

RECONSTRUCCIÓN 3D

Javier Finat Codes

10 de octubre de 2019

©2019MoBiVAP (Universidad de Valladolid) - www.mobivap.eu

Licenciado bajo Creative Commons No Comercial 3.0 (la "licencia"). Usted no debería utilizar este fichero si no está de acuerdo con los términos de la licencia. Puede obtener una copia de la licencia en <http://creativecommons.org/licenses/by-nc/3.0>. A menos que lo requiera la ley o de acuerdo con lo escrito, el software distribuido bajo la licencia se considera "TAL CUAL", SIN GARANTÍAS NO CONDICIONES DE NINGÚN TIPO, ni explícitas ni implícitas. Vea la licencia para consultar las limitaciones y permisos para cada idioma específico

Reconstrucción 3D. Una Introducción



Índice general

1 Reconstrucción 3D. Una Introducción	3
Índice general	3
1.1. Elementos básicos de Reconstrucción 3D	8
1.1.1. Elementos de modelado	8
1.1.2. Imágenes para la Reconstrucción 3D	11
1.1.3. Modelado semi-automático	14
1.1.4. Visión Estéreo	16
1.2. Estrategias de resolución para la Reconstrucción 3D	21
1.2.1. Una taxonomía vinculada al modelado	21
1.2.2. Hipótesis adicionales simplificadoras	22
1.2.3. Marco general para la unificación	25
1.3. Diseño de algoritmos	28
1.3.1. Modelado y resolución del problema óptico-geométrico	28
1.3.2. Algoritmos según tipo de agregación	28
1.3.3. Algoritmos de coste de pegado	28
1.3.4. Algoritmos de Optimización	28
1.3.5. Estimación y corrección de errores	29
1.4. Aspectos históricos	30
1.4.1. Reconstrucción 3D y Fotogrametría terrestre	30
1.4.2. Dos vistas	31
1.4.3. Una vista	31
1.4.4. Tres o más vistas	32

El módulo B_{32} de la materia B_3 (Visión Computacional) está enfocado hacia la Reconstrucción 3D a partir de una o varias vistas. Este tópico ha recibido una atención intensiva desde finales de los ochenta y sobre todo a lo largo de los noventa que culmina en la obra de R.Hartley y A.Zisserman publicado por primera vez en el año 2000. Este trabajo (citado en lo sucesivo como [Har00], a pesar de reediciones posteriores) proporciona un marco común a la pléyade de aproximaciones y resultados parciales existentes en la época. El carácter exhaustivo y la excelente organización de materiales contenidos en [Har00] hace difícil realizar aportaciones significativas en un curso de carácter introductorio como el actual.

Por ello, además de glosar algunos resultados contenidos en [Har00], el propósito adicional de este módulo es llamar la atención sobre aspectos que conectan con modelos y herramientas de otras materias de este repositorio. Estos modelos y herramientas están relacionadas con *aplicaciones de la Reconstrucción 3D* vinculadas a

- el *modelado avanzado 3D* incluyendo virtualización de escenarios arqueológicos, arquitectónicos o urbanos;
- la *renderización* incluyendo efectos vinculados a la iluminación y los mapas de reflectancia, incluyendo análisis y síntesis;
- las relaciones con *otros módulos del CEViC*. Entre estas últimas cabe destacar la utilidad para
 - el *movimiento* con especial atención a la reconstrucción a partir de vídeo (Structure/Shape from Motion);
 - el *reconocimiento de objetos y escenas* (sobre todo arquitectónicos), como base para una realimentación entre reconstrucción y reconocimiento de objetos/escenas);
 - el *Vídeo 3D* como Reconstrucción dinámica de objetos en movimiento (sobre todo personajes) orientado hacia la simplificación en tareas de post-producción digital.

Por ello y a pesar del espléndido trabajo realizado en [Har00], la Reconstrucción 3D está lejos de ser un área cerrada en relación con otras áreas. En particular, centramos la atención en aplicaciones vinculadas a visualización dinámica de entornos interactivos y asistencia a la producción de contenidos digitales. Desde finales de los noventa, la demanda creciente del sector de producción digital en áreas de ocio y entretenimiento es una fuente permanente de inspiración para la Reconstrucción 3D. Los avances recientes en el marco DeepLearning abren posibilidades antes insospechadas para la adaptación e interacción en entornos complejos.

El módulo B_{32} está organizado de acuerdo con el esquema siguiente:

1. Conceptos básicos. Modelos de perspectiva. Estimación de elementos estructurales.
2. Calibración. Geometría epipolar.
3. Visión Estéreo estática a partir de 2 vistas.
4. Reconstrucción a partir de múltiples vistas.
5. Reconstrucción a partir de una sola vista. Movimientos rígidos.
6. Reconstrucción a partir de video (Structure From Motion)

El *objetivo de la Reconstrucción 3D* es el modelado de objetos o de escenas a partir de la información (analógica o digital) proporcionada por cámaras o sensores de rango (infrarrojos, láser, rádar, LIDAR, rayos X, etc). Los inputs capturados por diferentes dispositivos condicionan a menudo el tipo de modelado o de representación.

- La percepción visual asociada a fotografías o planos es aparentemente continua y ello sugiere la utilización de modelos continuos. Sin embargo, el procesamiento automático de la información digital contenida en fotos dé lugar a toda clase de discontinuidades. Estamos interesados en las discontinuidades de la función de intensidad cuyo soporte está dado por líneas.
- Los inputs procedentes de los sensores de rango son discretos, por lo que la primera tarea del modelado para la Reconstrucción 3D consiste en asociar de forma automática modelos continuos a partir de nubes de puntos (con una densidad y una distribución muy variables).

La posibilidad de generar en tiempo real una reconstrucción 3D a partir de varias vistas está generando una demanda creciente de aplicaciones de altas prestaciones relacionadas con la navegación automática, la realidad virtual o herramientas de visualización avanzada que incluyan la posibilidad de consulta o la generación de informes a partir de la interacción con objetos 3D. L

Las altas prestaciones afectan a la calibración (precisión en las medidas de los objetos), a la visualización (calidad de las imágenes generadas de forma sintética para la representación 3D de la escena) y a la interacción (capacidad de respuesta en presencia de objetos reales o virtuales). Por ello, este área de la Visión Computacional se solapa con otras procedentes de la Fotogrametría, Informática Gráfica y Sistemas Expertos.

Si se atiende a los aspectos geométricos relativos a la escena, el problema fundamental es una fusión precisa y rápida de información 2D para la generación de modelos 3D a diferentes resoluciones. Para ello, la información 2D de cada vista se “reproyecta” (según estrategias coarse-to-fine) sobre modelos *continuos* asociados a una nube densa de puntos 3D.

En particular, la reproyección permite remapear no sólo elementos geométricos más finos, sino también propiedades radiométricas (incluyendo texturas complejas, p.e.) que, eventualmente, pueden ser diferentes de las originales. Esta estrategia es bien conocida por los artistas gráficos y los expertos en Inforgrafía, pero en su mayor parte se realiza de forma manual. De cara a aplicaciones inmersivas (de interés para cirugía asistida ó para videojuegos, p.e.) es deseable contar con una implementación computacional que pueda facilitar dicha reproyección en tiempo real.

Existe una gran diversidad de aproximaciones a la Reconstrucción 3D automática de un objeto o de una escena a partir de una o varias imágenes o bien a partir de una ó varias secuencias de video. En la primer parte de estas notas se desarrolla una introducción básica orientada a mostrar algunas de las aproximaciones más habituales de carácter estático, con especial atención a las basadas en aspectos geométricos con respecto a un plano común de referencia (geometría de las homografías).

Tomasi y Kanade (1991-92) proporcionan los primeros desarrollos de la Reconstrucción 3D asociada a una secuencia de video. Inicialmente utilizaron modelos de tipo Euclídeo (proyección ortográfica) que extendieron a continuación a marcos afines (compatibles con deformaciones aparentes). Estos tópicos se abordan utilizando técnicas SfM (Structure from Motion) basadas en el seguimiento de nubes semi-densas de puntos. Más recientemente, se han desarrollado herramientas para recuperar la trayectoria de la cámara utilizando la consistencia local asociada al seguimiento de un número bajo de puntos en una colección de vistas.

La introducción de sensores ópticos activos utilizan luz estructurada ó bien dispositivos láser 3D. La luz estructurada permite evaluar la curvatura de objetos de gran complejidad, mediante la proyección de un patrón plano sobre la superficie; la evaluación de los mapas de curvatura proporciona una medida aproximada de la forma de objetos complicados (como esculturas, p.e.) de una forma no-intrusiva.

De forma complementaria, los dispositivos láser proporciona nubes de decenas de miles o millones de puntos 3D con una precisión variable dependiendo de los principios físicos de captura (triangulación, diferencia de fase, tiempo de vuelo) o bien de propiedades geométricas (profundidad, orientación relativa, p.e.). Asimismo, la utilización de la información densa proporcionada

por estos dispositivos proporciona una geo-referencia para la localización (posición y orientación) relativa de cámaras convencionales con respecto a un modelo preciso de la escena.

Algunas de las líneas de investigación desarrolladas desde principios de siglo están orientadas hacia la fusión on-line de información procedente de imagen y de rango. La información de rango (procedente de infrarrojos, LED o dispositivos de escaneo láser de línea con direcciones prefijadas) está siendo incorporada en Robótica Móvil para la navegación automática de vehículos. En los años ochenta y noventa se utilizaron otros sensores activos de tipo infrarrojo o ultrasonidos, o bien sensores de tipo magnético (para dispositivos hápticos, p.e.), en relación con operaciones en entornos industriales (agarra, manipulación, pintado, soldadura, p.e.)

La interacción en tiempo real en entornos 3D semi-inmersivos requiere el desarrollo de modelos y herramientas que permitan empaquetar y gestionar la información en primitivas (geométricas o radiométricas) con objeto de simplificar la gestión de objetos o de escenas complejas en relación con requerimientos cambiantes del usuario y con un relativamente bajo coste computacional. Uno de los ejemplos más avanzados está relacionado con la asistencia a la cirugía, donde los objetos rígidos considerados en la mayor parte de este módulo se reemplazan por objetos deformables correspondientes a órganos internos, p.e..

En presencia de grandes volúmenes de datos procedentes, es necesario seleccionar la información “más significativa”. Para resolver este problema, se requieren técnicas de muestreo inteligente y generación de primitivas 3D a partir de nubes de datos agrupados en mallas deformables 3D. Para ello, es preciso identificar un número “suficiente” de puntos de control, tanto en la escena, como en los dispositivos encargados de la captura y del movimiento de la cámara

Asimismo, puede ser necesario generar una cámara virtual que sea capaz de visualizar la escena desde una localización diferente a la actual, en función de la interacción llevada a cabo por el experto. Un ejemplo elemental corresponde a los avatares de SecondLife de mediados de la primera década del siglo XXI, pero actualmente se dispone de soluciones para entornos realistas mucho más complejos y de personajes con una representación y una capacidad de respuesta mucho mayor.

Las necesidades actuales van más allá y requieren la incorporación en tiempo real de propiedades radiométricas (color y texturas) sobre formas 3D, en presencia de malas condiciones de iluminación, teniendo en cuenta que la apariencia cambia con la navegación o la inspección visual de los objetos 3D reales o simulados. Actualmente, ello sólo es posible limitándose a un pequeño número de objetos para los cuales se dispone de modelos mallados deformables gracias a la existencia de “buenas parametrizaciones” (dadas por mallas cuadrangulares ó cuboidales, p.e.).

La integración y gestión de esta información relativa a objetos sólidos B^b (correspondientes a la escena) y deformables (incluyendo agentes reales o virtuales que interactúan dentro de la escena 3D) plantea nuevos retos a los que estas notas pretenden contribuir, proporcionando un soporte desde el punto de vista del modelado físico-matemático y su implementación computacional. Para ello, desarrollamos una estrategia basada en la fusión de información geométrica y radiométrica, dando prioridad a la primera (a diferencia del enfoque llevado a cabo en el módulo B_{31}). A continuación reseñamos algunos de los problemas típicos a resolver que abordaremos en los primeros capítulos de este módulo.

Las *dificultades generales del modelo geométrico* afectan a

- Estimación de la *profundidad* a partir de información radiométrica (Baker y Binford, 1981), de la disparidad entre pares de vistas (Falkenhagen, 1984) o de información métrica en visión monocular (Criminisi, 1999).
- Identificación de hechos geométricos comunes y *puesta en correspondencia* automática entre varias vistas. Dos aproximaciones al problema:

- identificación en una imagen y búsqueda de correspondencias en la otra mediante p.e. procedimientos multi-escala y windowing con tamaños adaptables para las “ventanas” de búsqueda.
- detección independiente de hechos en cada vista y pegado por relajación”.
- Identificación de *elementos estructurales* en modelos de perspectiva (ó sus generalizaciones), estimación de la transformación proyectiva entre modelos y reconstrucción 3D salvo transformaciones proyectivas.

Las *dificultades generales del modelo radiométrico* afectan a

- Naturaleza discreta de la información: filtrado y agrupamiento.
- Identificación y extracción de datos radiométricos significativos relativos a regiones con independencia de la iluminación.
- Implementación de modelos para propagación y rellenado en imagen y visualización sobre el objeto reconstruido 3D.

A pesar de estas dificultades, el modelo radiométrico es de gran utilidad para imágenes en movimiento en las que resulta muy difícil capturar características geométricas de los objetos; esta cuestión se aborda de manera más sistemática en el módulo 5 del CEViC.

Una de las áreas en la que se están produciendo la integración de conocimientos procedentes de dichas áreas es el video 3D y sus aplicaciones a la TV digital 3D personalizada (navegable de forma individual por cada “agente interactivo”). La generación de modelos 3D interactivos consiste en ofrecer un “fondo volumétrico navegable” (extendiendo el tradicional background plano) que sea compatible con diferentes visualizaciones e interacciones llevadas a cabo por otros agentes.

Para ello, es imprescindible usar el marco proyectivo como un soporte en el que los actores 3D puedan ser sumergidos y con los que pueda haber diferentes grados de interacción. Ello requiere una adquisición 3D (basada en diferentes dispositivos de captura de imagen y de rango), una visualización semi-inmersiva 3D (herramientas para la navegación y consulta en escenarios 3D) y una capacidad de respuesta interactiva con actores o personajes virtuales en la escena (extendiendo los avatares sencillos ya existentes en SecondLife o similares).

1.1. Elementos básicos de Reconstrucción 3D

El enfoque bottom-up desarrollado en el módulo 1 de CEViC parte de la imagen como un mapa de bits sobre el que se tienen diferentes tipos de funciones con una distribución irregular que es necesario organizar y agrupar de acuerdo con diferentes propósitos. Por el contrario, en la mayor parte del módulo B_{32} se desarrolla una aproximación top-down en la que la imagen es el resultado de una proyección del espacio 3D sobre una porción del plano 2D.

Este cambio de metodología requiere reformular algunas nociones que aportan el enfoque geométrico (diferentes tipos de geometrías) que es complementario del óptico (basado en diferentes tipos de señales) desarrollado en el módulo B_{31} . En particular, este cambio afecta a un énfasis sobre las propiedades de los diferentes modelos a utilizar para la Reconstrucción 3D.

En la mayor parte de este módulo, una vista se considera como la proyección de un “trozo de mundo”, es decir, un subconjunto acotado del espacio ordinario \mathbb{R}^3 , se proyecta sobre el plano de la cámara. Esta aproximación al problema sugiere un enfoque basado en semirrectas ó rayos r_j que pasan por el centro C de la cámara cuya intersección con el plano π_C da lugar a los puntos de la imagen. Por ello, la *Geometría Projectiva* proporciona el *marco estructural* para la Reconstrucción 3D.

No obstante, cualquier enfoque computacional debe tener presente los algoritmos para procesar la información y las estructuras de datos que agrupan la información contenida en una o varias imágenes.

1.1.1. Elementos de modelado

En esta subsección se repasan brevemente algunas nociones presentadas en el módulo B_{31} (Procesamiento y Análisis de Imágenes Digitales). Por ello, se puede saltar para quienes estén familiarizados con dichos contenidos.

Visión Computacional para Reconstrucción 3D

El *objetivo fundamental* de la Visión Artificial o por Computador consiste en el desarrollo de modelos y algoritmos para extraer, procesar y analizar la información contenida en imágenes digitales. Con ello, se pretende proporcionar una asistencia para la identificación y representación de objetos o de escenas, así como mejorar la interacción de dispositivos artificiales, tanto en el caso estático, como en el caso de movimiento relativo entre el observador y la escena.

Los objetivos de la Visión Computacional requieren resolver cuestiones relativas a modelos, algoritmos y estructuras de datos. Estas cuestiones son similares a las de otras materias presentadas anteriormente como Mecánica Computacional B_1 o Robótica B_2 . De manera más específica afectan a:

- El desarrollo de *modelos físico-matemáticos* a partir de los datos obtenidos usando dispositivos de imagen o de rango.
- El *diseño e implementación de algoritmos* para detectar, extraer y describir hechos optico-geométricos y generar primitivas a diferentes niveles. Para ello, utiliza herramientas de procesamiento y análisis de la información contenida en una o más vistas.
- La gestión computacional de las *estructuras de datos* para la captura, gestión e interpretación de objetos y escenas, así como la capacidad para generar nuevas vistas de dichos objetos.
- La *validación del modelo* tanto en la estimación y corrección de errores, como en la visualización interactiva del modelo 3d generado.

En este módulo, se muestran herramientas básicas para la *reconstrucción 3D a partir de una o varias vistas* tomadas por una cámara estática o en movimiento. Para ello, se sigue una estrategia de complejidad creciente que incluye los pasos siguientes:

1. Detección y extracción de “hechos geométricos” (puntos, líneas) significativos a partir del análisis de información radiométrica contenida en imagen.
2. Agrupamiento de acuerdo con modelos geométricos para la escena, dados inicialmente por modelos de perspectiva.
3. Transformaciones entre modelos para facilitar la predicción y el seguimiento.

Este esquema afecta al desarrollo de estructuras de datos eventualmente móviles, al reconocimiento semi-automático de modelos simplificados y, sobre todo, a la implementación de algoritmos para obtener soluciones robustas para los modelos de escenas generados a partir de la información contenida en vistas o secuencias de video.

Modelado radiométrico

El *objetivo del modelado radiométrico* es la *segmentación radiométrica*, es decir, la generación de un mapa 2D para una imagen o 3D para la superficie X de un objeto sólido Y como unión disjunta de regiones localmente homogéneas desde el punto de vista radiométrico. La homogeneidad local se entiende módulo un umbral de tolerancia fijado por el usuario; por ello, un mismo objeto puede soportar diferentes segmentaciones radiométricas:

- La segmentación radiométrica de una imagen 2D puede ir referida a la escala de grises, el color o la luminancia/reflectancia, p.e.. La representación digital de los datos (según el número de bits) condiciona la “escala” de la representación. Así, p.e. la utilización de sólo dos bits para la escala de grises (representación binaria) dará lugar a una descomposición en blanco y negro, mientras que utilización de ocho bits para cada canal de color da lugar a una descomposición en 256^3 colores que en la práctica es imposible de manejar.
- La segmentación volumétrica de un objeto 3D tiene como objetivo la generación de mapas de color o de reflectancia correspondientes a las superficies que acotan los objetos. A priori, el análisis para los atributos de puntos en mallas 3D o vóxeles en escenas volumétricas es similar al mostrado para imágenes. Sin embargo, en este caso la localización relativa del objeto con respecto al sistema de captura interviene de una manera aún más decisiva para la respuesta de la superficie y las irregularidades en la distribución de los datos radiométricos. Por ello, en este caso es conveniente contar previamente con una segmentación volumétrica por tosca que pueda ser (en términos de orientaciones relativas o bien de concavidad y convexidad, p.e.)

Resumiendo, los objetos presentan propiedades radiométricas muy variables debido la interacción de la luz con la superficie del objeto; esta interacción se manifiesta en términos de la luminancia o la reflectancia del objeto y la sensibilidad del sensor utilizado para capturar dichas propiedades (CCD o CMOS en el caso digital). Estos factores condicionan la distribución de la información que resulta significativa para capturar aspectos fundamentales de las “apariencias” que permiten discriminar a unos objetos con respecto a otros.

El modelado radiométrico en imagen está soportado por mapas de píxeles con su correspondiente distribución en términos de propiedades relativas al color (escala de grises, algún tipo de representación cromática, texturas) o a otras propiedades (luminancia, reflectancia) de la superficie con respecto a la luz (incidente y reflejada). El modelado radiométrico en 3D puede ser de tipo continuo (mapas cromáticos sobre superficies) o discreto disperso (soportado por nubes de puntos). Existen varias formas (no completamente equivalentes, debido a la presencia de patologías) para representar las propiedades radiométricas.

Modelado optico-geométrico

El modelado optico-geométrico tiene como propósito la visualización del modelado geométrico en términos de la *envolvente visual* que es el cono $C_C M$ de rectas tangentes $t_C M$ al borde externo M del objeto volumétrico N trazado desde la posición correspondiente a una cámara (real o virtual) C de un observador ideal (el foco es tipo "pinhole" o agujero de alfiler). El vértice C del cono visual puede estar en el infinito, en cuyo caso, todas las generatrices del cono proyectante son paralelas a una dirección dada (es el caso de las representaciones correspondientes a la planta, alzado y vistas laterales utilizadas en la representación habitual planimétrica. Así, por ejemplo, si todas las generatrices son paralelas a la dirección del eje Oz , entonces el cono proyectante se obtiene resolviendo el sistema de ecuaciones

$$f(X, Y, Z) = 0, \quad \frac{\partial f(X, Y, Z)}{\partial Z} = 0$$

al que se llama el *lugar discriminante* de f con respecto a Z en este caso particular, permite discriminar entre las rectas secantes, tangentes y exteriores al objeto sólido: las rectas tangentes al contorno aparente son aquellas que pasan por el foco de la cámara C y su ecuación presenta al menos una raíz doble correspondiente a un punto de tangencia de la recta r con el objeto M .

El modelado optico-geométrico proporciona una mayor robustez al modelado geométrico obtenido a partir de información visual o de rango, pues permite minimizar problemas debidos a la reflectancia del objeto, aunque persistan problemas debidos a oclusiones parciales. La información $3D$ ligada a los conos proyectantes (algoritmos de envolventes visuales) cuya directriz es (una proyección de) el contorno aparente presentan una gran robustez y soportan algoritmos para la gestión de estructuras superpuestas al objeto (diferentes mallados triangulares, p.e.).

La introducción de octrees sobre una escena $3D$ se ha utilizado tradicionalmente para simplificar la gestión de información $3D$. Esta simplificación consiste en subdividir el espacio y jerarquizar la información atendiendo a atributos fácilmente verificables. Otra utilidad poco explorada es la obtención de mallas cuadrangulares procedentes de la selección de un punto de vista asociado a la localización virtual del observador. Las mallas cuadrangulares están ordenadas y son fácilmente parametrizables lo cual reduce de forma muy significativa la complejidad del modelado. No obstante, las mallas cuadrangulares presentan un mayor número de "patologías" que las triangulares. Para edificios es relativamente fácil adaptar octrees a los planos dominantes; sin embargo, el desarrollo de octrees adaptables a la geometría de objeto complejos es uno de los retos para el próximo futuro.

Modelado topológico

La evolución de mapas de perspectiva corresponde (variación en la localización de la cámara) a una "deformación aparente" de la escena que se controla de manera relativamente fácil mediante la estimación de transformaciones proyectivas. Para acelerar dicha estimación es frecuente incorporar restricciones adicionales correspondientes a características del movimiento rígido (traslaciones y rotaciones) llevado a cabo, o bien relativo a la conservación de características del flujo (de imagen o de escena).

La presencia de objetos deformables con deformación real (persona ó animales, típicamente), da lugar a complicaciones adicionales en el análisis de imagen que sólo se pueden abordar de forma precisa utilizando modelos topológicos. La utilización de este tipo de modelos presenta una complejidad bastante mayor que se desarrolla con más detalle en los dos módulos siguientes B_{33} (Movimiento) y B_{34} (Reconocimiento). No obstante, en este apartado se presentan algunas ideas básicas relacionadas con un tratamiento simplificado del problema en términos de envolventes cambiantes. Como siempre, el reto es el seguimiento en tiempo real de envolventes menos toscas de las presentadas aquí.

El *objetivo del modelado topológico* es la recuperación de aspectos topológicos de los objetos 2D y 3D ligados a la conexividad (número de componentes conexas) para regiones planares o el género de las superficies que acotan objetos volumétricos en el caso 3D. Como los datos son discretos, es necesario superponer estructuras adicionales que pueden ser expresadas en términos de mallas. Las Topologías Algebraica y Diferencial proporcionan herramientas para la obtención de invariantes continuos a partir de datos discretos y la identificación de los cambios cualitativos en la topología de los objetos a partir de la identificación de valores críticos con respecto a algún funcional (altura, distancia, energía, curvatura).

Las estructuras básicas proporcionadas por la Topología Algebraica Computacional (ver módulo B_{12} para detalles y referencias) son *complejos celulares*. Grosso modo, un complejo celular sobre un espacio X es una colección de células d -dimensionales $e_i^{(d)}$ con $i \in I_d$ para $0 \leq d \leq n = \dim(X)$ verificando condiciones de incidencia y compatibilidad, con objeto de que los operadores borde ∂ estén bien definidos (para más detalles ver cualquier manual de Topología Algebraica). Un *mallado triangular* es un caso particular de complejo celular lineal a trozos (PL o Piecewise Linear, en lo sucesivo). Un complejo celular presenta habitualmente un elevado grado de redundancia e interesa suprimirla introduciendo la noción de α -*shape*; dicha estructura se obtiene suprimiendo aristas y caras sin romper la condición de conexión de la estructura superpuesta a X . En CGAL 3.3.1 se puede ver una implementación de las α -shape y algunas aplicaciones para la descripción de objetos con una geometría compleja (esculturas, p.e.)

La Topología Diferencial Computacional (ver módulo B_{13} para detalles y referencias) proporciona herramientas para la recuperación de la topología de variedades M a partir de descomposiciones celulares obtenidas mediante el análisis de puntos críticos no-degenerados de una función $f : M \rightarrow \mathbb{R}$, como la función altura o profundidad, p.e. (hay extensiones asociadas a funcionales más interesantes como el cuadrado de la distancia, la energía o la curvatura). La variedad se reconstruye mediante la superposición de "rodajas" (slices) perpendiculares a una dirección dada; estas rodajas son superficies de nivel que pueden soportar asimismo PL-estructuras para facilitar la reducción de información de los PL-modelos superpuestos en términos de octrees o de representaciones volumétricas asociadas. Los cambios en la topología de las rodajas están determinados por los puntos críticos del funcional f . El índice de la forma cuadrática correspondiente al Hessiano en cada punto crítico determina la dimensión de la célula a añadir. Las estrategias de Topología Diferencial están siendo aplicadas a la Reconstrucción 3D a partir de Imágenes Biomédicas tipo Rayos X, CT (Computer Tomography), NMR (Nuclear Magnetic Resonance) o PET (Positron Emission Tomography), p.e.

1.1.2. Imágenes para la Reconstrucción 3D

Recordemos del módulo B_{31} que hay dos enfoques para la *noción de vista*:

- *enfoque geométrico*: una vista es una proyección cónica $\pi_C : W \rightarrow \Pi_C$ de una escena tridimensional acotada $W \subset \mathbb{R}^3$ (etiquetada como "mundo") sobre el plano Π_C de la cámara; para simplificar se supone que la proyección se realiza desde un punto C al que se llama centro de la cámara.
- *enfoque radiométrico*: una vista es un mapa de bits; en particular para una imagen en escala de grises de 8 bits está dada por una función $f : [a, b] \times [c, d] \rightarrow \mathbb{R}$ definida sobre un dominio rectangular $[a, b] \times [c, d]$ a valores naturales en el intervalo $[0, 255]$ (pues $2^8 = 256$, donde el valor 0 se toma como "negro puro" y 255 como "blanco puro"). Se representa mediante un array en el que cada píxel almacena información sobre la función de intensidad f_g en la escala de grises.

Por ello, una vista contiene información geométrica, es decir, relativa al modelo de proyección (posición, orientación) y radiométrica, es decir, relativa a las características ópticas de la

luz (intensidad de radiación, color y polarización). A lo largo del módulo B_{31} (Procesamiento y Análisis de Imagen) sólo hemos considerado el enfoque radiométrico. A lo largo del módulo B_{32} (Reconstrucción 3D) sólo consideramos el enfoque geométrico, aunque el radiométrico sea necesario para la extracción automática de información que suponemos ya realizada. No obstante, en los párrafos siguientes se recuerdan algunas nociones básicas que permiten relacionar ambos enfoques.

Sobre la doble naturaleza de la luz

Las relaciones entre los enfoques geométrico y radiométrico proceden nuevamente del carácter dual de la luz como fenómeno ondulatorio y corpuscular:

- las *características ondulatorias* se adaptan a la descripción simplificada de la luz en términos de rayos (un rayo es el conjunto de puntos del espacio que se proyectan sobre el mismo píxel de la cámara);
- la *naturaleza corpuscular* de la luz se adapta bien para describir los fenómenos ligados al comportamiento de las superficies en relación con la absorción/emisión de energía vinculados a la incidencia, así como la propagación en diferentes direcciones sobre la superficie.

Ambas cuestiones son cruciales (incluyendo su interacción) y encuentran un gran número de aplicaciones relacionadas con la Informática Gráfica (RayCasting / RayTracing) o la renderización/simulación avanzada (modelos de partículas). La Reconstrucción 3D clásica pone el acento en la representación basada en rayos, pero las aplicaciones más recientes a otros terrenos relacionados con la post-producción digital presentan un creciente uso de modelos que, de una u otra forma, utilizan fenómenos corpusculares en modelos avanzados de propagación sobre superficies (mapas de reflectancia bidireccionales, p.e.).

Procesamiento y análisis de vistas

La extracción de información de cada imagen digital se lleva a cabo usando los modelos y herramientas descritos en el módulo B_{31} . De una forma más precisa, suponemos conocidos los resultados básicos relacionados con filtrado y agrupamiento:

- El *filtrado* tiene como objetivo la detección y extracción de primitivas elementales de tipo geométrico (minisegmentos, p.e.) o radiométrico (mini-regiones, p.e.) utilizando propiedades de discontinuidad (filtros de baja frecuencia) o de continuidad (filtros de alta frecuencia) para la función de intensidad en la escala de grises.
- El *agrupamiento* se basa en la selección de datos que tienen parámetros similares para datos geométricos (segmentos con extremos comunes y pendientes similares, p.e.) o características comunes para datos radiométricos (intensidad o color similar) para su agregación (clustering) en "hechos" que permitan su interpretación. Las estrategias más frecuentes son de tipo métrico (comparación de parámetros modulo un umbral de tolerancia, p.e.), lógico (relajación de etiquetas, p.e.) o estadístico (identificación de correlaciones entre datos numéricos, p.e.).

Los minisegmentos se extraen mediante el filtro de Canny (1984). La extracción de minisegmentos y su agrupamiento en segmentos largos permite identificar líneas de perspectiva y agruparlas en haces. Cada haz pasa por un punto de fuga V_α en el que confluyen al menos cuatro semirrectas¹; se determina como un tipo de junta cuádruple, donde las discontinuidades

¹ Si sólo se dispone de tres, es necesario un análisis más fino para discriminar si la junta triple es tipo flecha (como las que aparecen en vértices externos de la envolvente planar de una proyección plana de un poliedro, p.e.) o bien un verdadero punto de fuga

soaciadas a la prolongación de las semirrectas se presentan de forma consecutiva. Para la identificación de juntas múltiples se utilizan filtros específicos (Deriche, Harris, SUSAN, p.e.) que se describen en OpenCV.

Objetivos de la Reconstrucción 3D

El *objetivo general* de la Reconstrucción 3D es la generación de

- una vista con apariencia volumétrica de un objeto/escena 3D (reconstrucción extrínseca); o bien
- la construcción volumétrica efectiva del objeto/escena independientemente de la localización de observador (reconstrucción intrínseca)

En el primer caso, proporciona soporte a la Renderización (que se desarrolla de forma independiente en el Curso certificado sobre técnicas de renderización), mientras que en el segundo caso proporciona soporte a la navegación interactiva dentro de la Visualización tridimensional. En ambos casos, las técnicas desarrolladas se pueden interpretar como una ingeniería inversa de la Informática Gráfica. Esta primera descripción muestra solapamientos e influencias mutuas de gran interés para diferentes aplicaciones relacionadas con animación, videojuegos y, en general, producción digital 3D (incluyendo efectos especiales).

Los *objetivos específicos* de la reconstrucción de una escena o de un objeto 3D a partir de una o múltiples vistas son

- La *comprensión de la estructura 3D* bajo la *restricción de unicidad*: El modelo geométrico obtenido debe ser único módulo la acción de un grupo de transformaciones que depende del marco geométrico elegido. En un contexto optico-geométrico más amplio, la *restricción de unicidad* se refiere a la determinación única de los parámetros asociados a las características geométricas de la escena y ópticas de la cámara, así como su evolución temporal para movimientos controlados de cámara.
- La *generación de representaciones geométricas coherentes con los datos* debe traducirse en la síntesis semi-automática de representaciones generadas mediante movimientos de ratón a partir de una o varias vistas; este objetivo afecta a la renderización asociada a propiedades radiométricas (color, textura, iluminación).
- La *visualización interactiva de un modelo u objeto 3D* (mediante movimientos de ratón) de escenas u objetos incluye la generación automática de modelos estáticos basados en información lineal a trozos (modelos de perspectiva), y la presentación de su evolución espacio-temporal a partir de la comparación entre datos contenidos en una o varias vistas tomadas de forma secuencial o simultánea.

El módulo B_{32} está centrado sobre todo en aspectos relacionados con los dos objetivos iniciales específicos.

La Reconstrucción 3D como Reproyección

La recuperación de información 3D a partir de una única vista es un problema “mal puesto”: Todos los puntos \mathbf{P}_{jk} contenidos en cada semirecta (rayo) r_j que pasan por el centro de la cámara \mathbf{C} se proyectan sobre el mismo punto $\mathbf{p}_j := \overline{\mathbf{C}\mathbf{P}_{jk}} \cap \Pi_C$ en el plano de imagen; más adelante, veremos que dos vistas tampoco son suficientes para la reconstrucción 3D de la región espacial visible desde ambas. Por ello, es necesario utilizar propiedades ligadas a la proyección del espacio sobre un plano que, una vez fijadas referencias, representaremos mediante matrices rectangulares.

De la misma forma que en Informática Gráfica, en la reconstrucción 3D basada en Geometría, los elementos más significativos son los “rayos”; intuitivamente corresponden a las “líneas de visión” que confluyen hacia “focos de expansión” para el foco de la cámara C ². Formalmente, los rayos son semirrectas que pasan por C que se proyectan sobre un punto de la imagen; ello genera una “pérdida” de la profundidad que se recupera usando Visión Estéreo (en el rango más próximo) o bien mediante pequeños desplazamientos (sacádicos) en el caso humano. Estas observaciones sugieren que la organización de la información en imagen y su visualización tridimensional hay que reconstruirla en términos de diferentes relaciones entre rectas (correspondientes a rayos visuales) y su proyección sobre el plano Π_C de la cámara.

Para mantener la coherencia de la representación, es necesario identificar y añadir “restricciones estructurales” a las líneas de visión. Inicialmente, en una escena arquitectónica de interior o exterior urbana, una estrategia básica consiste en identificar las relaciones de paralelismo para líneas en la escena 3D que corresponden a condiciones de convergencia para líneas en la vista 2D que confluyen en puntos de fuga V_i . El marco para describir estas relaciones está dado por la Geometría Projectiva.

Las líneas de perspectiva resultantes se agrupan en haces de rectas que proporcionan un soporte para la interpretación, la gestión (movimientos rígidos en la escena 3D asociados a los movimientos 2D del ratón) y la inmersión en la escena. Los puntos de fuga V_i y los haces λ_i de líneas de perspectiva ℓ_{ij} a través de v_i proporcionan los *elementos de perspectiva* de la escena. Las “transformaciones proyectivas” (ver más adelante para detalles) permiten gestionar globalmente toda la escena en términos de varios modelos de perspectiva.

Una *distinción básica* a tener en cuenta consiste en que cualquier *transformación proyectiva* tienen lugar entre espacios proyectivos \mathbb{P}^n (habitualmente $n = 2$ o $n = 3$) de la misma dimensión; es decir, están dadas por $(n + 1) \times (n + 1)$ -matrices regulares salvo factor de escala. Por el contrario una visualización en perspectiva se refiere habitualmente a una *visualización de una proyección* $\mathbb{P}^3 \rightarrow \mathbb{P}^2$ del espacio proyectivo tridimensional \mathbb{P}^3 en el plano proyectivo bidimensional \mathbb{P}^2 : es decir, están dadas por 3×4 -matrices regulares salvo factor de escala. Esta diferencia inicial es fundamental para distinguir entre dos formas de entender la Reconstrucción: como una representación planar de un objeto 3D a o bien como una representación de un objeto/escena que se pueda navegar de forma interactiva. En ambas se necesita una estimación de la profundidad.

1.1.3. Modelado semi-automático

El modelado 3D es un área de gran demanda para multitud de CAD/CAM aplicaciones relacionadas con Arquitectura, Urbanismo, Diseño Industrial, Informática Gráfica (incluyendo Realidad Virtual), Renderización ó Visualización, entre otras. La combinación de varias de estas técnicas se utilizan de forma intensiva en Videojuegos. Uno de los mayores retos es la generación interactiva en tiempo real de escenarios “aumentados” a partir de información parcial, incluyendo la inmersión de avatares en la escena con los que se pueda interactuar.

El enfoque tradicional del modelado utiliza de forma manual interactiva herramientas de diseño. Además de su interés intrínseco, la Reconstrucción 3D aporta elementos para asistir el modelado 3D, automatizar algunos procesos y abaratar los costes de producción resultantes. Para ello, utiliza modelos y herramientas que permiten extraer características geométricas a partir de un procesamiento y análisis de imagen centrado en las primitivas geométricas.

En esta sección se revisan algunas cuestiones generales relacionadas con diferentes formas de abordar el modelado que tienen interés para las áreas señaladas anteriormente. Para ello, se adopta un enfoque top-down, es decir, basado en modelos teóricos procedentes de Matemáticas ó Física. Esta aproximación es complementaria del enfoque bottom-up desarrollado en el módulo

²En la literatura se han descrito experimentos psicofísicos para observadores que orientan la atención de forma natural hacia los puntos de fuga

precedente B_{31} (Procesamiento y Análisis de Imagen), donde se desarrollan herramientas para el agrupamiento automático de datos discretos en torno a primitivas (geométricas, topológicas, radiométricas) como las presentadas a continuación.

Modelado geométrico

El *modelado geométrico tridimensional* se describe mediante objetos volumétricos B o bien mediante superficies S que los acotan (es decir, tales que $\partial B = S$, donde ∂ denota el operador “borde”). La representación matemática de las superficies S puede estar dado mediante

- ecuaciones implícitas (globales) de la forma $f(X, Y, Z) = 0$ donde f es inicialmente un polinomio de grado d , es decir, mediante una *función* $f : \mathbb{R}^3 \rightarrow \mathbb{R}$
- representaciones paramétricas (locales) de la forma $(X, Y, Z) = (X(u, v), Y(u, v), Z(u, v))$, es decir, mediante una aplicación $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ definida propiamente sobre un abierto $U \subset \mathbb{R}^2$ y tomando valores en el abierto $\varphi(U) \subset M$

Los inputs digitales para el modelado geométrico proceden de información discreta correspondiente al agrupamiento de mini-segmentos en curvas o de mini-regiones en trozos de superficies con características medibles. Por ello, *todos* los modelos descritos a partir de ecuaciones son ideales.

Habitualmente, no se dispone de ningún tipo de representación paramétrica o implícita de los objetos capturados mediante dispositivos de imagen (cámaras convencionales, omnidireccionales o de vídeo) ó de rango (láser, infrarrojos, dispositivos de tipo acústico, etc).

Algunos problemas del modelado

En la práctica, es necesario desarrollar herramientas (modelos y algoritmos) para la estimación y agrupamiento de datos discretos en objetos continuos y complementar esta información con otra procedente del modelado radiométrico o topológico (en presencia de deformaciones).

El *agrupamiento en torno a primitivas* (geométricas o radiométricas) no es sencillo, debido a diferentes problemas como

- *ajuste a formas* (prefijadas o “aprendidas”) que pueda requerir procesos de optimización en espacios de formas:
- *incertidumbre* sobre los valores de las características ó los parámetros de posibles deformaciones; puede requerir modelos de error que permitan acotar la incertidumbre;
- *información incompleta* relacionada con “inconvenientes” de naturaleza geométrica (occlusiones parciales, p.e.) o radiométrica (reflectancia de superficies con respecto a la emisión incidente, p.e.); puede requerir el desarrollo de técnicas para “completar” la geometría o restaurar la radiometría de los objetos.

Aproximación y Estimación

Para facilitar el ajuste de primitivas geométricas eventualmente curvadas, es conveniente analizar el comportamiento tangencial y normal a las superficies S que acotan los objetos. Dicho comportamiento corresponde a una aproximación de primer orden. Un análisis más fino, requiere la introducción de técnicas asociadas a elementos de orden dos que se expresan en términos de diferentes tipos de curvatura.

Como siempre, el problema radica en proporcionar herramientas estadísticas que faciliten la estimación de las aproximaciones de orden bajo. Es necesario introducir una estimación del error y desarrollar técnicas de optimización que permitan minimizar los errores observados

Modelado híbrido

Por lo general, el *modelado híbrido* de objetos complejos requiere combinar varias de las herramientas descritas más arriba, incluyendo no sólo aspectos relativos a la superficie, sino también a las componentes que forman parte de dichos objetos. Para la gestión de un objeto complejo en términos de sus componentes, es conveniente recurrir a representaciones simbólicas que se gestionan mediante *grafos*.

Un ejemplo típico es el *grafo dual* asociado a una segmentación planar de una imagen o volumétrica de un objeto: Cada región se representa mediante un punto interior al que se asocia un nodo del grafo; cada arista del grafo corresponde a dos regiones adyacentes. En general, el grafo resultante no es un árbol. Frecuentemente, es interesante convertir dicho grafo en un árbol, para lo cual es necesario suprimir aristas sin romper la conexión del grafo; en este caso, se obtiene un tipo de *esqueleto* que representa algunas de las relaciones de adyacencia entre las regiones correspondientes a la segmentación efectuada. El tratamiento de la segmentación de un objeto *2D* o *3D* en términos de grafos permite gestionar dichos objetos como si fueran documentos XML, pero relativos ahora a imágenes o a volúmenes. Ello facilita su gestión de forma remota y presenta un interés creciente en el marco de aplicaciones avanzadas que requieran una interacción con el sujeto o procesos de búsqueda en repositorios de imágenes o de objetos volumétricos.

En esta introducción se recuerdan algunas propiedades básicas de la Visión Computacional, se describen los objetivos, problemas y métodos para la Reconstrucción *3D* y se plantean las estrategias más usuales para el diseño de algoritmos.

1.1.4. Visión Estéreo

La percepción humana de escenas tridimensionales realiza una estimación densa de la profundidad obtenida a partir de la visión binocular o estereo. Una descripción tosca de los mecanismos utilizados por la retina y las primeras capas de la corteza visual se ha presentado en el primer capítulo del módulo B_{31} . En este módulo, sólo se desarrollan *propiedades geométricas* de proyecciones; en particular, no se consideran los efectos asociados a la distribución de conos y bastones en la retina, ni tampoco la diferente sensibilidad en relación con las zonas periféricas. En el último capítulo se incorporan aspectos dinámicos, es decir, correspondientes al movimiento relativo de las cámaras que puede considerarse como un “estéreo temporal”.

Desde un punto de vista geométrico, el *modelado biológico* para la Visión Estéreo utiliza la constancia de la “línea base” b (distancia entre los focos de las cámaras) para calcular la *disparidad*, es decir, la diferencia entre la localización de puntos homólogos contenidos en dos vistas rectiificadas. En una primera aproximación, la profundidad relativa se puede interpretar como la inversa de la disparidad. En otras palabras, distinguimos la “cercanía” de los objetos por una mayor disparidad; en particular, los objetos lejanos (a más de veinte metros para la Visión Humana, p.e.) no presentan disparidad y se dice que “corresponden al fondo” BG.

Cuando se toman vistas desde localizaciones arbitrarias en entornos no controlados, frecuentemente no es posible utilizar información sobre la línea base y es necesario utilizar información procedente sólo de puntos contenidos en pares de vistas que es necesario extraer y comparar de forma automática.

La información procedente de una única vista es insuficiente para una Reconstrucción *3D* de un objeto, debido al carácter incompleto de la información geométrica (falta de información sobre la profundidad, presencia de oclusiones, p.e.) y óptica (dependencia con respecto a iluminación, p.e.). Por ello, la Reconstrucción *3D* requiere al menos dos vistas. No obstante, la identificación de elementos de perspectiva en escenarios arquitectónicos permite inferir datos (salvo factor de escala) sobre la estructura de la escena a partir de una sola vista. En ocasiones sólo se dispone de una sola vista (pinturas, grabados, fotografías antiguas) para un objeto arquitectónico; de ahí el interés en extraer la máxima información posible a partir de una sola vista.

La información afecta a una puesta en correspondencia densa entre puntos homólogos ($\mathbf{p}_{i1}, \mathbf{p}_{i2}$) de cada vista (dos planos proyectivos \mathbb{P}^2 en el caso más general); la puesta en correspondencia da puntos del grafo $\Gamma_\varphi \subset \mathbb{P}^2 \times \mathbb{P}^2$ de la aplicación $\varphi : \Pi_1 \rightarrow \Pi_2$ asociada a la correspondencia entre planos de imagen. Este enfoque da lugar a un marco demasiado amplio que es necesario limitar imponiendo condiciones adicionales. En efecto, dado un punto $\mathbf{p}_{i1} \in \Pi_1$ la búsqueda (ó la generación) de su homólogo $\mathbf{p}_{i2} \in \Pi_2$ en otra vista tendría a priori una complejidad cuadrática. Por ello, es necesario introducir *restricciones* que restrinjan el rango de búsqueda. Hay dos tipos de restricciones según que tengan en cuenta el carácter "disperso" (basado en líneas estructuradas según modelos) o "denso" (basado en puntos sin información adicional) de los elementos a comparar en dos vistas:

Aproximación top-down

La aproximación top-down se basa en un modelo geométrico de la escena que se proyecta sobre un modelo para cada una de las vistas. La proyección sobre la vista es siempre una *proyección central* desde el foco \mathbf{C}_i de la i -ésima cámara sobre un plano Π_i (en la práctica puede tratarse de una única cámara en movimiento).

El espacio de rayos que pasan por un punto del espacio ordinario (representando el observador) se puede parametrizar por el cociente de una esfera \mathbb{S}^2 / \sim donde se han identificado los puntos diametralmente opuestos; la imagen por la aplicación antipodal $a : \mathbb{S}^2 \rightarrow \mathbb{S}^2 / \sim$ es el plano proyectivo \mathbb{P}^2 que proporciona el marco geométrico más general para una vista y, en particular, para cualquier representación en perspectiva.

Los elementos fundamentales para la aproximación top-down son las líneas 3D: las líneas que pasan por el centro de la cámara se proyectan sobre puntos en el plano imagen; todas las demás líneas 3D se proyectan sobre líneas 2D en el plano imagen. Por ello, el primer problema a resolver en la aproximación top-down es la extracción automática de líneas mediante el agrupamiento de mini-segmentos.

Los mini-segmentos se agrupan en segmentos largos y estos a su vez en líneas. Algunas de ellas son *líneas de perspectiva*. Ello se debe a que la mayor parte de las cámaras dan lugar a imágenes según modelos de perspectiva (las lentes esféricas no generan directamente modelos de perspectiva). Un haz λ_i de líneas de perspectiva es una familia uniparamétrica de líneas convergentes en la vista (proyección de líneas paralelas en la escena) que se intersectan en un punto del plano de la imagen, al que se llama *punto de fuga* \mathbf{v}_i . Una línea del horizonte ℓ_∞ es una línea que pasa por dos puntos de fuga $\overline{\mathbf{v}_i \mathbf{v}_j}$. Los puntos de fuga proporcionan información sobre la estructura de la imagen (de forma independiente de la proyección elegida) y, por consiguiente, de la estructura de la escena.

La determinación de puntos de fuga es relativamente fácil en escenas arquitectónicas, p.e., pero necesita restricciones estructurales adicionales para escenas naturales en las que no hay elementos arquitectónicos (paisajes, exploración planetaria). En el capítulo 2 veremos cómo a partir de dos vistas próximas ("pequeña baseline") de un montón de piedras o un paisaje natural es posible proporcionar una reconstrucción 3D de la escena, aunque las líneas de perspectiva no sean fáciles de identificar. Ello es posible gracias a la estimación de la "restricción epipolar", la única restricción estructural general entre pares de vistas que extiende el enfoque basado en elementos de perspectiva. La idea es muy simple: Dadas dos cámaras representadas por \mathbf{C}, \mathbf{C}' , para cada punto 3D general \mathbf{P}_i , se tienen dos imágenes $\mathbf{p}_i = \pi(\mathbf{P}_i)$, $\mathbf{p}'_i = \pi'(\mathbf{P}_i)$ y dos epipolos $\mathbf{e} = \pi(\mathbf{C}_2)$, $\mathbf{e}' = \pi'(\mathbf{C}_1)$

El plano $\Pi_i := \langle \mathbf{C}, \mathbf{C}, \mathbf{P}_i \rangle$ corta al plano Π de la primera cámara en una recta epipolar $\ell_i = \overline{\mathbf{e}_i \mathbf{p}_i}$ y al plano Π' de la segunda cámara en una recta epipolar $\ell'_i = \overline{\mathbf{e}'_i \mathbf{p}'_i}$. Las rectas epipolares se cortan en un punto que pertenece a la línea $\pi \cap \Pi'$ en la que se intersectan los planos de las cámaras. Por tanto, el punto homólogo \mathbf{p}'_i de \mathbf{p}_i se encuentra sobre una línea que es la traza del plano $\Pi_i := \langle \mathbf{C}, \mathbf{C}, \mathbf{P}_i \rangle$ sobre el plano Π' . Esto restringe el rango de búsqueda de homólogos a una línea epipolar y, por tanto, reduce la complejidad de rango cuadrático a lineal. La incorporación

de restricciones adicionales asociadas a "hechos salientes" (geométricos o radiométricos) permite acelerar la puesta en correspondencia, aunque no proporcionen elementos estructurales (independientes de la localización del observador)

De forma análoga, el espacio proyectivo \mathbb{P}^3 proporciona el marco geométrico más general para comparar diferentes representaciones o visualizaciones y, por consiguiente, la aproximación más general a la escena basada en modelos.

La relación entre elementos homólogos de dos vistas se expresa mediante una correspondencia bilineal en el marco proyectivo. Sin embargo, esta correspondencia no tiene por qué ser una transformación del plano proyectivo \mathbb{P}^2 (una homografía planar). El espacio proyectivo permite gestionar modelos de perspectiva arbitrarios a partir de la identificación de los elementos estructurales (líneas de perspectiva y puntos de fuga). La identificación de elementos invariantes (asociados a elementos de perspectiva o a la métrica del espacio) permite reducir el marco proyectivo (conservación de la incidencia y de la razón doble), al afín (conservación del paralelismo y, en el caso normalizado, del área orientada) o al euclídeo (conservación de la ortogonalidad y de la distancia orientada).

Aproximación bottom-up

La aproximación bottom-up se basa en propiedades de la imagen. En este caso, para la Reconstrucción 3D se utilizan propiedades geométricas relativas a "hechos salientes", asociados a propiedades geométricas (junturas donde confluyen dos o más segmentos, p.e.) o radiométricas (puntos en los que aparece un valor extremal para la función de intensidad). Es preciso tener en cuenta que hay junturas procedentes asociadas a discontinuidades de propiedades radiométricas.

El número de junturas dobles en una imagen puede ser de varios cientos o miles. Por ello, la comparación entre elementos homólogos de dos vistas no es elemental; debe tener en cuenta propiedades radiométricas, localización aproximada y restricciones estructurales asociadas a los candidatos.

La formulación intrínseca (es decir, independiente del sistema de coordenadas elegido) de las restricciones estructurales se lleva a cabo en términos de expresiones k -lineales invariantes (donde k es el número de vistas) que son tensores multilineales. Así, p.e.

- para dos vistas existe una restricción estructural (restricción epipolar) que se expresa mediante un tensor bilineal que representa la condición a verificar por elementos homólogos (puntos o rectas) de dos vistas;
- para tres vistas la restricción estructural está dada por un tensor trilineal que representa la condición a verificar por elementos homólogos (puntos o rectas) de tres vistas

y así sucesivamente; una cuestión clave consiste en que dichos tensores representan las correspondencias entre haces proyectivos, lo cual limita el rango de búsqueda para elementos homólogos. Esta idea tan simple es la que permite conectar con el enfoque de la Fotogrametría tradicional basada en el ajuste de haces.

La Geometría Proyectiva como marco

La unificación de los diferentes enfoques para la Reconstrucción 3D ha sido llevada a cabo en el marco de una *extensión de la Geometría Proyectiva* y sus transformaciones regulares (también llamadas homografías). La idea de considerar direcciones como puntos del espacio se encuentra ya en los primeros tratados del Renacimiento y alcanza una formalización definitiva en el tratado de Poncelet (1826). Esta formulación es de uso común en Informática Gráfica, donde los elementos se describen como "rayos" (correspondientes a semirectas en el marco cartesiano convencional).

En particular, la *Geometría Afín* se obtiene fijando un hiperplano del infinito H_i^∞ (una recta del horizonte, en el caso del plano proyectivo \mathbb{P}^2) y considerando las transformaciones que la dejan invariante; análogamente, la *Geometría Euclídea* se obtiene fijando la cuádrca absoluta Q^∞ dada por $\sum_{i=0}^n x_i^2 = 0$ para coordenadas homogéneas $[x_0 : \dots : x_n]$, y considerando las transformaciones proyectivas que dejan invariante dicha cuádrca.

La *extensión* de la Geometría Proyectiva incluye no sólo objetos no-lineales (inmersiones de Segre y de Veronese) consideradas ya en la Geometría Algebraica Clásica de finales del s.XIX, sino también la identificación de tensores estructurales para la puesta en correspondencia entre dos vistas para el marco afín (Matrix Fundamental F) y el Euclídeo (Matrix Esencial E).

Este enfoque fue extendido en la segunda mitad de los noventa a tres o más vistas, mediante la introducción del tensor trifocal. A continuación, se extendió a tensores multilineales que permiten expresar de forma compacta las relaciones multilineales entre elementos homólogos (puntos y líneas) contenidos en múltiples vistas. La inclusión de líneas (más robustas que los puntos, pero menos precisas para el seguimiento) es una extensión de los principios clásicos de dualidad, pero con una extensión significativa al caso de tensores.

La introducción de métodos para el cálculo explícito de tensores trifocales y posteriormente multilineales se lleva a cabo a finales de los noventa, utilizando herramientas computacionales avanzadas para el cálculo con matrices dispersas. En el caso de tres vistas, este tipo de métodos permitieron suprimir la ambigüedad en la Reconstrucción 3D a partir de dos vistas señalada por O.Faugeras y S.Maybank (1992).

Asimismo, estas extensiones permitiendo unificar las estrategias de ajuste de haces utilizadas en Fotogrametría, evitando una tediosa puesta en correspondencia manual que se utilizaba hasta entonces. Asimismo, facilitaron un soporte para la autocalibración (Triggs, 1997) y, por consiguiente, para la Reconstrucción métrica a partir de secuencias de imágenes (Heyden y Astrom, 1997), en ausencia de información sobre los parámetros intrínsecos de la cámara (Pollefeys, Koch y Van Gool, 1998).

Una conexión con la Fotogrametría terrestre

La Fotogrametría terrestre se ocupa de extraer *información métrica* a partir de imágenes usando cámaras calibradas. Para ello es necesario identificar referencias absolutas para la localización (posición y orientación). Esto se lleva a cabo mediante una "estación total" que a partir de una pequeña colección de puntos 3D de referencia, permite construir una triangulación de la escena con respecto a la cual se lleva a cabo una Reconstrucción tridimensional muy precisa de la escena o de los objetos a documentar.

A diferencia de las técnicas basadas en fotogrametría, la reconstrucción 3D de los puntos basada en Visión Estéreo convencional *no es exacta*. La utilización de los mismos principios de Geometría Proyectiva (ajuste de haces), la automatización de la puesta en correspondencia entre elementos homólogos contenidos en múltiples vistas [Har00], el tratamiento estadístico masivo de la información obtenida y la formulación de invariantes intrínsecos (tensores de estructura) para múltiples vistas, han propiciado una convergencia entre los métodos basados en Fotogrametría y en Visión Computacional para la Reconstrucción 3D desde finales de los noventa. Actualmente, se usan de forma indistinta, salvo para trabajos que requieren una precisión submilimétrica.

En cualquier caso, es importante una *estimación del error* y el desarrollo de métodos que permitan minimizarlo de forma automática. En el caso de la Visión Estéreo convencional (sólo dos cámaras) que se desarrolla en el capítulo 3, todos los hechos extraídos a partir de una o varias vistas heredan la incertidumbre procedente de la cámara y añaden otras asociadas a la manipulación (procesamiento y análisis) de la información contenida en imágenes. Esta limitación da lugar a que las covarianzas entre datos puedan comportarse de forma diferente, sobre todo en el caso de reconstrucciones centradas en el objeto.

La incertidumbre es menor cuando se adoptan reconstrucciones centradas en la cámara, especialmente en el caso de cámaras calibradas (conocimiento de parámetros intrínsecos y extrínsecos); la calibración permite transferir propiedades de la imagen a propiedades de los rayos visuales, es decir, las rectas que pasan por el centro C de la cámara. Dependiendo de la aplicación a desarrollar y de los requerimientos del cliente una vez estimado y acotado el error, es necesario elegir entre las diferentes formas de realizar la reconstrucción.

1.2. Estrategias de resolución para la Reconstrucción 3D

Los datos que se toman como referencia pueden ser relativos a la percepción visual, los objetos contenidos en las vistas ó en la escena. De acuerdo con esta distinción, los *problemas fundamentales* de Visión Computacional que afectan a la Reconstrucción 3D aparecen asociados (Banerjee, 2001) a la

1. *Puesta en correspondencia*: Identificar, comparar y pegar elementos homólogos en diferentes vistas procedentes de diferentes proyecciones de un escenario 3D. Un pegado coherente debe estimar y corregir de forma automática las distorsiones; en el caso artificial es necesario llevar a cabo una “corrección” en términos de la “calibración” de la cámara (estimación de parámetros).
2. *Estimación de restricciones estructurales* para diferentes vistas de un objeto o de la escena obtenidas a partir de proyecciones. En el caso de dos vistas próximas, estimación de matriz fundamental F para el caso afín y de la matriz esencial E para el caso euclídeo.
3. *Reproyección*: Generación de nuevas vistas que sean consistentes con las precedentes. Pegado de vistas para generar objetos volumétricos eventualmente curvados.

No existe una única estrategia para resolver estos problemas y en este módulo se describe una introducción a algunas de las más utilizadas. Una referencia bastante más completa es Hartley y Zisserman (2000).

1.2.1. Una taxonomía vinculada al modelado

Criterios para la Reconstrucción 3D

Hay diferentes estrategias para la Reconstrucción 3D. De una forma tosca, se pueden clasificar de acuerdo con diferentes criterios:

- Carácter continuo (líneas, planos) o disperso (puntos, vértices) de inputs y outputs: reconstrucción á partir de regiones (pegado de parches, p.e.) o de contornos (intersección de conos proyectantes, p.e.)
- Caso calibrado o no-calibrado: Conocimiento previo de la calibración o determinación on-line.
- Visión monocameral o estéreo (geometría epipolar, tensores multilíneales).
- Visión estática o dinámica: Estimación de la estructura a partir del movimiento (SFM) o del movimiento a partir de la estructura (MFS) a partir del análisis de video.

Estrategias de análisis

En este curso se desarrollan diferentes *estrategias de análisis* y el correspondiente *diseño de algoritmos*. Algunas *características de las estrategias adoptadas* son:

1. *híbrida* basada en modelos geométricos de perspectiva mapas densos de profundidad y modelos ópticos basados en propiedades de la imagen,
2. *incremental* desde una a varias vistas
3. *multirresolución* para facilitar la gestión de modelos complicados o “muy pesados” desde el punto de vista computacional.
4. *tipo bottom-up* pues se realiza una Reconstrucción 3D sin conocimiento previo del modelo

5. *tipo top-down* pues se fija el marco geométrico (proyectivo, afín o euclídeo) y el modelo de perspectiva a utilizar para la visualización planar.
6. *iterativa* según modelos de complejidad creciente desde los modelos proyectivos más toscos a los euclídeos más finos.

Calibración de una cámara

La calibración se refiere a la estimación de los parámetros intrínsecos (5) y extrínsecos (6) de la cámara. La estimación se lleva a cabo en Fotogrametría usando parrillas de calibración; uno de los métodos más frecuentes procede del algoritmo de Tsai (ver OpenCV) que se comenta en el capítulo 2 de este módulo.

La estimación de los parámetros intrínsecos y extrínsecos de una cámara permite corregir de forma automática las distorsiones más frecuentes producidas por una cámara. Esta estimación fotogramétrica es equivalente a estimar las 11 entradas (12 salvo escala) de la 3×4 -matriz de proyección de una escena tridimensional sobre la imagen plana .

Incorporando información de rango

La disponibilidad de sensores cada vez más precisos y con mayor alcance permite incorporar de manera cada vez más eficiente información relativa a la profundidad. Asimismo, la disponibilidad de información previa sobre la escena o la identificación de marcas (esquinas, máximos de intensidad, p.e.) es asimismo de gran utilidad. El problema no es elemental, pues pequeñas perturbaciones en la escena pueden dar lugar a grandes perturbaciones en la estimación de los parámetros extrínsecos de la cámara.³

Una de las mayores dificultades radica en la estimación de las características cinemáticas a partir de secuencias de video. La segmentación estática es relativamente simple, pero la segmentación dinámica asociada a una diferente variabilidad de las regiones móviles) aún es un reto a resolver para escenas complejas. Por ello, es conveniente utilizar sensores adicionales (barrido continuo horizontal, p.e.) que puedan proporcionar información en tiempo real de características cinemáticas aisladas relativas a los cambios en el mapa de profundidad.

1.2.2. Hipótesis adicionales simplificadoras

La información previa sobre el tipo de escenas o de objetos, simplifica la aplicación de algunas técnicas de Reconstrucción 3D y su eficiencia computacional. En una primera aproximación, podemos distinguir entre hipótesis relativas a la cámara, la escena ó los objetos a reconstruir. y las características del movimiento.

Aunque el marco para la mayor parte de los desarrollos realizados en este módulo B_{32} es casi siempre estático, la extensión al caso cinemático (velocidades y aceleraciones variables) es casi inmediata. Para ello, basta interpretar los planos detectados en la escena como soporte de 1-formas diferenciales y las trayectorias de movimiento observadas en imagen como la proyección sobre el plano Π_C de la cámara C correspondientes a los desplazamientos reales en la escena.

La *simplificación de la escena* afecta sobre todo al cálculo de una *envolvente cuadrangular o cuboidal* asociada a los elementos estructurales de la escena. La dualidad entre rectas y planos del espacio ambiente (bien conocida de la Geometría Analítica Líneal en el espacio cartesiano) se extiende de forma natural a una dualidad entre campos vectoriales y 1-formas diferenciales

³Una vez mostrada la calibración intrínseca de la cámara en el capítulo 2, supondremos que es constante; por ello, el problema se reduce a estimar la calibración extrínseca (posición y orientación del centro) que es un problema de naturaleza Euclídea.

en Geometría Diferencial. Esta dualidad es compatible (aunque debe ser reformulada) en los contextos afín y proyectivo que extienden el marco cartesiano tradicional.

Obviamente, dicha envolvente es una aproximación muy tosca a la forma de los objetos presentes en una escena, pero facilita un marco geométrico que es fácil de actualizar, propagar y seguir. Por ello, facilita el soporte para desarrollos relacionados con cámaras móviles que se desarrollan en los módulos B_{33} (Movimiento) y B_{34} (Reconocimiento) en presencia de cámaras móviles. Las demás simplificaciones se tratan en los apartados específicos que se comentan más abajo.

Hipótesis sobre la cámara

Es frecuente que los parámetros intrínsecos de la cámara permanezcan invariantes a lo largo de la toma de datos. Las hipótesis más frecuentes para una cámara corresponden a que los píxeles de las vistas sean cuadrados (ratio de aspecto unidad) y el punto principal coincide con el centro de la imagen (desviación nula). Estas hipótesis sólo son válidas para cámaras de elevada calidad, donde además los parámetros intrínsecos se mantienen constantes.

Si la constancia de los parámetros intrínsecos no puede ser garantizada (la mayor parte de las cámaras digitales realizan un auto-enfoque o no tienen píxeles cuadrados, p.e.), hay que llevar a cabo evaluaciones métricas o bien incorporar restricciones euclídeas evaluadas directamente sobre la escena, recuperando una Reconstrucción euclídea a partir de la proyectiva directamente:

Habitualmente es necesario realizar una calibración (previa o bien on-line) o especificar e implementar los algoritmos correspondientes a la Reconstrucción 3d para el caso no-calibrado.

Simplificando la escena

Para obtener modelos que pueden ser actualizados e interpretados en tiempo real, es necesario llevar a cabo simplificaciones drásticas de la escena a partir de los datos proporcionados por sensores de imagen y de rango. En escenas construidas por el hombre, la simplificación más drástica de la escena está dada por representaciones cuadrangulares o cuboidales asociadas a mapas de perspectiva \mathbb{P}_m .⁴

Un ejemplo típico corresponde al mapa de perspectiva \mathbb{P}_M de una escena con tres puntos de fuga \mathbf{V}^α para $1 \leq \alpha \leq 3$, con sus correspondientes haces Λ^α de líneas de perspectiva. La unión $\Lambda^{\alpha_1} \cup \Lambda^{\alpha_2}$ de dos haces proporciona un mapa de cuadriláteros, mientras que la unión de tres haces proporciona un mapa de cuboides, en los que cada cara tiene como aristas dos líneas que pasan por dos puntos de fuga.

Por ello, conocidos tres puntos de fuga, el transvase entre información planar y volumétrica es inmediato. Además, en presencia de movimiento se tienen mapas de perspectiva móviles $\mathbb{P}_m(t)$ que están generados por (al menos) puntos de fuga móviles $\mathbf{V}^\alpha(t)$ para $1 \leq \alpha \leq 3$. Una ventaja obvia de este enfoque consiste en que la actualización del modelo estructural se lleva a cabo a partir de un número muy reducido de puntos de fuga.

En el marco cinemático, la información relativa al flujo de imagen (resp. escena) se “codifica” en términos de cajas cuadrangulares (resp. cuboidales) asociada a elementos estructurales en la escena. En particular, el paso de modelos cinemáticos 2D a modelos 3D se lleva a cabo en términos del producto exterior de una forma dada por la dual del campo en la dirección del flujo observado para el movimiento planar. Recíprocamente, la contracción del flujo de escena al flujo de imagen se lleva a cabo en términos de la contracción de una 3-forma diferencial a través por un campo cuya curva integral es la proyección del movimiento observado sobre el plano de imagen.

⁴Una región cuadrangular (resp. simplicial) es la imagen del cuadrado (resp. cubo) unidad por una aplicación afín (resp. proyectiva).

Simplificando objetos

A diferencia de los elementos estructurales que aparecen en la escena, los objetos pueden presentar una diversidad morfológica y una variabilidad bastante mayor. Afectan a diferentes tipos de agentes, que incluyen mobiliario (de interior o urbano), personas, animales, vehículos, etc. Para disminuir la complejidad en el tratamiento computacional, suponemos que ya se ha llevado a cabo una segmentación de imagen, con un etiquetado semántico en el que se ha suprimido (usando operadores morfológicos, p.e.) las regiones no significativas correspondientes a vegetación, cielo, etc.

El análisis de la información proporcionada por los sensores de imagen o de rango para los objetos presentes en la escena, da una colección finita de siluetas en diferentes planos de profundidad superpuestos a la escena. Dichas siluetas se pueden describir en términos de diferentes tipos de envolventes (cuadrangulares, convexas, visuales) según el orden de complejidad que se requiera para la aplicación a desarrollar.

En nuestro caso, la *simplificación de los objetos* afecta sobre todo a la envolvente convexa \mathcal{H} para objetos planares b^β ó volumétricos B^β . Las envolventes convexas tienen diferentes ventajas tales como un cálculo bastante rápido, solución única en problemas de optimización, actualización rápida y disponibilidad de herramientas software eficientes para su tratamiento computacional.

Una aproximación más tosca compatible con la simplificación de escenas (ver apartado precedente) utiliza *bounding boxes* para dichos objetos que se denotan como bb^β y BB^β , respectivamente. Esta aproximación es útil cuando se requiere incorporar en tiempo real la variación de las características cinemáticas asociada a diferentes agentes móviles en escenas de perspectiva cambiantes.

Simplificando el movimiento

Los métodos habituales de Reconstrucción 3D están basados en dispositivos estáticos de imagen (cámaras calibradas o no) o de rango (diferentes tipos de escaneo láser, p.e.). La fusión de ambos tipos de información se desarrolla a partir del capítulo 3 de este módulo para el caso estático. Aunque con menor precisión, esta fusión se puede realizar para el caso dinámico, tal y como veremos en el capítulo 5 (SfM), donde se presentan herramientas de seguimiento para nubes “semi-densas” de puntos capturadas por una cámara móvil, tanto en escenas de interior como de exterior.

Desde principios de los noventa se está utilizando el vídeo para la captura de propiedades de la estructura, la forma o la variabilidad en las sombras a partir del movimiento. Todas estas estrategias se etiquetan como SfM, donde S corresponde a Structure, Shape, Shadow, según el contexto. El enfoque basado en SfM se desarrolla a partir del capítulo 5 de este módulo. El problema a resolver radica en la extracción de características cinemáticas del movimiento propio a partir del seguimiento de un “número suficiente” de puntos móviles (con movimiento real o aparente).

Las variaciones en la velocidad de una cámara móvil $C(t)$ pueden dar lugar a irregularidades en el flujo de imagen y, por consiguiente, en la percepción del flujo de escena. Por ello, es conveniente realizar una segmentación móvil que pueda incorporar y, si es posible, corregir el tratamiento de información móvil de acuerdo con sistemas distribuidos. Para ello, es frecuente introducir simplificaciones para la cinemática del movimiento en las que se supone que la velocidad de la cámara es constante para cada uno de los “segmentos” temporales de una secuencia de vídeo.

A partir del capítulo 5 se introducen herramientas tipo SLAM (Simultaneous Localization and Mapping) que permiten estimar de forma simultánea la localización (posición y orientación) de la cámara, junto con características del movimiento observador sobre nubes semi-densas de “puntos

significativos”. Para facilitar la reinterpretación de esta información, basta re proyectar las nubes de puntos sobre los mapas de perspectiva móviles \mathbb{P}_m que se calculan de forma simultánea.

El *reto actual* es el desarrollo de Sistemas Expertos que sean capaces de llevar a cabo una interpretación automática de los mapas de perspectiva móviles “aumentados” con los objetos \mathbb{AP}_m , incluyendo herramientas para la actualización, seguimiento y predicción. Este tópico es bastante más avanzado

1.2.3. Marco general para la unificación

Desde principios de los noventa se ha desarrollado una unificación en el marco proyectivo que ha ido incorporando aspectos relativos a la radiometría y al movimiento con múltiples aplicaciones en relación con la generación semi-automáticas de modelos tridimensionales para objetos y escenas, la reproyección de propiedades radiométricas (escala de grises, color, iluminación), la incorporación de características cinemáticas del movimiento y la fusión de información procedente de múltiples cámaras. En los párrafos siguientes comentamos estos aspectos por separado.

En la práctica es aconsejable utilizar una combinación de ambas aproximaciones (geométricas y radiométricas), sobre todo para aplicaciones relacionadas con el soporte a la post-producción. En cualquier caso, una vez generado el modelo básico es necesario disponer de un módulo de renderización basado en imagen compatible con la precisión métrica y que proporcione información suficientemente realista del objeto 3D.

Dado el carácter introductorio de este curso y para minimizar la dependencia con respecto al módulo B_{31} (Procesamiento y Análisis de Imagen) y, en menor medida, el módulo B_{33} (Movimiento) de la materia B_3 (Visión Computacional), se desarrollan sobre todo herramientas basadas en el *marco proyectivo* para el caso *estático*. En el módulo B_{33} se incorporan características cinemáticas del movimiento.

Unificación geométrica

El marco general para las *unificación de aspectos geométricos* está dado por una extensión de la Geometría Proyectiva que incluye soluciones automatizadas para

- *Visión estéreo*: binocular estática o a partir del movimiento,
- Métodos para la *puesta en correspondencia* (caso calibrado o no) de elementos homólogos, incluyendo modelos lineales (proyección ortográfica en marco Euclídeo) y no-lineales (marco afín)
- Incorporación de *modelos de perspectiva cambiantes* asociadas a cámaras móviles y predicción a partir del análisis en tiempo real.
- *Fusión de información geométrica* procedente de múltiples vistas procedentes de la captura realizada por varias cámaras.

Unificación radiométrica

A diferencia del enfoque presentado en el módulo B_{31} (Procesamiento y Análisis de Imagen), el marco general para la *unificación de aspectos radiométricos* tiene como objetivo la reproyección de información radiométrica (grises, color, iluminación, p.e.) sobre modelos de objetos o de escenas; para ello, se puede desarrollar una estrategia top-down (evaluando características de la “forma” usando la geometría eventualmente variable de contornos o regiones) o bien en términos

de una estrategia bottom-up (agrupamiento de regiones usando una casi-homogeneidad radiométrica (variación por debajo de un umbral)⁵.

Todos estos aspectos se extienden de forma natural a la *Renderización* de uso común en Informática Gráfica Avanzada (ver materia B_4). La renderización se puede entender como una “reproyección” sobre una vista generada artificialmente (modelo sintético) a partir de la información disponible. La renderización se puede obtener a partir de imágenes estáticas (IBR: Image Based Rendering) o bien a partir de vídeo digital (VBR: Video Based Rendering). Incluye herramientas avanzadas de procesamiento y análisis vinculadas a modelos de propagación tales como

- *Filtrado adaptativo* con respecto a contornos adaptables en el plano de imagen (usando snakes, p.e.).
- *Filtrado supervisado* para regiones creciendo de forma competitiva con condiciones frontera (los diagramas de Voronoi proporcionan una PL-aproximación).
- *Modelos de difusión-reacción* que utilizan Ecuaciones en Derivadas Parciales (PDE) mixtas con condiciones iniciales para la fase de inicialización y condiciones frontera para controlar la propagación, p.e.

Unificación cinemática

La estimación de características cinemáticas (velocidad, aceleración del movimiento (lineal y angular) permite extender los aspectos geométricos y radiométricos presentados en los dos apartados anteriores. La estimación de estas características se lleva a cabo en el módulo B_{33} ; no obstante, es conveniente tener presente esta extensión para que el diseño de las estructuras de datos y la implementación de algoritmos sean compatibles con el caso móvil o incluso con el de múltiples cámaras capturando información de manera simultánea. Ello requiere una capacidad de cálculo que posiblemente esté disponible en los ordenadores embarcados en vehículos en los próximos años.

A la vista de la elevada complejidad de la información, es necesario reducir la información utilizando mapas de perspectiva muy sencillos sobre los cuales se puedan superponer diferentes capas de información correspondientes a la evolución espacio-temporal (real o aparente) de la escena. A finales de los noventa se desarrollan algoritmos de Cinemática Computacional que extienden los de Geometría Computacional aplicados a estructuras cináticas de datos (KDS) ⁶

La aproximación basada en KDS para el análisis cinemático de escenas cambiantes utiliza información semidensa contenida en regiones móviles. La mayor limitación de este enfoque es el carácter disperso de la información que hace difícil una reinterpretación en términos de modelos “más estructurados”. La mayor ventaja de los modelos estructurados es la facilidad para su estimación y el seguimiento; el ejemplo típico está dado por mapas de perspectiva.

Obviamente, los modelos basados en la matriz fundamental (para el caso Afín) ó la matriz esencial (para el caso Euclídeo) son aplicables a situaciones más generales y robustos. Asimismo, permiten generar nuevas vistas del objeto o de la escena, insertando una “cámara virtual” para la visualización desde puntos de vista próximos. Sin embargo, actualmente no se dispone de herramientas para la estimación en tiempo real que se requiere para su aplicación a escenas de tráfico, p.e.

⁵Este método se utiliza en el trabajo conjunto con A.Viloria y M.Gonzalo sobre regiones móviles de Voronoi para la secuencia de tráfico bajo lluvia.

⁶Las KDS fueron desarrolladas inicialmente por el grupo de L.Guibas en Stanford y han sido incorporadas a CGAL desde principios del siglo 21.

Fusión de información

La fusión de todas las características (geométricas, radiométricas, cinemáticas) en escenas de exterior arbitrarias es el problema más difícil. El *enfoque estructural basado en tensores* (para múltiples vistas o para el movimiento), aún presenta cuellos de botella importantes relacionados con la estimación en tiempo real de dichos tensores. Además, una predicción más robusta requiere una interpretación, al menos a nivel semántico (edificios, calzada, vegetación, personas, p.e.) de la escena.

La interpretación semántica requiere sistemas expertos a “bajo nivel” basados en la extracción de regiones radiométricamente casi-homogéneas (variabilidad por debajo de un umbral). La *estimación de campos del movimiento* para cada una de las regiones móviles (con movimiento real o aparente) facilita un soporte para el seguimiento en tiempo real de su evolución espacio-temporal.

Sin embargo, la estimación del movimiento de regiones casi-homogéneas no proporciona un modelo cinemático robusto que permita predecir o bien simular la evolución espacio-temporal de la escena. La estrategia que desarrollamos en los capítulos más avanzados de los módulos B_{32} (Reconstrucción) y B_{33} (Movimiento) utiliza los modelos más básicos dados por mapas de perspectiva móviles. Una limitación obvia de este enfoque radica en que sólo es aplicable a escenas estructuradas (construidas por el hombre), en las que el mapa de perspectiva se puede generar fácilmente. Una ventaja obvia es la generación en tiempo real de dichos mapas.

El *reto más inmediato* es el reconocimiento automático de los elementos de perspectiva (puntos de fuga y líneas del horizonte) por parte de Sistemas Expertos. Este reconocimiento, simplificaría el seguimiento y la predicción del movimiento como soporte a la conducción automática en escenas de tráfico que es una de las áreas más relevantes de aplicación. Para ello, se requiere combinar modelos y herramientas presentadas en los módulos B_{32} (Reconstrucción para las escenas de tráfico), B_{33} (movimiento relacionando flujo de imagen con flujo de escena) y B_{34} (Reconocimiento de elementos estructurales en mapas de perspectiva).

1.3. Diseño de algoritmos

Los algoritmos a desarrollar dependen del tipo de requerimientos planteados: velocidad vs. precisión, dispersión vs. densidad, precisión vs. exactitud, etc. El *diseño de algoritmos para la Reconstrucción 3D* debe tener en cuenta diferentes aspectos y, como ya se ha comentado en otras ocasiones, no hay una estrategia única, utilizándose frecuentemente una aproximación híbrida que debe ajustarse dependiendo de las características del problema, los inputs disponibles y el objeto/escena a reconstruir.

1.3.1. Modelado y resolución del problema óptico-geométrico

- Selección del marco geométrico (proyectivo, afín, euclídeo).
- Caracterización de restricciones estructurales (matrices fundamental y esencial)
- Reproyección y generación de vistas sintéticas.
- Renderización avanzada (mapeado de texturas, iluminación, etc)

1.3.2. Algoritmos según tipo de agregación

El tipo de agregación elegido puede estar basado en

- características de la señal extraídas mediante diferentes tipos de filtrado,
- subdivisiones de la pantalla (ventanas, parrillas) o del objeto (mallas, parametrizaciones) para información rasterizada o vectorial.
- características topológicas de la imagen de tipo estático (componentes conexas, bordes) o dinámico (modelos de propagación, reacción-difusión)
- características de la escena (agregación $3d$ iterativa) asociada a la resolución de problemas de visibilidad, consistencia local de objetos, coherencia global de la escena
- una representación simbólica (diferentes tipos de grafos) de las regiones $2d$ en pantalla o $3d$ en la escena para facilitar la gestión de la información

1.3.3. Algoritmos de coste de pegado

El *coste del pegado para datos homólogos* depende del filtro utilizado y está asociado a

- Identificación de hechos (segmentos, esquinas, bordes, regiones)
- Puesta en correspondencia entre hechos homólogos
- Estimación de restricciones estructurales entre hechos homólogos.

1.3.4. Algoritmos de Optimización

Hay diferentes *tipos de optimización* para mejorar los datos obtenidos:

- *Muestreo*: WTA (Winner-Takes-All), consenso (RANSAC, IMPSAC)
- *Programación Dinámica*: No lineal (Newton, cuadrática, etc), Convexa (Localizadores, elipsoide, etc), Lineal (Dualidad, Método del Símplice, Punto Interior), Entera (soluciones enteras) y Combinatoria (identificación de un politopo convexo que acote las soluciones, en ausencia de información precisa sobre las restricciones)
- *Regularización* para resolver problemas mal-puestos (ill-posed).

- *Operaciones sobre grafos: Corte, Contracciones/expansiones*

En el módulo 2 nos centramos sobre todo en los dos primeros. Los algoritmos de regularización se abordan sobre todo en relación con estimación del movimiento (módulo 3) y los relacionados con operaciones sobre grafos con cuestiones de visión estéreo (módulo 5)

1.3.5. Estimación y corrección de errores

La *estimación y corrección de errores* para puntos 3D se basa en

- Ajuste de haces proyectivos de líneas 2D.
- Cálculo de puntos de intersección para líneas 3D (variantes de LSM)
- Triangulación 3D (Delaunay y extensiones)
- Rotación y traslación (localización relativa) entre cámaras (representación mediante cuaterniones, p.e.)
- Reproyección y generación de vistas sintéticas (algoritmos asociados a tratamiento estadístico de los datos tipo Monte-Carlo, p.e.)

1.4. Aspectos históricos

En esta subsección se describen algunas de las contribuciones más significativas al problema de la Reconstrucción 3D a partir de una o varias vistas, sin tener en cuenta la posibilidad de movimiento de cámara y de obtención de la estructura a partir de un movimiento controlado de la(s) cámara(s).

1.4.1. Reconstrucción 3D y Fotogrametría terrestre

La Reconstrucción 3D se desarrolla inicialmente siguiendo una metodología muy similar a la Fotogrametría terrestre (Brown, 1976). Más adelante, en la última década del siglo XX se pone de manifiesto cómo el ajuste de haces proporciona el marco teórico común para la Fotogrametría (medida a partir de fotos) y la Reconstrucción 3D y se desarrollan algoritmos de Optimización (y de minimización de errores) que utilizan las ventajas de ambas aproximaciones.

El desarrollo de las herramientas de procesamiento digital a lo largo de los años ochenta (con fines inicialmente militares) proporciona el impulso para el desarrollo e hibridación de técnicas procedentes de diferentes áreas. La estrategia habitual seguida en Fotogrametría consiste en determinar los parámetros intrínsecos y extrínsecos de la cámara:

- Los *parámetros intrínsecos* (longitud focal, punto principal, desviación, ratio de aspecto) se determinan off-line en el laboratorio utilizando una parrilla de calibración.
- Los *parámetros extrínsecos* correspondientes a la localización (posición, orientación) del observador se determinan on-line utilizando una red topográfica con respecto a la que se referencia las vistas.

Una vez obtenida la información relativa a la calibración de la cámara, se utiliza dicha información para el pegado y evaluación métrica de la escena o del objeto a partir de una o varias vistas. La evaluación métrica es posible cuando el formato de la foto soporta información vectorial; en caso contrario (formatos TIFF, p.e.), sólo se dispone inicialmente de información raster, lo cual requiere un post-procesamiento de la información.

Lamentablemente, la Fotogrametría tradicional presenta una baja automatización, requiere una elevada especialización y tiene un coste temporal elevado. Además, los algoritmos utilizados en Fotogrametría sólo proporcionan resultados robustos cuando algunos de los parámetros han sido estimados previamente (off-line) con una gran precisión; pequeñas desviaciones (debidas incluso a variaciones en la temperatura o mecánicas) pueden dar lugar a estimaciones sesgadas que dan lugar a errores significativos en el modelo resultante global. Por otro lado, en un gran número de vistas (más aún en grabados o pinturas) puede requerirse la utilización de fotos cuyos parámetros se desconocen (intrínsecos del dispositivo o extrínsecos del observador). En estos casos, es preciso desarrollar una aproximación alternativa basada en información sobre la(s) vista(s) que faciliten la puesta en correspondencia (semi-)automática de datos homólogos de cara a la visualización avanzada de vistas generadas de forma sintética a partir de movimientos de ratón.

No obstante, el marco geométrico y la metodología utilizada en Fotogrametría y en Visión Computacional son muy similares: La metodología se refiere a las estrategias óptimas para la toma de datos y el modelo común para la puesta en correspondencia. El modelo común viene dado por el ajuste de haces de líneas proyectivas en cada vista, que es básicamente una transformación proyectiva entre dos (haces de) rectas proyectivas en el modelo proyectivo \mathbb{P}^2 de cada vista. Ello facilita la transferencia de información entre ambas disciplinas. El modelo geométrico es bastante simple, pero su estimación es altamente no-trivial. Los problemas más importantes en el ajuste de haces afectan a la puesta en correspondencia automática de primitivas comunes entre varias vistas, la estimación de errores y, sobre todo, los procesos de optimización asociados. El ajuste de haces proporciona la aproximación más robusta y precisa al problema de la Reconstrucción 3D y proporciona el soporte común para la conexión con la Fotogrametría terrestre.

1.4.2. Dos vistas

Algunos de las aportaciones más relevantes para la Reconstrucción 3d para la visión estéreo a partir de dos vistas son

- Longuet-Higgins (1981): introducción de la *restricción epipolar* como restricción estructural para calcular la traslación y la rotación entre dos localizaciones de la cámara.
- Faugeras y Toscani (1986): Calibración a partir de dos vistas.
- Koenderink y VanDorn (1989): recuperación de la *estructura afin* de la escena a partir de dos vistas utilizando un invariante de la forma calculado a partir de un *plano de referencia*.
- Maybank (1990): Análisis proyectivo de la ambigüedad para la Reconstrucción 3D.
- Faugeras (1992): reconstrucción *proyectiva* a partir de un sensor débilmente calibrado, supuesta conocida la geometría epipolar.
- Hartley, Gupta y Chang (1992): reconstrucción *proyectiva* (salvo transformación proyectiva) a partir de un sensor débilmente calibrado.
- Mohr, Boufama y Brand (1993): *Reconstrucción proyectiva* a partir de un sensor débilmente calibrado, utilizando una estimación global de incógnitas mediante la minimización de errores residuales entre puntos de imagen y sus retro-proyecciones.
- Faugeras, Luong y Maybank (1992): recuperación de la *estructura euclídea* a partir de la invarianza por semejanzas de la cónica absoluta (es decir, la imagen cónica sólo depende de las propiedades de la cámara). Las ecuaciones de Kruppa (1913) proporcionan la información necesaria. Sin embargo, la estimación de los parámetros de las ecuaciones es un problema difícil de resolver.
- Hartley (1997): Las ecuaciones de Kruppa se pueden obtener a partir a partir de la descomposición en valores singulares (SVD) de la matriz fundamental. Para resolverlas, hay que calcular 3 matrices fundamentales que se obtienen mediante un desplazamiento controlado de un cabezal estéreo.

1.4.3. Una vista

Inicialmente, se adopta un enfoque similar al de la Fotogrametría clásica, incluyendo la rectificación de la vista asociada a la calibración de la cámara. Con respecto a los parámetros intrínsecos, una cámara métrica tiene ratio de aspecto unidad (píxeles cuadrados), desviación nula (coincidencia del punto principal y el centro de la imagen). A partir de estas hipótesis se determinan los restantes parámetros de la cámara. Algunas de las contribuciones más importantes son:

- Tsai (1987): Reconstrucción euclídea a partir de la estimación de la longitud focal de la cámara y de los parámetros extrínsecos a partir de una parrilla de calibración plana.
- Ayache y Hansen (1988): Alineamiento de imágenes en Visión Estéreo.
- Caprile y Torre (1990): Cálculo y utilización de los puntos de fuga para la estimación de los parámetros de la cámara; en particular, el punto principal de la cámara es el ortocentro del triángulo formado por los puntos de fuga correspondientes a tres direcciones ortogonales.
- Faugeras (1993): Desarrollo de métodos (lineales y no-lineales) para la estimación de la 3×4 -matriz de proyección, la robustez de la estimación y la buena elección de los puntos de referencia.

- Rothwell, Forsyth, Zisserman y Mundy (1993): Extracción de la estructura proyectiva 3d a partir de puntos en una imagen
- Shashua (1995): Evaluación de propiedades métricas a partir de una sola vista.
- Cipolla, Drummond y Robertson (1998): Calibración en escenas arquitectónicas a partir de la estimación de puntos de fuga.
- Liebowitz y Zisserman (1998): Autocalibración obtenida a partir de estructuras planares (fachadas de edificios, típicamente) obtenida a partir de relaciones de paralelismo o de ortogonalidad entre líneas y ratios de longitudes en segmentos observados (no se requiere calibración previa).
- Criminisi (1999): Reconstrucción 3D sin calibración previa mediante extracción de información afín a partir de una sola vista, representación algebraica del problema y recuperación de la información algebraica a partir de dicha representación algebraica.
- Criminisi, Reid y Zisserman (1999): Metrología a partir de una sola vista.
- Sturm et al: Reconstrucción proyectiva válida salvo transformación proyectiva (caso no calibrado) a partir de la identificación de un paralelepípedo rectángulo en la escena.

1.4.4. Tres o más vistas

Los hitos más significativos para la reconstrucción proyectiva basada en tres vistas de la escena son

- Shashua (1994): Introducción del tensor trilineal para evitar problemas de ambigüedad mediante "alineamiento", transfiriendo información de dos vistas a una tercera (forma canónica de las matrices de proyección).
- Hartley (1994): Utilización de puntos y líneas para la recuperación de la estructura proyectiva mediante *algoritmo lineal*.
- Armstrong, Zisserman y Hartley (1996): Auto-calibración a partir de tres vistas
- Torr y Zisserman (1998): Estimación robusta del tensor trifocal
- Heyden y Astrom (1997, ICPR Vienna):
- Hartley y Zisserman (2000): Geometría de Una y Múltiples Vistas (Biblia).