, y se selecciona la única que minimiza la mediana de los residuos al cuadrado (expresión (B.1)). Puesto que la mediana no incluye el 50% de los mayores errores, (B.1) no se ve afectada por los espurios.

Una vez seleccionada la muestra que proporciona dicha solución, se procede a clasificar cada elemento del conjunto inicial como dato válido o espurio, en función de la magnitud de su residuo con respecto al modelo seleccionado, esto es:

$$\begin{cases} \text{espurio} & r_i^2 > T^2 \\ \text{válido} & r_i^2 \leq T^2 \end{cases} \quad (\text{B.4})$$

donde T es el umbral de corte.

Finalmente, para obtener una solución eficiente en presencia de ruido, se utiliza el conjunto de datos válidos obtenido con (B.4) para calcular una regresión de mínimos cuadrados.

B.2 RANSAC

Dadas n muestras, el método de regresión robusta RANSAC (Random Simple Consensus) [31] resuelve el siguiente problema de minimización:

$$\min \left\{ \sum_i \rho(r_i^2) \right\} \quad (\text{B.5})$$

donde ρ es una función “todo o nada” de la forma:

$$\rho(r^2) = \begin{cases} 0 & r^2 < T^2 \\ 1 & r^2 \geq T^2 \end{cases} \quad (\text{B.6})$$

El proceso de estimación es análogo al LMedS: se comparan las m diferentes soluciones obtenidas a partir de otros tantos subconjuntos aleatorios de p muestras, y se selecciona la única que minimiza el número de espurios

(expresión (B.5)). Cuando se selecciona la mejor solución, se clasifica cada elemento como válido o espurio, en función de su residuo con respecto al modelo seleccionado (expresión (B.4)). Finalmente, se utiliza el conjunto de muestras válidas para calcular una regresión de mínimos cuadrados.

Nótese que es posible, ajustar adaptativamente el valor de m en cada iteración con objeto de reducir el tiempo de estimación. Podemos, por tanto, reescribir la expresión (B.3) como sigue:

$$m_{j+1} = \frac{\log(1 - P)}{\log\left(1 - \left(1 - \frac{q_j}{n}\right)^p\right)} \quad (\text{B.7})$$

donde q_j denota el número muestras válidas para la iteración j .

B.3 MAPSAC

El algoritmo RANSAC ha demostrado ser muy eficiente para la estimación robusta de parámetros, sin embargo, si el umbral de corte T es demasiado alto, la estimación deja de ser robusta [102]. Como detalla en la sección B.2, RANSAC trata de minimizar la función de coste “todo o nada” (B.6). Si T^2 es demasiado alto todas las muestras serían válidas y por tanto, todas las soluciones tendrían el mismo coste. De igual forma, es posible que aunque no todas las muestras fuesen válidas, podría haber varias soluciones con el mismo coste ¿cual sería, por tanto, la mejor solución? En [101], Torr y Zisserman proponen una nueva función de coste que trata de resolver esta limitación. Para ello, plantean el siguiente problema de minimización:

$$\min \left\{ \sum_i \rho(r_i^2) \right\} \quad (\text{B.8})$$

donde ρ es una función de la forma:

$$\rho(r^2) = \begin{cases} r^2 & r^2 < T^2 \\ T^2 & r^2 \geq T^2 \end{cases} \quad (\text{B.9})$$

Como puede observarse, los espurios son penalizados con un valor fijo,

mientras que las muestras válidas son puntuados en función de cómo se ajustan a los datos.

Este nuevo algoritmo recibe el nombre de MAPSAC (Maximum a Posteriori Consensus) y es una generalización de MLESAC para el caso Bayesiano. El proceso de estimación es análogo al propuesto para el RANSAC.

Apéndice C

Topología algebraica

En este apéndice se introducen conceptos básicos de topología algebraica tales como: símplex, complejo simplicial abstracto, complejo simplicial, etc. utilizados para la formalización del registro lineal por trozos, así como también, del proceso de optimización (topológica y geométrica) propuesto en esta tesis. La topología algebraica, o más concretamente los complejos simplicial, se utilizan en multitud de disciplinas científicas, incluyendo: ingeniería del software [87], gráficos por ordenador [49], ciencias sociales [14], etc. ya que permiten modelar espacios topológicos n -dimensionales; sus invariantes: huecos, túneles; y las relaciones entre sus símplexes constituyentes: adyacencia, conexión, etc. El lector puede encontrar más información sobre el contenido de este apéndice en [55].

C.1 Simplex

Un símplex o n -símplex es el análogo en n dimensiones de un triángulo. Más exactamente, un símplex (en plural, símplexes) es la envoltura convexa de un conjunto de $(n + 1)$ puntos afinmente independientes (esto es, en posición general) en un espacio Euclídeo de dimensión n o mayor. Así por ejemplo, un 0-símplex es un punto, un 1-símplex un segmento, un 2-símplex un triángulo, un 3-símplex un tetraedro, etc. (en cada caso, con su interior) (ver fig. C.1). Nótese que, cada uno de estos n -símplexes pueden estar definidos en un

espacio de dimensión n o superior.

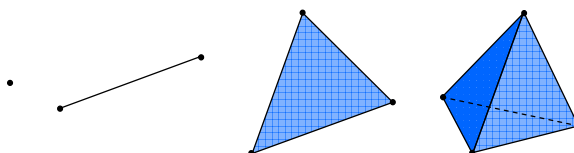


Figura C.1: n -símplices, con $n = 0, 1, 2, 3$.

La envoltura convexa de cualquier conjunto no-vacío de los $n + 1$ puntos que definen un n -simplex se llama *cara* del simplex. Se denomina m -cara del n -simplex la envoltura convexa de $m + 1$ puntos de éste. Más concretamente, las 0-caras se llaman vértices, las 1-caras lados, las $(n - 1)$ -caras facetas y la única n -cara es el n -simplex en sí mismo. Las caras también son símplices. Por ejemplo, las caras de un triángulo son sus vértices, sus aristas y el propio triángulo. Se dice que un símplice es *regular* si también es un politopo regular, por ejemplo, un triángulo equilátero.

El n -símplice *estándar* o *canónico* es el subconjunto de \mathbb{R}^{n+1} dado por:

$$\Delta^n = \left\{ (t_0, \dots, t_n)^\top \in \mathbb{R}^{n+1} \mid \sum_{i=0}^n t_i = 1 \text{ y } t_i \geq 0 \right\} \quad (\text{C.1})$$

El simplex estándar Δ^n cae en el hiperplano afín que se obtiene al eliminar la restricción $t_i \geq 0$ en (C.1), y es claramente regular (ver fig. C.2).

Los vértices de un n -simplex estándar son los puntos n -dimensionales:

$$\begin{aligned} e_0 &= (0, 0, 0, \dots, 0)^\top, \\ e_1 &= (1, 0, 0, \dots, 0)^\top, \\ &\vdots \\ e_n &= (0, 0, 0, \dots, 1)^\top. \end{aligned}$$

Existe un aplicación canónica f de un n -simplex estándar a un n -simplex

arbitrario con vértices $\{v_0, \dots, v_n\}$ dada por:

$$f : (t_0, \dots, t_n)^\top \mapsto \sum_{i=0}^n t_i v_i \quad (\text{C.2})$$

Los coeficientes t_i son las *coordenadas baricéntricas* de un punto en el n -símplex. Este símplex arbitrario se denomina habitualmente n -símplex afín, para enfatizar que la aplicación canónica es una transformación afín.

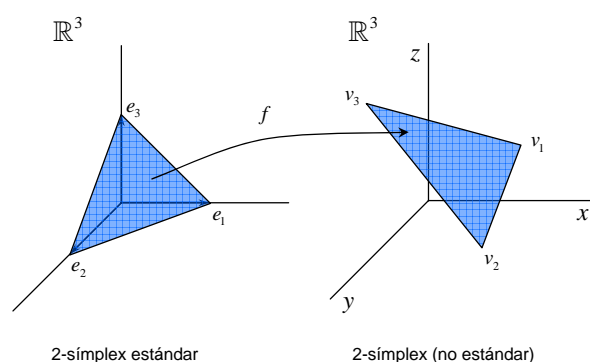


Figura C.2: Transformación canónica de un 2-símplex estándar a un 2-símplex cualquiera.

C.2 Complejo simplicial abstracto

Un complejo simplicial abstracto K consta de dos conjuntos: un conjunto de m vértices $\{1, \dots, m\}$ y un conjunto de subconjuntos de estos vértices, llamados caras o símplices de K , tales que, cualquier subconjunto de exactamente un vértice es un símplex de K y cada subconjunto de un símplex de K también es un símplex de K . Formalmente:

Definición C.1 Dado un conjunto universal S , y una familia K de subconjuntos no vacíos de S , K es un *complejo simplicial abstracto* si:

$$\forall \sigma \subset S, \text{ si } \sigma \in K, \text{ entonces } \forall \tau \subset \sigma \text{ se verifica que } \tau \in K$$

La dimensión de un simplex $\sigma \in K$ es el número de sus vértices menos uno (esto es, $\dim(\sigma) = \text{card}(\sigma) - 1$) y la dimensión de K es el máximo de las dimensiones de sus símplexes (esto es, $\dim(K) = \max \{\dim(\sigma), \sigma \in K\}$). La notación σ^d denota el símplex σ de dimensión d .

La figura C.3 muestra un complejo simplicial abstracto de dimensión 2, K^2 , junto con sus correspondientes vértices, aristas y triángulos.

Definición C.2 La *estrella* de un símplex $\sigma^d \in K$, denotado como $\mathbf{star}(\sigma^d)$, es el conjunto de símplexes de los que σ^d es cara (ver fig. C.3).

Gracias a la definición de estrella de un d -símplex es posible definir diferentes símplexes y complejos simpliciales característicos, estos son:

- Los $(d + 1)$ -símplexes de $\mathbf{star}(\sigma^d)$ son los *padres* de σ^d :
 - σ^d es *frontera* si tiene sólo un padre y *compartido* en caso contrario.
 - σ^d es *libre* si tiene al menos un padre y *máximo* en caso contrario.
- Finalmente, se dice que complejo simplicial K es *regular* si todos los símplexes máximos tienen la misma dimensión.

Definición C.3 El *cierre* de un símplex $\sigma^d \in K$, denotado como $\mathbf{closure}(\sigma^d)$, es el complejo simplicial más pequeño que contiene todos los símplexes de σ^d (ver fig. C.3).

C.3 Complejo simplicial o poliedro

Para generar una triangulación a partir del complejo simplicial abstracto K se procede de forma análoga a lo descrito en la sección C.1 para el símplex estándar o canónico. En primer lugar, se determina el espacio topológico asociado a K y en segundo lugar, se mapea a un espacio geométrico, típicamente \mathbb{R}^3 .

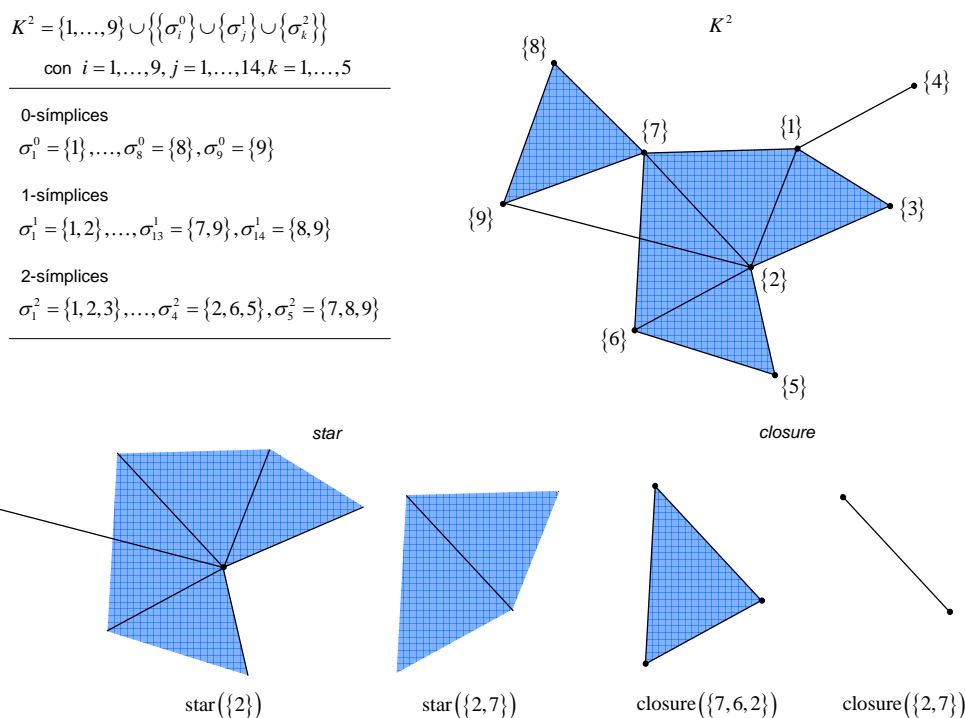


Figura C.3: Ilustración de un complejo simplicial abstracto de dimensión 2 y algunos de sus subconjuntos: estrella, cierre.

Dado el complejo simplicial abstracto K , se identifican sus m vértices $\{1, \dots, m\}$ con los vectores bases estándar $\{e_1, \dots, e_m\}, e_j \in \mathbb{R}^m$. Para todo simplex $\sigma \in K$, se define su correspondiente *simplex abierto*, denotado como $|\sigma| \subset \mathbb{R}^m$, como el interior de la envolvente convexa de sus vértices:

$$|\sigma| = \left\{ (t_1, \dots, t_m)^T \in \mathbb{R}^m \mid \sum_{i=1}^m t_i = 1 \text{ y } t_i \geq 0 \right\} \tag{C.3}$$

Definición C.4 La *realización topológica* de K , denotada como $|K|$, se define como la unión de los simplexes abiertos de K , esto es, $|K| = \cup_{\sigma \in K} |\sigma|$.

Definición C.5 La *realización geométrica* de K es la imagen $\phi_V(|K|)$, donde $\phi_V : \mathbb{R}^m \mapsto \mathbb{R}^3$ es una aplicación lineal que mapea los $e_j \in \mathbb{R}^m$ a $v_j \in \mathbb{R}^3$.

El subíndice V en ϕ_V enfatiza el hecho de que la aplicación viene especificada por el conjunto V de posiciones de los vértices. No todo conjunto $V = \{v_1, \dots, v_m\}$ da lugar a un *embedding* de $\phi_V(|K|) \subset \mathbb{R}^3$. Si ϕ_V es un homeomorfismo entre $|K|$ y su imagen $\phi_V(|K|)$, la realización geométrica se denomina *complejo simplicial* o *poliedro* (ver fig. C.4). Formalmente:

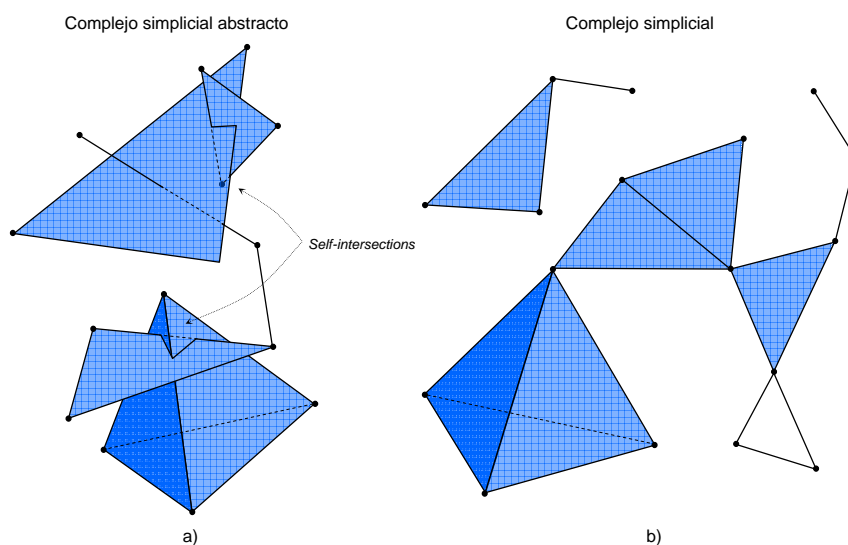


Figura C.4: Ilustración de un complejo simplicial de dimensión 3. La figura ilustra gráficamente las diferencias entre: (a) un complejo simplicial abstracto y (b) un complejo simplicial. En un complejo simplicial abstracto la intersección de dos símplex puede no ser ni el conjunto vacío ni una cara de ambos, dando lugar a intersecciones propias o *self-intersections*, mientras que en el complejo simplicial los símplex sólo se cortan en las caras.

Definición C.6 Dado un conjunto universal S , y una familia K de subconjuntos no vacíos de S , K es un *complejo simplicial* si verifica lo siguiente:

1. $\forall \sigma \subset S$, si $\sigma \in K$, entonces $\forall \tau \subset \sigma$ se verifica que $\tau \in K$ y
2. si $\sigma, \tau \in K$, entonces $\sigma \cap \tau = \emptyset$ o bien es una cara común de σ y τ .

La figura C.5 muestra gráficamente los conceptos de realización topológica y geométrica de un complejo simplicial. Concretamente, la figura ilustra la

generación de una triangulación a partir de un complejo simplicial abstracto

$$K^1 = \{1, 2, 3\} \cup \{\{\sigma_i^0\} \cup \{\sigma_j^1\}\}$$

con $i = 1, 2, 3; j = 1, 2$

tal que

$$\sigma_i^0 : \sigma_1^0 = \{1\}, \sigma_2^0 = \{2\}, \sigma_3^0 = \{3\}$$

$$\sigma_j^1 : \sigma_1^1 = \{1, 2\}, \sigma_2^1 = \{2, 3\}$$

esto es, dos aristas conectadas.

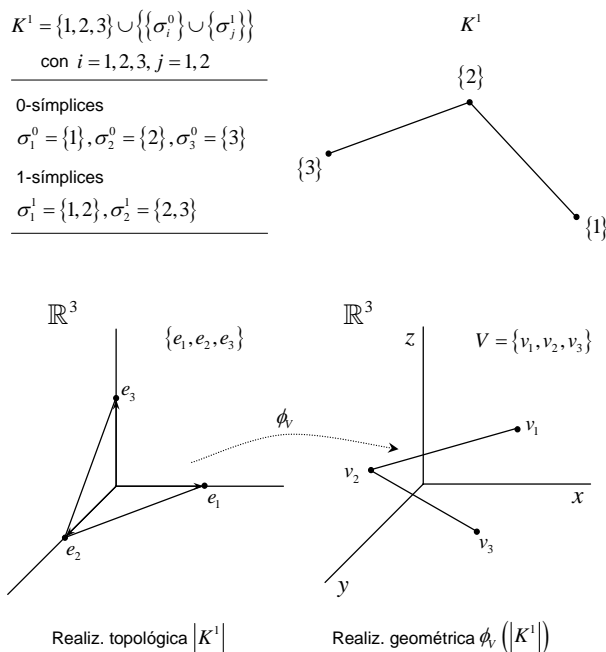


Figura C.5: Realización topológica y geométrica de un complejo simplicial.

En primer lugar se procede a identificar los vértices de K mediante los vectores base estándar $\{e_1, e_2, e_3\}, e_j \in \mathbb{R}^3$, lo que equivale a unir los simplices abiertos de K (esto es, la realización topológica), en segundo lugar, se construye una aplicación canónica ϕ_V que mapea los vectores base estándar de \mathbb{R}^3 al conjunto $V = \{v_1, v_2, v_3\}, v_j \in \mathbb{R}^3$, de posiciones de los vértices (esto es, la realización geométrica). Como se observa en la ilustración, esta

aplicación es inyectiva uno-a-uno (esto es, ϕ_V es un *embedding*) y por tanto, la imagen $\phi_V(|K^1|)$ es un complejo simplicial.

Nota: No confundir simplex abierto con simplex estándar, mientras que $|\sigma| \subset \mathbb{R}^m$, $\Delta^n \subset \mathbb{R}^{n+1}$ (ver fig. C.6).

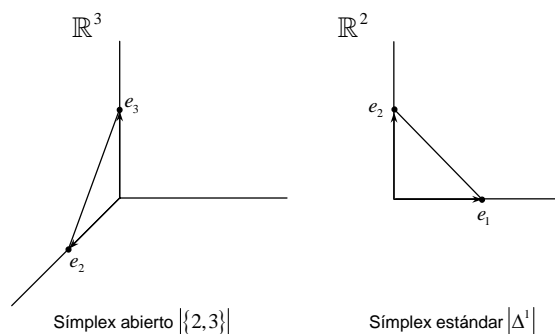


Figura C.6: Simplex abierto $|\sigma| \subset \mathbb{R}^3$ vs. Simplex estándar $\Delta^n \subset \mathbb{R}^2$.

C.3.1 Invariantes topológicos

Un complejo simplicial tiene ciertos atributos que son invariantes topológicos, así por ejemplo, todos los poliedros topológicamente equivalentes (esto es, existe un homeomorfismo entre ambos complejos) poseen los mismos invariantes.

El primero de ellos se conoce con el nombre de *primer vector de estructura* del complejo y permite tener una visión local del complejo, recurriendo al concepto de q -conectividad. Este concepto y sus aplicaciones en ciencias sociales fueron desarrolladas por el matemático Ronald Atkin en [14].

Dados dos símlices de un complejo K , σ_p y σ_r :

- Decimos que σ_p y σ_r son “ q -adyacentes” si existe al menos una cara común entre ellos que es un q -símplex. Obviamente, si σ_p y σ_r son q -adyacentes entonces son $(q-1)$ -, $(q-2)$ -, ..., 1-, y 0-adyacentes. Sea δ_q la relación definida como “es q -adyacente con”. Dicha relación es reflexiva y simétrica pero no transitiva.
- Decimos que σ_p y σ_r están “ q -conectados”:

1. si σ_p y σ_r son q -adyacentes o
2. si existe una secuencia de sımplices de K , $\sigma_1, \dots, \sigma_n$ tal que σ_p es q -adyacente con σ_1 , σ_1 es q -adyacente con σ_2 , \dots , σ_n es q -adyacente con σ_r .

Si σ_p y σ_r son q -conectados entonces ellos tambiın son $(q - 1), \dots, 1, 0$ -conectados en K . Si se define la relaciın γ_q como “es q -conectado con” entonces γ_q es una relaciın de equivalencia sobre los sımplices de K .

El segundo invariante topologico permite tener una visiın global del complejo, calculando los “agujeros” y “tuneles” que posee el poliedro en distintas dimensiones.

De esta manera a un complejo simplicial K^n tiene asociado un conjunto de valores $\beta = \langle \beta_0, \dots, \beta_n \rangle$, donde β_i es el correspondiente numero de Betti para la dimensiın i , con $i = 1, \dots, n$ [55]. β_0 es el numero de componentes arco-conexas, β_1 es el numero de tuneles linealmente independientes (esto es, 2-agujeros), β_2 es el numero de agujeros de dimensiın 3 (esto es, 3-agujeros) del complejo; en general, β_i es el numero de agujeros de dimensiın $(i + 1)$ (esto es, $(i + 1)$ -agujeros) del complejo.

Para terminar, en topologıa algebraica, cualquier complejo simplicial K se caracteriza topologicamente por el numero de agujeros, tuneles, ası como el numero de componentes conexas. Se define, por tanto, la *caracterıstica* o *numero de Euler* de K , como:

$$\chi(K) = (\# \text{ componentes}) - (\# \text{ tuneles}) + (\# \text{ agujeros}) \quad (\text{C.4})$$

Para el caso particular de las redes triangulares se verifica que:

$$\chi(K) = (\# \text{ triangulos}) - (\# \text{ aristas}) + (\# \text{ vertices}) = 2 \quad (\text{C.5})$$

C.3.2 Representaciın

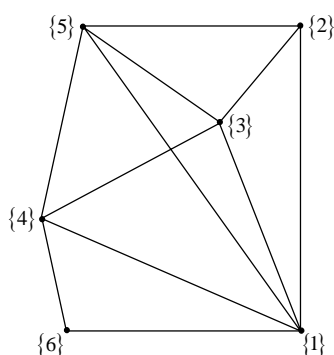
El procedimiento utilizado, tıpicamente, para representar un complejo simplicial, ası como sus invariantes topologicos, son las *matrices de incidencia*,

conexión, etc. Este sistema de representación facilita, desde un punto de vista operativo, el manejo y actualización de su estructura. Para ilustrar su utilización considérese el complejo simplicial abstracto

$$K^1 = \{1, \dots, 6\} \cup \{\{\sigma_i^0\} \cup \{\sigma_j^1\} \cup \{\sigma_k^2\}\}$$

con $i = 1, \dots, 6; j = 1, \dots, 11; k = 1, \dots, 7$

representado en la siguiente figura



Para construir la *matriz de incidencia* de K se genera una tabla de dimensión $N \times M$, donde las $N = (\text{card}(\{\sigma_j^1\}) + \text{card}(\{\sigma_k^2\}))$ representan los 1- y 2- símplices; y las $M = \text{card}(\{\sigma_i^0\})$ columnas los vértices, tal como se ilustra en la tabla C.1. En esta matriz, un 1 representa la pertenencia de cada vértice a cada cara (arista o triángulo) del complejo, y 0 lo contrario.

Los invariantes topológicos se pueden representar de forma análoga a la descrita para el complejo simplicial, así por ejemplo, si se desea representar la *matriz de 1-conexión* del complejo simplicial K , γ_1 , se procede como sigue: se genera una matriz cuadrada de dimensión $N \times N$, donde $N = \text{card}(\{\sigma_j^1\}) + \text{card}(\{\sigma_k^2\})$ representan los 1- y 2-símplices 1-conectados, como se ilustra en la tabla C.2 (por claridad, sólo se incluyen un subconjunto de símplices). En esta matriz, un 1 indica que ambos símplices están 1-conectados y 0 lo contrario. Recuérdese que: 1) si dos símplices están 1-conectados también están 0-conectados; 2) un símplice está 1-conectado consigo mismo.

Este tipo de estructuras facilita el acceso a los distintos símplices del complejo, su actualización, y el recorrido de éstos para su procesamiento.

Tabla C.1: Matriz de incidencia de un complejo simplicial.

	1	2	3	4	5	6
σ_1^1	1	1	0	0	0	0
σ_2^1	1	0	1	0	0	0
σ_3^1	1	0	0	1	0	0
σ_4^1	1	0	0	0	1	0
σ_5^1	1	0	0	0	0	1
σ_6^1	0	1	1	0	0	0
σ_7^1	0	1	0	0	1	0
σ_8^1	0	0	1	1	0	0
σ_9^1	0	0	1	0	1	0
σ_{10}^1	0	0	0	1	1	0
σ_{11}^1	0	0	0	1	0	1
σ_1^2	1	1	1	0	0	0
σ_2^2	1	0	1	1	0	0
σ_3^2	1	0	1	0	1	0
σ_4^2	1	0	0	1	1	0
σ_5^2	1	0	0	1	0	1
σ_6^2	0	1	1	0	1	0
σ_7^2	0	0	1	1	1	0

Tabla C.2: Matriz de 1-conexión (γ_1) de un complejo simplicial.

	σ_1^1	σ_1^2	σ_2^2	σ_3^2	σ_4^2
σ_1^1	1	0	0	0	0
σ_1^2	0	1	1	1	0
σ_2^2	0	1	1	1	0
σ_3^2	0	1	1	1	0
σ_4^2	0	0	0	0	1

Apéndice D

Detectores de esquinas: Harris y Lowe

En este apéndice se describen detalladamente los detectores de Harris y Lowe, dos técnicas ampliamente utilizadas en visión por computador y que se combinan en esta tesis para detectar y emparejar de forma robusta puntos de interés. Estas técnicas se caracterizan fundamentalmente por lo siguiente:

- El detector de Harris, posee un coste computacional bajo, pero la estabilidad de sus esquinas se ve muy afectada por los cambios en la escala.
- El detector de Lowe, por el contrario, tiene un coste computacional significativamente más alto (especialmente, la etapa de búsqueda de extremos) pero proporciona esquinas invariantes a la escala y la rotación de la imagen, así como parcialmente invariantes a cambios en la iluminación. Para dotarlas de esta invarianza crea descriptores, denominados SIFT, basados en información del gradiente.

La detección de esquinas mediante Harris y su caracterización mediante descriptores SIFT, como se propone en la sección 4.5.1, proporciona un mecanismo rápido y eficaz para detectar pares de correspondencias en imágenes captadas desde distintos puntos de observación y en condiciones de iluminación diferentes.

D.1 El detector de Harris

El detector bordes y esquinas propuesto por Harris en 1988 es simple, inmune al ruido y eficaz, lo que lo hace especialmente popular. Basado en el detector de Moravec [73], este detector se basa en la función de auto-correlación local de la imagen, la cual mide los cambios locales de la imagen en diferentes direcciones.

Dado un desplazamiento $(\Delta x, \Delta y)$ y un punto (x, y) , la función de auto-correlación c se define como:

$$c(x, y) = \sum_W [I(x_i, y_i) - I(x_i + \Delta x, y_i + \Delta y)]^2 \quad (\text{D.1})$$

donde $I(x, y)$ representa la imagen y (x_i, y_i) es el punto en la ventana gaussiana W centrado en (x, y) (por claridad el factor $e^{-(x^2+y^2)/(2\sigma^2)}$ de la ventana gaussiana se ha omitido en este desarrollo).

Si se aproxima $c(x, y)$ truncando su expansión en series de Taylor en el término de primer orden, se obtiene:

$$I(x_i + \Delta x, y_i + \Delta y) \approx I(x_i, y_i) + \begin{bmatrix} I_x(x_i, y_i) & I_y(x_i, y_i) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \quad (\text{D.2})$$

donde $I_x(x_i, y_i)$ e $I_y(x_i, y_i)$ son las derivadas parciales respecto a x e y , respectivamente.

Sustituyendo la expresión (D.2) en (D.1) se obtiene:

$$c(x, y) = \begin{bmatrix} \Delta x & \Delta y \end{bmatrix} \begin{bmatrix} \sum_W (I_x(x_i, y_i))^2 & \sum_W I_x(x_i, y_i) I_y(x_i, y_i) \\ \sum_W I_x(x_i, y_i) I_y(x_i, y_i) & \sum_W (I_y(x_i, y_i))^2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \quad (\text{D.3})$$

Finalmente, se denota como matriz de auto-correlación C a la expresión:

$$C(x, y) = \begin{bmatrix} \sum_W (I_x(x_i, y_i))^2 & \sum_W I_x(x_i, y_i) I_y(x_i, y_i) \\ \sum_W I_x(x_i, y_i) I_y(x_i, y_i) & \sum_W (I_y(x_i, y_i))^2 \end{bmatrix} \quad (D.4)$$

Nótese que la matriz $C(x, y)$ contiene información relativa a la variación de intensidad en torno al pixel (x, y) . De la correcta interpretación de los autovalores de C , se puede determinar si (x, y) es una esquina (λ_1 y λ_2 grandes), forma parte de un borde ($\lambda_1 > \lambda_2$, o viceversa), o de una zona homogénea (λ_1 y λ_2 aproximadamente 0). Como se puede deducir de (D.4), la información proporcionada por los autovalores es invariante a la rotación, pero no al cambio de escala.

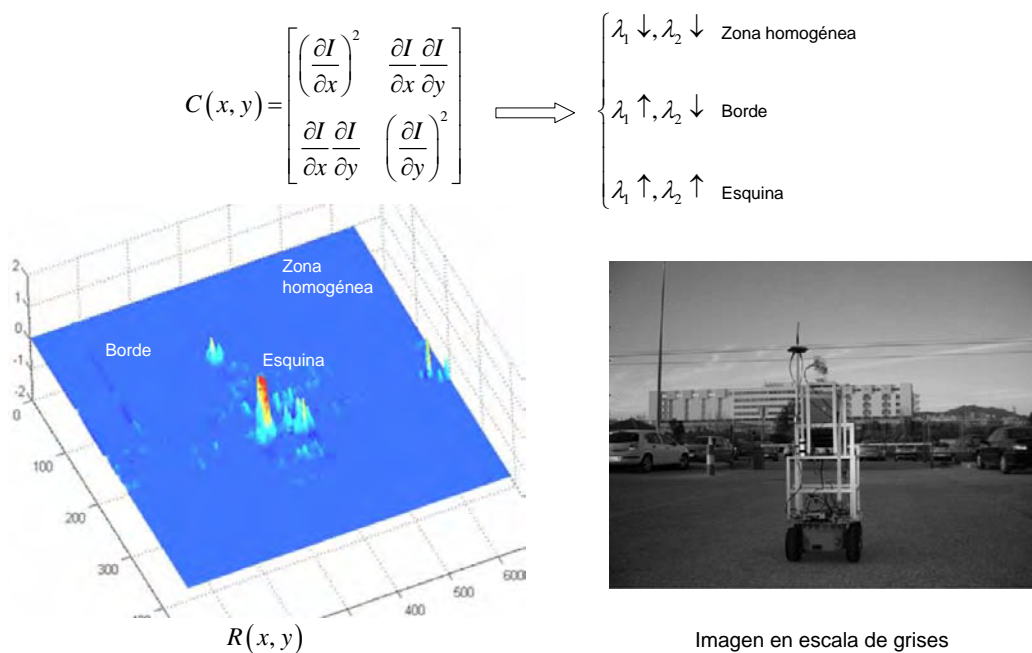


Figura D.1: Proceso de detección de esquinas propuesto por Harris.

Harris propone una función para detectar esquinas sin tener que obtener explícitamente los autovalores de $C(x, y)$. La función de detección viene dada

por la siguiente expresión:

$$R = \det(C) - k(\text{trace}(C))^2 \quad (\text{D.5})$$

donde k es un parámetro igual a 0.04 (por sugerencia del autor).

Las esquinas son definidas como máximos locales de la función de detección R (ver fig. D.1). La precisión subpíxel se logra a través de una aproximación cuadrática del entorno del máximo local.

En la práctica, el número de esquinas proporcionado por este operador suele ser considerable. En este caso suele ser habitual que antes de iniciar el proceso de emparejamiento, limitar el número de esquinas seleccionadas mediante zonas de supresión de máximos, así como también, su distribución. Una estrategia ampliamente utilizada consiste en seleccionar n esquinas “fuertes” (esto es, aquellas que presentan valores de R significativos) uniformemente distribuidas o asegurar una distancia mínima entre ellas mediante supresión local de máximos [100].

D.2 El detector de Lowe (SIFT)

En [63], Lowe propone un método para detectar esquinas invariantes a la escala y la rotación de la imagen, así como parcialmente invariantes a cambios en la iluminación y otras transformaciones afines. Para dotarlo de robustez, Lowe asigna a cada esquina un descriptor basado en el histograma de orientaciones de su entorno, denominado *Scale-Invariant Feature Transform* (SIFT). Un identificador altamente distintivo que facilita su posterior emparejamiento. A continuación se detallan las fases de la que consta esta técnica.

En primer lugar se buscan puntos estables a lo largo de diferentes escalas de la imagen para conseguir la invarianza de las características a los cambios de escala. Para lograr esto Lowe propone una búsqueda de características a lo largo de todas las posibles escalas, utilizando para ello, una función continua de la escala conocida como *espacio de escalas*. El espacio de escalas de una imagen está definido por la función $L(x, y, \sigma)$, que se obtiene de la convolución de una variable de escala Gaussiana $G(x, y, \sigma)$ con la imagen de

entrada $I(x, y)$.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (\text{D.6})$$

donde $*$ es la operación de convolución en x e y , y

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (\text{D.7})$$

Para detectar eficientemente características estables en el espacio de escalas, Lowe propone localizar extremos (máximos y mínimos locales) en la función DoG (Diferencia de Gaussianas) convolucionada con la imagen $D(x, y, \sigma)$, que se obtiene de la diferencia de dos escalas muy próximas separadas por un factor multiplicador k :

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (\text{D.8})$$

donde k es un parámetro igual a $\sqrt{2}$ (por sugerencia del autor).

La figura D.2 muestra gráficamente el proceso de construcción del espacio de escalas. La imagen inicial es convolucionada incrementalmente con Gaussianas para producir imágenes separadas por k en el espacio de escalas. Si se divide cada octava del espacio de escalas en un número entero de intervalos s , tal que $k = 2^{1/s}$, es necesario generar $s + 3$ imágenes en la pila de imágenes suavizadas para cada octava, hasta que la detección final de extremos abarque una octava completa. A continuación, las imágenes de escalas adyacentes son sustraídas para generar la DoG. Una vez que una octava completa ha sido procesada, se reduce a la mitad la imagen Gaussiana que tiene dos veces el valor inicial de σ y se repite el proceso.

A continuación, para detectar los extremos en la imagen $D(x, y, \sigma)$, cada punto candidato es comparado con los ocho vecinos de su misma escala, con los nueve vecinos en la escala superior y con los nueve vecinos de la escala inferior. De tal forma que un punto es seleccionado si y sólo si, es más grande

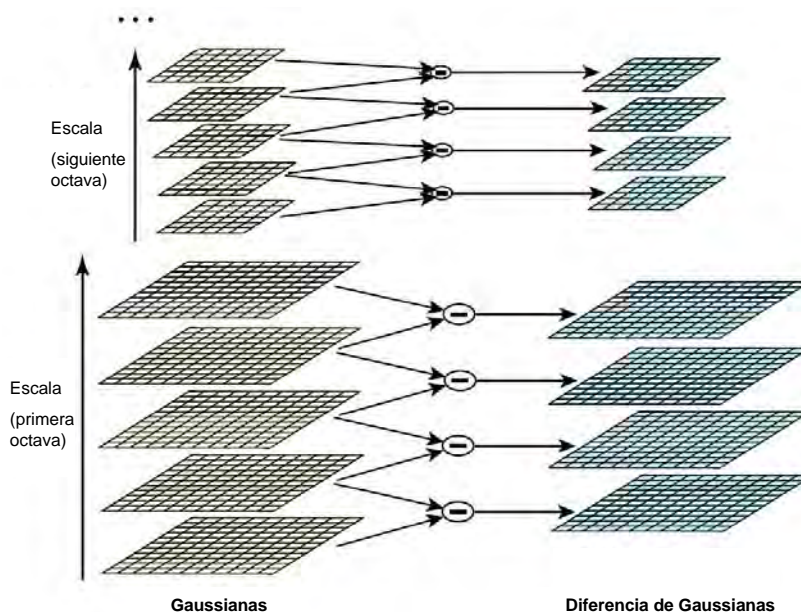


Figura D.2: Proceso de construcción del espacio de escalas.

o más pequeño que todos sus vecinos. La figura D.3 ilustra el proceso de selección.

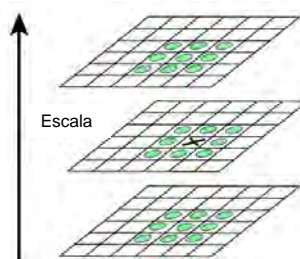


Figura D.3: Proceso de búsqueda de extremos en las octavas construidas.

Una vez finalizado el proceso de detección, se crea un descriptor por cada característica localizada. Para construir el descriptor se procede del siguiente modo: en primer lugar, se calcula la magnitud del gradiente y la orientación en cada punto de su vecindario (fig. D.4(a)); en segundo lugar, se ponderan con una ventana Gaussiana the radio σ igual a la mitad de la anchura del descriptor (la ventana de ponderación se representa mediante una circunferencia); y en tercer lugar, las muestras ya ponderadas son acumuladas

en histogramas de orientación (fig. D.4(b)). El proceso de acumulación se realiza por subregiones de tal forma que la magnitud de cada entrada del histograma es proporcional a la suma de las magnitudes del gradiente en dicha orientación.

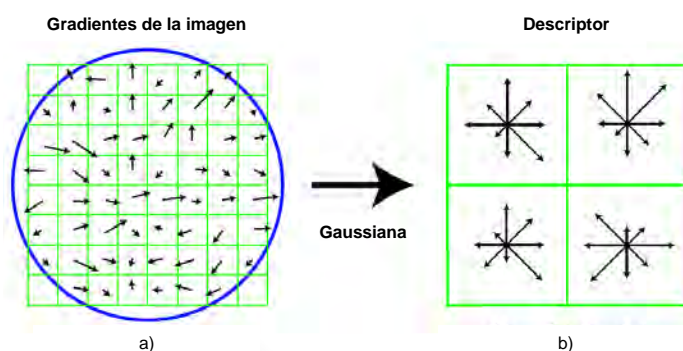


Figura D.4: Descriptor de longitud $2 \times 2 \times 8 = 32$ obtenido a partir de regiones de 8×8 muestras.

La dimensión del descriptor viene dada por la expresión $N \times N \times B$, donde B es el número de orientaciones consideradas en los histogramas (esto es, el número de *bins*). La figura D.4 muestra un vecindario de dimensión 8×8 del que se obtienen 2×2 histogramas con 8 *bins* cada uno, produciendo descriptores de dimensión $2 \times 2 \times 8 = 32$.

Finalmente, el vector descriptor es normalizado con objeto de reducir la influencia de los cambios de iluminación (contraste, brillo y diferencias no lineales) en el descriptor. El lector puede encontrar en [63] más detalles sobre el proceso de normalización y los valores aconsejados para cada uno de los parámetros.

Bibliografía

- [1] G. Ambrosio, J. González, and V. Arévalo. Detección de Cambios Urbanos Mediante Comparación de Imágenes Multitemporales. In *X Congreso de Métodos Cuantitativos, Sistemas de Información Geográfica y Teledetección*, pages CD-ROM, Valladolid, Spain, Sep 2002.
- [2] V. Arévalo and J. Gonzalez. An Experimental Evaluation of Non-Rigid Registration Techniques on QuickBird Satellite Imagery. *International Journal of Remote Sensing*, In press, 2007.
- [3] V. Arévalo, J. Gonzalez, and G. Ambrosio. Shadow Detection in Colour High-Resolution Satellite Images. *International Journal of Remote Sensing*, In press, 2007.
- [4] V. Arévalo and J. González. A Comparison of Registration Techniques on QuickBird Satellite Images. In *ISPRS Mid-term Symposium 2006. Remote Sensing: From Pixels to Processes*, pages CD-ROM, Enschede, The Netherlands, May 2006.
- [5] V. Arévalo and J. González. Improving Piecewise-Linear Registration of High-Resolution satellite Images through Mesh Optimization. *IEEE Transactions on Geoscience and Remote Sensing*, 2007.
- [6] V. Arévalo and J. González. Improving Piecewise-Linear Registration through Mesh Optimization. *3rd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'07), Part II*, 4478:122–129, 2007.
- [7] V. Arévalo, J. González, and G. Ambrosio. Corrección Geométrica de Imágenes de Satélite de Distinta Resolución. In *XII Congreso de Méto-*

- dos Cuantitativos, Sistemas de Información Geográfica y Teledetección*, pages 247–260, Murcia, Spain, September 2004.
- [8] V. Arévalo, J. González, and G. Ambrosio. La Librería de Visión Artificial OpenCV. Aplicación a la Docencia e Investigación. In *The Open Source International Conference*, pages CD-ROM, Málaga, Spain, February 2004.
- [9] V. Arévalo, J. González, and G. Ambrosio. La Librería de Visión Artificial OpenCV. Aplicación a la Docencia e Investigación. *Base Informática*, 40:61–66, 2004.
- [10] V. Arévalo, J. González, and G. Ambrosio. OpenCV. La Librería Open Source de Visión Artificial. *Linux Free Magazine*, 10:141–147, 2005.
- [11] V. Arévalo, J. González, and G. Ambrosio. Registro Elástico de Imágenes de Satélite de Alta Resolución Mediante Optimización de Redes Triangulares. In *XXVI Jornadas de Automática*, pages 1135–1141, Alicante, Spain, September 2005.
- [12] V. Arévalo, J. González, J. Valdés, and G. Ambrosio. Detecting Shadows in QuickBird Satellite Images. In *ISPRS Mid-term Symposium 2006. Remote Sensing: From Pixels to Processes*, pages CD-ROM, Enschede, The Netherlands, May 2006.
- [13] J. Astola and I. Virtanen. *Entropy Correlation Coefficient, a Measure of Statistical Dependence for Categorized Data*. Lappeenranta teknillinen korkeakoulu, first edition, 1981.
- [14] R.H. Atkin. From Cohomology in Physics to q-connectivity in Social Science. *International Journal of Man-Machine Studies*, 4:139–167, 1972.
- [15] S. Battiato, S. Nicotra, and A. Palazzo. Self Extracting SVG Rendering of Compressed Raster Images Vectorized by DDT Triangulation. In *International Conference on Scalable Vector Graphics*, volume 1, pages CD-ROM, Tokyo, Japan, 2007.

- [16] A. Baumberg. Reliable Feature Matching Across Widely Separated Views. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'00)*, volume 1, pages 774–781. Hilton Head, South Carolina, 2000.
- [17] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded Up Robust Features. In *European Conference on Computer Vision (ECCV'06)*, volume 13, pages 404–417, 2006.
- [18] R.K. Beatson, W.A. Light, and S.D. Billings. Domain Decomposition Methods for Solution of the Radial Basis Function Interpolation Equations. *SIAM Journal on Scientific Computing*, 22:1717–1740, 2001.
- [19] F. Bookstein. Principal Warps: Thin-Plate-Splines and the Decomposition of Deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.
- [20] L. G. Brown. A Survey of Image Registration Techniques. *ACM Computing Surveys*, 24(4):325–376, 1992.
- [21] M. Brown and D.G. Lowe. Invariant Features from Interest Point Groups. In *British Conference on Machine Vision (BMVC'02)*, volume 1, pages 656–665, Cardiff, Wales, 2002.
- [22] J. F. Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.
- [23] H.M. Chen, P.K. Varshney, and M.K. Arora. Performance of Mutual Information Similarity Measure for Registration of Multitemporal Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 41(11):2445–2454, 2003.
- [24] A.A. Cole-Rhodes, K.L. Johnson, J. LeMoigne, and I. Zavorin. Multi-resolution Registration of Remote Sensing Imagery by Optimization of Mutual Information using a Stochastic Gradient. *IEEE Transactions on Image Processing*, 12(12):1495–1511, 2003.

- [25] T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press, second edition, 2001.
- [26] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, NY, USA, first edition, 1991.
- [27] P.W. de Bruin, P.M. van Meeteren, F.M. Vos, A.M. Vossepoel, and F.H. Post. Accurate and High Quality Triangle Models from 3D Grey Scale Images. *Medical Image Computing and Computer-Assisted Intervention (MICCAI'02), Part II*, 2489:348–355, 2002.
- [28] N. Dyn, D. Levin, and S. Rippa. Data Dependent Triangulations for Piecewise Linear Interpolation. *IMA Journal of Numerical Analysis*, 10:137–154, 1990.
- [29] M. Ehlers and D.N. Fogel. High-Precision Geometric Correction of Airborne Remote Sensing Revisited: The Multiquadric Interpolation. In *SPIE: Image and Signal Processing for Remote Sensing*, volume 2315, pages 814–824, September 1994.
- [30] F.M. Estrada, F. Yamazaki, and M. Matsuoka. Use of LANDSAT images for the identification of damage due to the 1999 Kocaeli, Turkey Earthquake. In *Asian Conference on Remote Sensing*, Taipei, Taiwan, 2000.
- [31] M.A. Fischler and R.C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography. *Communications of the ACM*, 24:381–395, 1981.
- [32] J. Folkesson and H. Christensen. SIFT Based Graphical SLAM on a Packbot. In *The International Conference on Field and Service Robots (FSR'07)*, Chamonix, France, July 2007.
- [33] M. Fornefett, K. Rohr, and H.S. Stiehl. Elastic Registration of Medical Images Using Radial Basis Functions with Compact Support. In *IEEE International Conference on Computer Vision and Pattern Recognition*

- (*CVPR'99*), volume 1, page 1402, Los Alamitos, CA, USA, 1999. IEEE Computer Society.
- [34] K.J. Friston, J. Ashburner, J.B. Poline, C.D. Frith, J.D. Heather, and R. Frackowiak. Spatial Registration and Normalisation of Images. *Human Brain Mapping*, 2:165–189, 1995.
- [35] W. Förstner and E. Gülch. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features. In *ISPRS Workshop on Fast Processing of Photogrammetric Data*, volume 1, pages 281–305. Interlaken, Switzerland, September 1986.
- [36] J.M. Geusebroek, G.J. Burghouts, and A.W.M. Smeulders. The Amsterdam Library of Object Images. *International Journal of Computer Vision*, 61(1):103–112, 2005.
- [37] J. González, G. Ambrosio, and V. Arévalo. Automatic Urban Change Detection from the IRS-1D PAN. In *IEEE-ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, pages 320–323, Rome, Italy, November 2001.
- [38] J. González, C. Galindo, V. Arévalo, and G. Ambrosio. Applying Image Analysis and Probabilistic Techniques for Counting Olive Trees in High-Resolution Satellite Images. *Advanced Concepts for Intelligent Vision Systems (ACIVS'07)*, 4678:920–931, 2007.
- [39] A. Goshtasby. Piecewise Linear Mapping Functions for Image Registration. *Pattern Recognition*, 19(6):459–466, 1986.
- [40] A. Goshtasby. Piecewise Cubic Mapping Functions for Image Registration. *Pattern Recognition*, 20(5):525–533, 1987.
- [41] A. Goshtasby. Registration of Images with Geometric Distortions. *IEEE Transactions on Geoscience and Remote Sensing*, 26:60–64, 1988.
- [42] A. Goshtasby. *2D and 3D Image Registration*. Wiley-Interscience, New York, NY, USA, first edition, 2005.

- [43] J.B. Hajnal, N. Saeed, A. Oatridge, E.J. Williams, I.R. Young, and G.M. Bydder. Detection of Subtle Brain Changes using Subvoxel Registration and Subtraction of Serial MR Images. *Journal of Computer Assisted Tomography*, 19:677–91, 1995.
- [44] J.B. Hajnal, N. Saeed, E.J. Soar, A. Oatridge, I.R. Young, and G.M. Bydder. A Registration and Interpolation Procedure for Subvoxel Matching of Serially Acquired MR Images. *Journal of Computer Assisted Tomography*, 19:289–96, 1995.
- [45] R.L. Harder and R.N. Desmarais. Interpolation using Surface Splines. *Journal of Aircraft*, 9:189–191, 1972.
- [46] C.J. Harris and M. Stephens. A Combined Corner and Edge Detector. In *Alvey Vision Conference*, volume 4, pages 147–151, Manchester, UK, 1988.
- [47] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, second edition, 2004.
- [48] G. Hermosillo, C. Chéfd’Hotel, and O. Faugeras. Variational Methods for Multimodal Image Matching. *International Journal of Computer Vision*, 50(3):329–343, 2002.
- [49] H. Hoppe. Progressive Meshes. *Computer Graphics*, 30(Annual Conference Series):99–108, 1996.
- [50] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle. Mesh Optimization. *Computer Graphics*, 27(Annual Conference Series):19–26, 1993.
- [51] Image Fusion Systems Research. Transformation Functions for Image Registration. Technical report, IFSR Technical Report, 2003.
- [52] J. Inglada and A. Giros. On the Possibility of Automatic Multisensor Image Registration. *IEEE Transactions on Geoscience and Remote Sensing*, 42(10):2104–2120, 2004.

- [53] Intel® Integrated Performance Primitives. <http://www.intel.com/cd/software/products/asm-na/eng/302910.htm>.
- [54] ITT Visual Information Solutions. <http://www.ittvis.com/envi>.
- [55] J.R. James. *Elements of Algebraic Topology*. Addison-Wesley, first edition, 1984.
- [56] Y. Ke and R. Sukthankar. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'04)*, volume 2, pages 506–513, Los Alamitos, CA, USA, 2004.
- [57] J.P. Kern and M.S. Pattichis. Robust Multispectral Image Registration Using Mutual-Information Models. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5):1494–1505, 2007.
- [58] L. Kitchen and A. Rosenfeld. Gray-level Corner Detection. *Pattern Recognition Letters*, 1:95–102, 1982.
- [59] J. Kybic and M. Unser. Fast Parametric Elastic Image Registration. *IEEE Transactions on Image Processing*, 12(11):1427–1442, 2003.
- [60] B. Lehner, G. Umlauf, and B. Hamann. Image Compression using Data-Dependent Triangulations. *3rd International Symposium on Visual Computing (ISVC'07), Part I*, 4841:351–362, 2007.
- [61] Leica Geosystems. <http://gi.leica-geosystems.com>.
- [62] L. Lemieux and G.J. Barker. Measurement of Small Inter-Scan Fluctuations in Voxel Dimensions in Magnetic Resonance Images using Registration. *Medical Physics*, 25:1049–1054, 1998.
- [63] D.G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 20:91–110, 2003.
- [64] J.M. López, J. González, C. Galindo, and V. Arévalo. Reconocimiento Automático de Matrículas usando Operadores Morfológicos y Múltiples

- Redes Neuronales. In *XXVII Jornadas de Automática*, pages 627–632, Almería, Spain, September 2006.
- [65] B.D. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.
- [66] W. Madych. Miscellaneous Error Bounds for Multi-quadric and Related Interpolators. *Computers Math. Applic.*, 24:121–138, 1992.
- [67] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality Image Registration by Maximization of Mutual Information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.
- [68] F. Maes, D. Vandermeulen, and P. Suetens. Comparative Evaluation of Multiresolution Optimization Strategies for Multimodality Image Registration by Maximization of Mutual Information. *Medical Image Analysis*, 3(4):373–386, 1999.
- [69] B.J. Matuszewski, J.K. Shen, and L.K. Shark. Elastic Image Matching with Embedded Rigid Structures using Spring-Mass System. In *IEEE International Conference on Image Processing*, volume 2, pages 937–940, September 2003.
- [70] E.M. Mikhail, J.S. Bethel, and J.C. McGlone. *Modern Photogrammetry*. John Wiley & Sons, New York, NY, USA, first edition, 2001.
- [71] K. Mikolajczyk and C. Schmid. An Affine Invariant Interest Point Detector. In *European Conference on Computer Vision (ECCV'02)*, volume 1, pages 128–142, Copenhagen, Denmark, 2002.
- [72] J.M.M. Montiel and D. Ortín. Indoor Robot Motion Based on Monocular Images. *Robotica*, 19:331–342, 2001.
- [73] H. Moravec. Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover. Technical Report CMU-RI-TR-3, Carnegie-Mellon University, Robotics Institute, 1980.

- [74] P. Moreels and P. Perona. Evaluation of Features Detectors and Descriptors Based on 3D Objects. In *IEEE International Conference on Computer Vision (ICCV'05)*, volume 1, pages 800–807, Los Alamitos, CA, USA, October 2005. IEEE Computer Society.
- [75] F.A. Moreno, J.L. Blanco, and J. Gonzalez. An Efficient Closed-form Solution to Probabilistic 6D Visual Odometry for a Stereo Camera. *Advanced Concepts for Intelligent Vision Systems (ACIVS'07)*, 4678:932–942, 2007.
- [76] F.A. Moreno, J.L. Blanco, J. González, V. Arévalo, and J.A. Fernández-Madrigal. Un Enfoque Probabilístico para la Estimación del Desplazamiento de un Robot Móvil Dotado de Visión Estéreo. In *XXVII Jornadas de Automática*, pages 1439–1446, Almería, Spain, September 2006.
- [77] D.D. Morris and T. Kanade. Image-Consistent Surface Triangulation. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'00)*, pages 332–338, Hilton Head, SC, USA, June 2000.
- [78] A.Ñakatuji, Y. Sugaya, and K. Kanatani. Mesh Optimization using an Inconsistency Detection Template. In *IEEE International Conference on Computer Vision (ICCV'05)*, volume 2, pages 1148–1153, Beijing, China, October 2005.
- [79] National Library of Medicine Insight Segmentation and Registration Toolkit. <http://www.itk.org>.
- [80] K.Ñovak. Rectification of Digital Imagery. *Photogrammetric Engineering and Remote Sensing*, 58(3):339–344, 1992.
- [81] OpenVIDIA : Parallel GPU Computer Vision. <http://openvidia.sourceforge.net/>.

- [82] J.S. Perrier, G. Agin, and P. Cohen. Image-Based View Synthesis for Enhanced Perception in Teleoperation. In *SPIE: Enhanced and Synthetic Vision*, volume 4023, pages 332–338, April 2000.
- [83] Peter Kovesei's Functions for Computer Vision. <http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/index.html>.
- [84] RapidMind®. <http://www.rapidmind.net/>.
- [85] K. Rohr, H.S Stiehl, R. Sprengel, W. Beil, T.M. Buzug, J. Weese, and M.H. Kuhn. Point-Based Elastic Registration of Medical Image Data Using Approximating Thin-Plate Splines. In *International Conference on Visualization in Biomedical Computing (VBC'96)*, Hamburg, Germany, Sep 1996. Springer.
- [86] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. Wiley-Interscience, New York, NY, USA, first edition, 2005.
- [87] L.E.M. Salazar. Una Aproximación Topológica al Diseño Modular en Ingeniería de Software. *Sistemas & Telemática. Revista de la Facultad de Ingeniería*, 2003.
- [88] F. Schaffalitzky and A. Zisserman. Multi-View Matching for Unordered Image Sets, or 'How do I Organize my Holiday Snaps? In *European Conference on Computer Vision (ECCV'02)*, volume 1, pages 414–431, Copenhagen, Denmark, May 2002.
- [89] D. Schlesinger, B. Flach, and A. Shekhovtsov. A Higher Order MRF-Model for Stereo-Reconstruction. *Pattern Recognition*, 3175:440–446, 2004.
- [90] C. Schmid and R. Mohr. Local Gray-Value Invariants for Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–534, 1997.
- [91] L.L. Schumaker. Computing Optimal Triangulations using Simulated Annealing. *Computer Aided Geometric Design*, 10(3-4):329–345, 1993.

- [92] M. Servais, T. Vlachos, and T. Davies. Bi-Directional Affine Motion Compensation using a Content-Based, Non-Connected, Triangular Mesh. In *European Conference on Visual Media Production (CVMP'04)*, pages 49–58, London, UK, March 2004.
- [93] J.R. Shewchuk. Lecture Notes on Delaunay Mesh Generation. Technical Report 3, University of California at Berkeley, 1999.
- [94] J. Shi and C. Tomasi. Good Features to Track. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, volume 1, pages 593–600, Seattle, USA, June 1994.
- [95] SiftGPU: A GPU Implementation of David Lowe's Scale Invariant Feature Transform (SIFT). <http://cs.unc.edu/~ccwu/siftgpu/>.
- [96] G. Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, Wellesley, MA, USA, first edition, 1986.
- [97] C. Studholme, D.J. Hawkes, and D.L.G. Hill. An Overlap Invariant Entropy Measure of 3D Medical Image Alignment. *Pattern Recognition*, 32:71–86, 1999.
- [98] C. Studholme, D.L.G. Hill, and D.J. Hawkes. Multiresolution Voxel Similarity Measures for MRI-PET Registration. In *Information Processing in Medical Imaging*, pages 287–298, Ile de Berder, 1995. Kluwer Academic.
- [99] The Mathworks. <http://www.mathworks.com>.
- [100] The Open Source Computer Vision Library. <http://www.sourceforge.net/projects/opencvlibrary>.
- [101] P. Torr and A. Zisserman. Robust Computation and Parametrization of Multiple View Relations. In *IEEE International Conference on Computer Vision (ICCV'98)*, pages 727–732, 1998.

- [102] P.H.S. Torr. A Structure and Motion Toolkit in Matlab: Interactive Adventures in S and M. Technical Report MSR-TR-2002-56, Microsoft Research, 2002.
- [103] Z. Toth, I. Viola, A. Ferko, and E. Gröller. N-dimensional Data-Dependent Reconstruction Using Topological Changes. *Topology-based Methods in Visualization*, pages 183–198, 2007.
- [104] P. Viola and W.M. Wells. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision*, 24(2):137–154, 1997.
- [105] G. Vogiatzis, P. Torr, and R. Cipolla. Bayesian Stochastic Mesh Optimisation for 3D Reconstruction. In *British Conference on Machine Vision (BMVC'03)*, pages 711–718, Norwich, UK, September 2003.
- [106] R.P. Woods, S.R. Cherry, and J.C. Mazziotta. Rapid Automated Algorithm for Aligning and Reslicing PET Images. *Journal of Computer Assisted Tomography*, 16:620–633, 1992.
- [107] R.P. Woods, S.T. Grafton, C.J. Holmes, S.R. Cherry, and J.C. Mazziotta. Automated Image Registration: I. General Methods and Intrasubject, Intramodality Validation. *Journal of Computer Assisted Tomography*, 22:139–152, 1998.
- [108] R.P. Woods, J.C. Mazziotta, and S.R. Cherry. MRI-PET Registration with Automated Algorithm. *Journal of Computer Assisted Tomography*, 17(4):536–546, 1993.
- [109] Hua Xie, L.E. Pierce, and F.T. Ulaby. Mutual Information Based Registration of SAR Images. In *IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS'03)*, volume 6, pages 4028–4031, July 2003.
- [110] G. Xu and Z. Zhang. *Epipolar Geometry in Stereo, Motion, and Object Recognition: A Unified Approach*. Kluwer Academic Publishers, Norwell, MA, USA, first edition, 1996.

- [111] B. Yip and J.S. Jin. Image Registration Using Triangular Mesh. *Advances in Multimedia Information Processing*, 3331:298–303, 2004.
- [112] Xiaohua Yu, B.S. Bryan, and T.W. Sederberg. Image Reconstruction using Data-Dependent Triangulation. *Computer Graphics and Applications*, 21(3):62–68, 2001.
- [113] B. Zitová and J. Flusser. Image Registration Methods: A Survey. *Image and Vision Computing*, 21(11):977–1000, 2003.