Reconstrucción a partir de dos vistas

Javier Finat y Fco Javier Delgado del Hoyo

18 de mayo de 2015

Índice general

1 Reconstrucción a partir de dos vistas

La percepción humana de la profundidad y de la volumetría utiliza la Visión Estéreo basada en pares de vistas sincronizadas y el control del movimiento ocular o de la cabeza para la captura de información tridimensional para los objetos del mundo real. Este control se ve facilitado por el "alineamiento" de elementos comunes sobre una "misma línea". En el caso artificial, la Reconstrucción 3D a partir de dos o más vistas trata de imitar este comportamiento mediante la generación semiautomática de

1. nuevas vistas $2D$ a partir de otras conocidas y
2. un modelo volumétrico navegable de forma interactiva.

El primer objetivo requiere calcular la localización (posición y orientación) de la cámara a partir de la información contenida en cada vista; el cambio en la localización proporciona una estimación del "movimiento" y la discretización del camino que conecta ambas localizaciones proporciona vistas sintéticas intermedias. El segundo objetivo incluye al primero, pues la introducción de una cámara virtual permite obtener nuevas vistas mediante proyección sobre un plano de imagen que es transversal (no necesariamente perpendicular) a la línea de visión. Sin embargo, tiene un mayor coste computacional y el modelo resultante sigue siendo habitualmente incompleto [1]. Dependiendo de los requerimientos de la aplicación se fija el alcance de la herramienta a desarrollar.

En el capítulo 1 se han mostrado modelos y herramientas para la visualización de escenas o de objetos generados por el hombre usando diferentes tipos de perspectiva; para ello, se utiliza información implícita asociada a elementos de perspectiva (puntos de fuga, líneas del horizonte), pero los modelos obtenidos carecen de precisión métrica y pueden presentar deformaciones adicionales asociadas al modelo (habitualmente no-lineal) de perspectiva. En el capítulo 2 se ha abordado la Reconstrucción *métrica* de la parte visible de un objeto o de una escena estática a partir de *una vista* utilizando una cámara calibrada; el carácter métrico del modelo permite estimar medidas y ángulos sobre el objeto o la escena [2]. La "cantidad" de información disponible en una sola vista es habitualmente muy reducida (salvo para cámaras tipo ojo de pez), pero se puede completar mediante el "pegado" de dos o más vistas.

La generación de modelos de perspectiva en escenarios naturales u objetos con geometría muy irregular (vegetación, montones de piedra, p.e.) hace muy difícil aplicar los métodos presentados en el primer capítulo basados en modelos de perspectiva. Por otro lado, los requerimientos asociados a la calibración métrica (descritos en el capítulo 2) presentan un elevado coste computacional que hace difícil una realimentación en tiempo real [3]. En consecuencia, es necesario desarrollar modelos

[1] las zonas cóncavas no se pueden recuperar debido a auto-oclusiones, p.e.

[2] La autocalibración basada en la estimación de la cónica absoluta proporciona la posibilidad de "medir salvo escala"

[3] Una motivación para la necesidad de resultados en tiempo real aparece asociada a la fusión de información para retransmisiones por televisión de video 3D; en presencia de diferentes tipos de cámaras o de una cámara móvil con zoom cambiante, es necesario disponer de herramientas que faciliten la puesta en correspondencia de elementos homólogos

y herramientas software que permitan pegar datos y generar nuevas imágenes a partir de otras imágenes sin recurrir a modelos de perspectiva ni a una calibración K previamente conocida; para ello, es necesario incorporar *restricciones estructurales* (sencillas de estimar) para los elementos homólogos contenidos en varias vistas.

Una *pipeline* típica para dos o más vistas consiste en los pasos siguientes:

1. *Detección y pegado de hechos* mediante filtrado y correspondencias
2. *Estimación de restricciones estructurales* usando Geometría Epipolar
3. *Auto-Calibración* (ver cap. 2) vs *Rectificación simultánea*
4. *Pegado denso* de hechos bajo restricciones del marco geométrico
5. *Reconstrucción 3D* del modelo, de la estructura o de la escena.

A lo largo de todo este capítulo se trabaja bajo las hipótesis siguientes:

- *H1:* Cada vista es la imagen de una proyección central π_i con centro en un punto $\mathbf{C}_i$. Por ello, una vez resuelto el problema de correspondencias, de una forma ideal basta con identificar la matriz de la proyección y aplicar las transformaciones sobre el espacio ambiente y el plano de imagen que muestren cómo se transforman los elementos homólogos.
- *H2:* La *calibración* de la cámara es *desconocida*, es decir, se enmarca dentro de la metodología que hemos etiquetado como *auto-calibración* en el capítulo anterior.

En la práctica las cosas no son tan sencillas, pues ni los objetos ni las escenas están aislados, ni su geometría o su topología son conocidas a priori, ni se dispone de ninguna información sobre la(s) cámara(s) que capturan las imágenes, ni los datos son suficientemente precisos, ni las herramientas computacionales proporcionan resultados que se ajusten de forma "exacta a la realidad". Este capítulo está dedicado al desarrollo de herramientas geométricas para la resolución de estos problemas a partir de imágenes capturadas desde localizaciones próximas.

La primera sección está dedicada a esbozar la jerarquía básica entre los diferentes tipos de Reconstrucción $3D$ con especial atención a la Reconstrucción Proyectiva; este enfoque es muy diferente al presentado en el capítulo anterior, donde se ponía el acento sobre la Reconstrucción Euclídea, es decir, la recuperación de información métrica de la escena en relación con la calibración (obtenida off-line).

En la segunda sección se desarrolla una extensión de elementos de perspectiva y su aproximación mediante diferentes tipos de aproximaciones bilineales que facilitan una solución computacionalmente eficiente, sencilla de implementar y significativa desde el punto de vista de visualización. De una manera intuitiva, si se toman n vistas $V_1,\ldots,V_n$ alrededor de un objeto, para reconstruir todo el objeto, es necesario "enlazar" las matrices fundamentales $\mathbf{F}_{12},\ldots,\mathbf{F}_{n-1,n}$ y $\mathbf{F}_{n1}$, lo cual da lugar a un "camino circular" (al que se llama *lazo* en la variedad $\mathcal{F}$ para la reconstrucción afín (el argumento es análogo para el caso euclídeo reemplazando la matriz fundamental por la esencial). Esta idea intuitiva requiere algunos refinamientos, pues es necesario suprimir la "ambigüedad" asociada a la reconstrucción y garantizar que los datos obtenidos al cerrar el "lazo" son coherentes.

Aún siendo conscientes de la simplificación abusiva asociada a la representación de los rayos de luz como rectas, a lo largo de todo este capítulo supondremos que la luz se propaga a lo largo de rectas. Este enfoque permite conectar directamente con el enfoque de Informática Gráfica basado e Ray Casting y Ray Tracing, como Ingeniería Inversa de la Reconstrucción $3D$; actualmente se dispone ya de algoritmos para Ray Tracing en Tiempo Real (RTRT o RT^2) de gran interés para videojuegos;

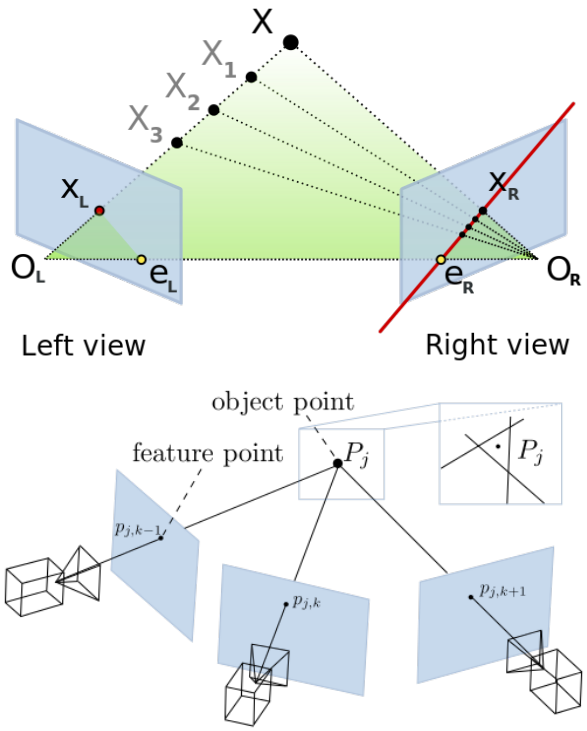

Figura 1.1: Estimación de la matriz fundamental a partir de dos vistas y triangulación para tres vistas

la necesidad de interactuar con escenarios que se exploran plantea el reto para desarrollar Reconstrucción 3D en tiempo real. Estas cuestiones se esbozan en la sección 3 de este capítulo. La última sección está dedicada a presentar los métodos más robustos para la estimación (variantes de Ransac) y algunas de las estrategias de optimización vinculadas a la estructura diferencial de las variedades que parametrizan las restricciones estructurales a las que llamaremos variedad fundamental y variedad esencial. Esta última sección es la más avanzada desde el punto de vista matemático y puede ser saltada en primera lectura.

Nuevamente, la referencia más completa para este módulo es el libro de R.Hartley y A. Zisserman sobre Geometría basada en Múltiples Vistas. Las lecturas 2 a 8 del curso de Marc Pollefeys proporcionan materiales gráficos más intuitivos, si bien los contenidos frecuentemente son más avanzados que los expuestos en este curso. Es altamente recomendable utilizar ambos recursos a lo largo de todo este módulo 2. La figura 1 ilustra algunos de los conceptos que se explican en este capítulo [4].

A lo largo de todo el capítulo es conveniente tener presente objetos volumétricos o escenas de complejidad creciente para los que se desea proporcionar una reconstrucción 3*D*. En el repositorio del grupo LFA-DAVAP [5] hay una gran cantidad de materiales relacionadas con objetos de Patrimonio que presentan una elevada variabilidad morfológica. Por ello, permiten hacerse una idea de las dificultades prácticas que se van a encontrar en relación con las técnicas presentadas.

Como siempre, la referencia más completa para la mayor parte de las cuestiones que se abordan en este capítulo es el libro de Hartley y Zisserman (2000, 2nd ed 2002); el formalismo matemático puede resultar una barrera inicial, pero una lectura pausada ayuda a entender y justificar la mayor parte de los desarrollos llevados a cabo desde mediados de los noventa. El Curso sobre Reconstrucción 3*D*

[4] http://imagine.enpc.fr/~moulonp/openMVG
[5] http://157.88.193.21/~lfa-davap

desarrollado por M.Pollefeys (disponible en la red) es asimismo de gran utilidad y sus presentaciones altamente recomendables pues ilustran la mayor parte de los desarrollos del capítulo.

1.1. Correspondencias y Reconstrucción proyectiva

El problema de correspondencias se refiere a la estimación de elementos homólogos en diferentes vistas. Para ello, es importante que las imágenes a comparar tengan una zona "suficiente" de solapamiento y que los hechos a detectar como candidatos a homólogos sean "suficientemente" robustos para minimizar la ambigüedad en la puesta en correspondencia. Si partimos de cámaras idénticas, un parámetros crucial es la *línea base* b_{ij} que se define como la distancia entre dos localizaciones (posición y orientación) $\mathbf{L}_i$ y $\mathbf{L}_j$ de la(s) cámara(s) en el espacio. Cuando b es "pequeña", la Geometría Epipolar proporciona un primer marco estructural para la puesta en correspondencia de elementos homólogos.

Cuando la línea base b_{ij} es "amplia" el procedimiento descrito da lugar a una degradación en la calidad de los resultados obtenido. Por ello, es necesario recurrir a una aproximación "escalonada" en la que inicialmente fijamos la región de búsqueda (mediante criterios estadísticos de agrupamiento radiométrico o geométrico) y, a continuación, se realiza una búsqueda más fina; en este último caso, no hay una única estrategia, sino una diversidad considerable de procedimientos y algoritmos que se exponen con más detalle en el capítulo 5 de este módulo (para una cámara en movimiento en torno a un objeto rígido), en el módulo 3 (para una cámara fija y objetos móviles) y en el módulo 5 (para diferentes cámaras eventualmente móviles con objetos móviles y eventualmente deformables).

Para empezar nos restringimos al caso en el que la línea base b es "pequeña". La *estrategia general* para realizar una reconstrucción proyectiva es siempre la misma:

1. *Calcular las correspondencias*, es decir, identificar los elementos homólogos. En el caso $0D$ se trata de pares de puntos $(\mathbf{p}, \mathbf{p}' \in \Pi \times \Pi'$ tales que existe $\mathbf{P} \in \mathbb{P}^3$ con $\pi(\mathbf{P}) = \mathbf{p} \in \Pi \simeq \mathbb{P}^2$ y $\pi'(\mathbf{P}) = \mathbf{p}' \in \Pi' \simeq \mathbb{P}^2$ a los que se llama puntos homólogos.

2. Una vez identificada una cantidad "suficiente" de pares de puntos homólogos, estimar la *localización* de cada cámara con centro $\mathbf{C}$.

3. *Construir la estructura de la escena* estimando una cantidad suficientemente densa de puntos $\mathbf{P} = \overline{\mathbf{Cp}} \cap \overline{\mathbf{C}'\mathbf{p}'}$

La puesta en correspondencia para elementos candidatos a homólogos se realiza habitualmente en Visión Estéreo para localizaciones próximas de las cámaras ("pequeña" línea base b), pues es necesario garantizar la existencia de una amplia zona de solapamiento; en este caso, existen relaciones bilineales que afectan a elementos homólogos que proporcionan las restricciones estructurales a verificar. Una vez concluido este proceso, es posible generar una nueva vista para cada localización de la cámara generada de forma virtual mediante movimiento de ratón. Para resolver los tres pasos citados es necesario estimar

1. Las *relaciones bilineales* que deben verificar los pares de puntos homólogos $(\mathbf{p}, \mathbf{p}' \in \Pi \times \times Pi' \simeq \mathbb{P}^2 \times \mathbb{P}^2$ [6]. La existencia de restricciones estructurales para dicha aplicación bilineal acelera la estimación de la matriz fundamental (para el caso afín) o esencial (para el caso euclídeo) que facilita dicha relación estructural a la que se llama *restricción epipolar*

2. La *localización de cada cámara* se ha presentado anteriormente como el núcleo de la matriz de proyección, por lo que basta estimar la matriz de proyección usando los métodos de autocalibración que han sido presentados en el capítulo anterior.

[6] Recordemos que la imagen del producto de Segre $\mathbb{P}^2 \times \mathbb{P}^2 \hookrightarrow \mathbb{P}^8$ parametriza *todas* las aplicaciones bilineales entre dos planos proyectivos

3. *Nuevas vistas* que intuitivamente consideramos como una interpolación continua entre vistas conocidas V_0 y V_1 que se consideran como "extremos de un camino" que conecta las vistas V_0 y V_1. La *generación semi-automática* (mediante movimientos de ratón) de nuevas vistas se resuelve mediante cálculo óptimo de un camino entre la localización inicial de la cámara y la seleccionada por el usuario de forma interactiva. La obtención del "mejor camino" (más suave y de longitud mínima) requiere resolver problemas de optimización en la variedad fundamental o esencial, o bien sobre el variedad asociada a los grupos de matrices que son significativos para cada uno de los marcos geométricos elegidos.

En este capítulo se desarrolla un enfoque que esta basado en la geometría lineal asociada a puntos o rectas del espacio, y sus proyecciones sobre diferentes planos de imagen. Los puntos "salientes" (vértices de poligonales, junturas, p.e.) y rectas "significativas" (líneas de perspectiva, líneas homólogas, p.e.) se obtienen a partir del procesamiento y análisis automáticos de imágenes. En el mejor de los casos, el punto de partida de la Visión Estéreo tradicional está dado por pares fotogramétricos tomados con dos cámaras calibradas idénticas montadas sobre un dispositivo métrico (par estéreo) con una distancia conocida entre los focos de las cámaras a la que se llama *línea base* y se denota mediante b; este valor y la convergencia de las líneas de visión en un punto (que se supone previamente conocido) condicionan el rango en el que los objetos se perciben como volumétricos [7].

La *novedad* de este capítulo radica en que, a diferencia de los métodos de Fotogrametría (ver capítulo 2), es posible llevar a cabo la Reconstrucción 3D sin recurrir a la calibración de las cámaras utilizando solamente información contenida en múltiples vistas (Faugeras, 1992; Hartley, 1992). Si los parámetros intrínsecos de la cámara (calibración interna) se mantienen fijos, la recuperación de la estructura métrica salvo escala se lleva a cabo mediante una estimación de (la proyección sobre el plano de imagen de) la cónica absoluta ω_∞. La justificación geométrica es muy simple: La cónica absoluta es la única cónica que permanece invariante por cualquier transformación euclídea; la justificación analítica es algo más sofisticada: la matriz de calibración y la cónica absoluta están relacionadas por un producto de matrices inversibles [8]. Por ello, el punto clave para relacionar la reconstrucción proyectiva con la métrica es la estimación de la cónica absoluta.

1.1.1. Marcos, modelos y datos geométricos

Desde el punto de vista topológico, las correspondencias entre imágenes se pueden realizar entre:

1. *elementos* $0D$ como vértices/esquinas (correspondencias geométricas) o máximos de intensidad (correspondencias radiométricas), p.e.

2. *elementos* $1D$ basadas en siluetas S_α o en grafos asociados (esqueletos, Reeb) a los objetos detectados en imagen;

3. *elementos* $2D$ con atributos radiométricos (distribución del color o de otras propiedades de la luz)

El caso más difícil corresponde al tratamiento basado en información $1D$ y se aborda en el módulo en el marco del Reconocimiento de objetos. De forma complementaria, se pueden incorporar criterios asociados a:

[7] En el caso humano la línea base es aproximadamente 6.5 cm y el rango de profundidad en el que se percibe la volumetría es de aproximadamente 18 m

[8] Ver la sección sobre autocalibración del capítulo 2 para detalles

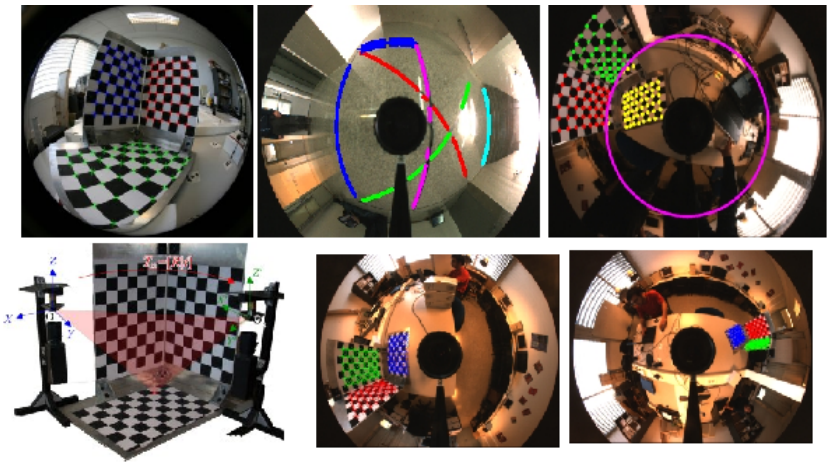

Figura 1.2: Calibración de dispositivos omnidireccionales (catadióptricos y ojo de pez)

- El *procesamiento de la información* a bajo nivel vinculada a los histogramas (distribuciones de la forma y sus generalizaciones, p.e.)
- *Transformaciones* realizadas sobre la imagen en el marco del dominio de la frecuencia
- La *escena* tales como la profundidad relativa o algún descriptor asociado al campo de luz [9]
- *Hechos específicos* tanto de tipo geométrico, como análisis espectral con sus correspondientes estrategias de segmentación.

Esta diversidad tan grande de aproximaciones da lugar a que no exista un estándar común. Por ello, para fijar ideas nos restringimos al caso geométrico en el que las correspondencias se llevan a cabo entre hechos $0D$ geométricos en el dominio espacial, pudiendo adoptarse diferentes marcos para la representación. Esta elección simplifica el tipo de transformaciones a utilizar que supondremos lineales, es decir, dadas por un subgrupo del grupo lineal general o su proyectivizado. Bajo estas restricciones, la *estrategia habitual* consiste en

1. *fijar el marco geométrico* (proyectivo, afín, euclídeo) a utilizar y estimar, junto con sus jerarquías;
2. *fijar el modelo* que incluye el tipo de cámara: cámara de perspectiva (proyección central), omnidireccional (catadióptrico vs ojo de pez, espejos cóncavos, p.e.) ó multivista (cabezal con cámaras sincronizadas con disposición circular, p.e.).
3. *especificar los datos geométricos* a agrupar: puntos, líneas, cónicas. etc y sus transformadas.

Los datos proceden de una extracción y agrupamiento de "hechos locales"; de forma intuitiva un hecho local es un objeto contenido en imagen que "difiere" de los contenidos en los píxeles más próximos. Más formalmente en el dominio espacial de una imagen digital es el soporte de una discontinuidad (geométrica o radiométrica) para alguna función definida sobre la imagen que sea "invariante" con respecto a las transformaciones consideradas.

Para facilitar el pegado el marco más general es el proyectivo; para visualizar los resultados usaremos el marco afín. Para algunos tipos de cámara es conveniente utilizar la geometría cilíndrica (para

[9] La función plenóptica se desarrolla en el capítulo 4 en relación con cuestiones de renderización

vistas panorámicas, p.e.) ó bien la geometría esférica (para dispositivos omnidireccionales; la figura 1.1.1 del grupo de Visión Computacional de la Universidad de Zaragoza ilustra los tipos de distorsión asociados a dispositivos omnidireccionales [10]. El modelo elegido afecta al tipo de reconstrucción (proyectiva vs afín, para cámaras de proyección perspectiva) y el tipo de transformaciones y proyecciones no sólo del marco geométrico, sino también de los objetos que se desea estimar y que deben ser invariantes con respecto a homografías [11].

Los *datos* utilizados en este capítulo son geométricos y corresponden inicialmente a puntos $\mathbf{p}_i^\alpha \in \Pi_\alpha$ y líneas $\ell_j \subset \Pi_\alpha$ en el plano de imagen Π_α correspondiente a un modelo de cámara asociado a una proyección central con centro $\mathbf{C}_\alpha$. Frecuentemente, se consideran datos híbridos, es decir, colecciones de puntos y líneas. *Todos los datos* están *contaminados* por el ruido, las distorsiones debidas al proceso de captura y los artefactos asociados a las primeras fases de Procesamiento y Análisis de la información. Ello implica que es preciso llevar a cabo un proceso de rectificación de las imágenes utilizadas y de optimización en todas las fases del proceso de Reconstrucción.

Para fijar ideas y especificar el modelo se utiliza inicialmente una cantidad minimal de datos. Debido a los errores procedentes del sistema de captura o inducidos por las primeras fases del procesamiento y análisis de imágenes, así como a la necesidad de actualizar información en función de la visibilidad o de geometrías complicadas para objetos, es necesario utilizar información redundante. Inicialmente se supone que dicha información está soportada sobre puntos en diferentes imágenes. Incluso si la puesta en correspondencia se resuelve de forma precisa (verificando las restricciones epipolares), la elevación de la información para reconstruir objetos o escenas volumétricas da lugar a todo tipo de errores que es necesario corregir. Esta corrección requiere combinar

- *modelos geométricos* vinculados al modelo de perspectiva vs omnidireccional, p.e.;
- *técnicas estadísticas* para el muestreo y agrupamiento (para ajuste de datos); y
- *técnicas de optimización* para garantizar la convergencia rápida hacia modelos robustos.

Una adecuada combinación de las diferentes técnicas modifica por lo general las colecciones de datos bajo restricciones geométricas que refuerzan la consistencia local de los objetos y la coherencia global de las escenas.

La *visualización* de un objeto o una escena requiere una cantidad "densa" de puntos; no se puede representar mediante una cantidad minimal de puntos o de rectas como en los modelos simplificados de perspectiva. Es imprescindible utilizar modelos densos. Por ello, los modelos matemáticos para el tratamiento de la información deben ser compatibles con la gestión de información redundante (incluyendo diferentes tipos de remuestreo, p.e.) y su optimización (diferentes modelos para distribuciones de error, distintas funciones de coste, diferentes procedimientos, etc). La introducción de restricciones estructurales de tipo multilineal (matriz fundamental/esencial, tensor trifocal o tensor multilineal) para $N \geq 2$ vistas afecta a una cantidad "densa" de puntos. Por ello permite abordar la reconstrucción densa utilizando técnicas de optimización apropiadas que se exponen en el resto del capítulo y en el apéndice (para la reconstrucción a partir de $N \geq 3$ vistas).

Una reformulación matemática del modelo

Desde el punto de vista del modelado matemático, la Reconstrucción $3D$ se basa en

[10] Este enfoque se aborda con más detalle en el cap.4

[11] Esta condición motiva y justifica una revisión de la Transformación Directa Lineal (DLT) en términos de datos normalizados para garantizar su invariancia; ver más abajo

- Una colección finita (eventualmente densa) de puntos $\mathbf{P}_i \in \mathbb{P}^3$ para $1 \leq i \leq N$
- Una colección finita de proyecciones centrales $\pi^j : \mathbb{P}^3 \to \Pi^j \simeq \mathbb{P}^2$ sobre el plano Π_j de cada cámara para $1 \leq j \leq c$
- Una colección finita (eventualmente densa) de puntos $\mathbf{p}_i^j \in \Pi^j$ para $1 \leq i \leq N$ y $1 \leq j \leq c$ con $\pi^j(\mathbf{P}_i) = \mathbf{p}_i^j$

Los *problemas a resolver* consisten en establecer las correspondencias, estimar los puntos espaciales $\mathbf{P}_i$ y las matrices M_j que representan las proyecciones lineales π_j. Las correspondencia se establecen a partir de una colección redundante de puntos significativos $\mathbf{p}_i^j \in \Pi_j$ en diferentes vistas, candidatos a "homólogos". Para simplificar nos reducimos a dos vistas próximas; la *restricción epipolar* impone una restricción que deben verificar los pares $(\mathbf{p}_i^1, \mathbf{p}_i^2) \in \Pi_1 \times \Pi_2$ que "proceden de un mismo punto" $\mathbf{P}_i \in \mathbb{P}^3$, es decir, tales que $\pi_1(\mathbf{P}_i) = \mathbf{p}_i^1$ y $\pi_2(\mathbf{P}_i) = \mathbf{p}_i^2$.

Para simplificar la notación usaremos $(\mathbf{p}_i, \mathbf{p}_i')$ para denotar $(\mathbf{p}_i^1, \mathbf{p}_i^2)$ con coordenadas $(\mathbf{x}_i, \mathbf{p}_i')$ para los candidatos a homólogos, reservando los superíndices (que hacen referencia a la j-ésima cámara) para el caso de más de dos vistas (estas cuestiones se abordan en el apéndice al capítulo 3) [12]

La presencia de errores y ruido en las fases de captura y procesamiento/análisis de la información requiere diseñar e implementar procedimientos que facilitan una selección "inteligente" de información de acuerdo con los objetivos requeridos y una aceleración en la forma de procesarla que faciliten la generación de modelos con diferentes niveles de detalle. La primera cuestión afecta al análisis de imagen y la selección de "hechos salientes" para una reconstrucción "dispersa"; la segunda cuestión afecta a la optimización

La generación de una nueva imagen a partir de otras conocidas, se entiende como la construcción de una nueva matriz de proyección cuya imagen se muestra en pantalla. Esta idea resulta muy atractiva, pero presenta la desventaja de tener que elegir de forma cuidadosa un sistema coordenado en $3D$ para expresar cada matriz de proyección. Esta selección puede resultar natural en el caso de Reconstrucción Euclídea (para control de calidad relativa a objetos volumétricos, p.e.). Sin embargo, fija grados de libertad en el marco euclídeo, lo cual resulta poco útil para la Reconstrucción (afín o proyectiva) en el caso no-calibrado. En el capítulo siguiente se lleva a cabo un análisis exhaustivo de las matrices de proyección para la reconstrucción métrica a partir de una sola vista.

La *estrategia alternativa* que se desarrolla en este capítulo consiste en un paso previo asociado a la generación de nuevas vistas usando correspondencias entre haces de líneas "epipolares" contenidos en vistas; la innovación con respecto al capítulo 1 consiste en que dichos haces de líneas *no son* haces de líneas de perspectiva, sino haces sobre los que se localizan los puntos homólogos. La idea consiste en trasladar la geometría del haz de planos Λ^b que pasa por la línea base $b = \mathbf{C} \times \mathbf{C}'$ a cada uno de los planos de proyección Π y Π' (asociados a cada cámara $\mathbf{C}$ y $\mathbf{C}'$). La intersección del haz de planos con cada plano de cámara da lugar a dos haces de rectas

$$\Lambda_e := \Lambda^b \cap \Pi \quad , \quad \Lambda_e' := \Lambda^b \cap \Pi'$$

a las que se llaman *líneas epipolares* parametrizadas por $\mathbf{P}^1$. Cada punto significativo $\mathbf{p}_i$ (resp. $\mathbf{p}_i'$) se encuentra sobre una línea epipolar λ_i^e (resp. $\lambda_i^{e'}$). Por ello, el ambiente para la puesta en correspondencia es el más simple posible y corresponde a las homografías de una recta proyectiva.

[12] En la literatura frecuentemente se denota mediante $(\mathbf{x}_i, \mathbf{x}_i')$ a pares de puntos homólogos en pares de imágenes, reservándose la notación $\mathbf{P}_i$ para representar la matriz de proyección; a pesar de un uso muy extendido, esta notación tiene dos inconvenientes: Se identifican puntos con sistemas de coordenadas (olvidando que un mismo punto se representa de muchas formas con respecto a muchos sistemas de coordenadas) y nos quedamos sin notación para representar los puntos en el espacio ambiente 3D

Relación con el modelado y calibración

El modelado de vistas $2D$ depende del tipo de proyección seleccionado y la calibración del tipo de dispositivo para la captura. Inicialmente, se supone que la proyección ideal π_j es central tipo pinhole, es decir, el foco se representa mediante un punto material (centro de proyección) $\mathbf{C}_j$ sobre un plano de imagen Π_j. Este modelo de proyección es ideal y poco realista, pues limita enormemente el campo de visión, pero se adopta como una primera aproximación cuyas limitaciones se corrigen utilizando aproximaciones vinculadas a la calibración y un modelado más realista. Algunas *observaciones previas* a tener en cuenta son:

- *Relación con el modelado:* La generación semi-automática de vistas no incluye necesariamente un modelo $3D$ del objeto, pues sólo se pretende mostrar *nuevas vistas* de un objeto B_α o de una escena tridimensional $\mathcal{E}$ sin recurrir a un modelo previo; esta observación cual marca una primera diferencia importante con respecto a las estrategias de modelado gráfico $3D$ o con respecto al diseño asistido (CAD/CAM). La interrelación con ambas áreas (modelado y diseño) forma parte de los tópicos que se pueden desarrollar como prácticas.

- *Calibración:* El caso ideal corresponde a cámaras cuya calibración interna se conoce, lo cual reduce el problema de la calibración a estimar la localización (posición y orientación). Sin embargo, las imágenes que se utilizan habitualmente para la Reconstrucción $3D$ no proceden de la misma cámara, ni están tomadas bajo condiciones similares, ni existe información disponible sobre los parámetros de las cámaras. Por ello, la única posibilidad para el "pegado" de la información disponible consiste en referir toda la información al objeto o la escena y "alinear" las imágenes antes de proceder a cualquier tipo de "interpolación" entre vistas próximas.

En el capítulo anterior se ha mostrado el papel que juegan la calibración para la Reconstrucción $3D$ en e caso euclídeo en relación con las condiciones de coplanariedad de rayos homólogos con la línea base. En este capítulo centramos la atención en una interpretación afín de dichas condiciones de coplanariedad para obtener la reconstrucción $3D$ en el caso no-calibrado.

En el marco proyectivo, el conocimiento a priori de la matriz de calibración $\mathbf{K}$ permite suponer (multiplicando por la inversa) que la calibración está representada por la matriz identidad $\mathbf{I}$. Cuando esta calibración es desconocida a priori o no se puede estimar sobre la marcha, se plantea la cuestión de cómo generar nuevas vistas y cómo recuperar la estructura de la escena módulo transformación proyectiva (Faugeras, 1992; Hartley, 1994). El problema se puede abordar de dos formas complementarias

- Como una cámara no-calibrada moviéndose en un espacio rectificado.

- Una cámara calibrada moviéndose en un espacio distorsionado.

Para fijar ideas, nos restringimos a la primera interpretación. En este caso, el movimiento euclídeo representado por pares $(\mathbf{R}, \mathbf{t})$ se representa para una cámara no-calibrada mediante $(\mathbf{KRK}^{-1}, \mathbf{Kt})$. La figura 1.1.1 ilustra esta idea.

Matrices de las cámaras

Como se desea realizar esta generación de nuevas vistas de forma continua, es conveniente modelar la colección de nuevas vistas como un camino $\gamma : [0,1] \to \mathcal{M}(3\times 4;\mathbb{R})$, donde $\gamma(0) = M_\pi$ es la vista inicial (habitualmente la izquierda) y $\gamma(1) = M_{\pi'}$ es la vista final (habitualmente la derecha).

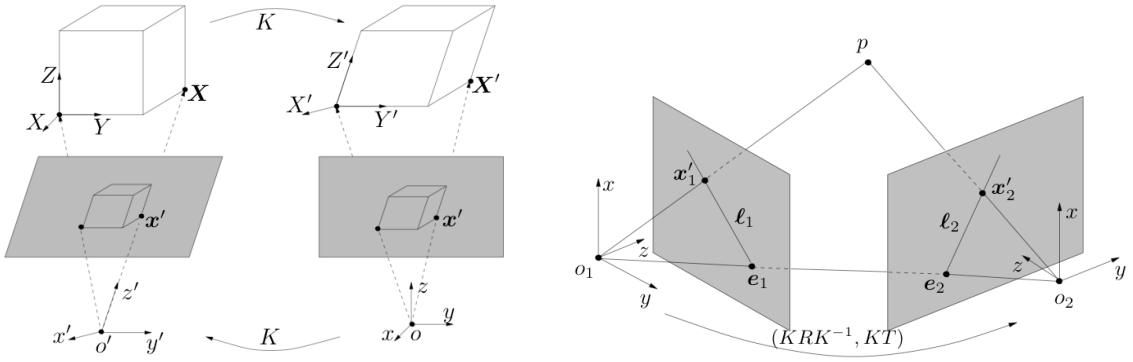

Figura 1.3: Transformación mediante movimiento euclídeo entre un par de imágenes

La acción del grupo estructural permite elegir coordenadas en los espacios de partida y llegada de modo que la matriz $\gamma(0)$ sea la más simple posible, es decir, el "origen" de coordenadas en el espacio vectorial de las matrices $\mathcal{M}(3\times 4;\mathbb{R})$.

Las matrices de proyección asociadas a las cámaras se calculan de la forma siguiente:

Dada la matriz de proyección $\mathbf{P} = [\mathbf{I}|0]$ y $\forall$ vector $\mathbf{v}$ para $\mathbf{P}'$ (similar para P'') se tiene

$$\mathbf{P}' = [[\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3]\mathbf{e}'' + \mathbf{e}'\mathbf{v}^\top|\lambda\mathbf{e}'],\ \lambda \in \mathbb{R}.$$

Elegimos ahora

$$\mathbf{P}' = [[\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3]e''|\mathbf{e}']$$

bajo la condición, $\mathbf{a}_i = \mathbf{T}_i\mathbf{e}''$ se fija de forma única $\mathbf{P}''$.

Sustituyendo en $\mathbf{T}_i = \mathbf{a}_i\mathbf{b}_4^\top - \mathbf{a}_4\mathbf{b}_i^\top$ se obtiene

$$\mathbf{T}_i = \mathbf{T}_i\mathbf{e}''\mathbf{e}''^T - \mathbf{e}'\mathbf{b}_i^\top \Rightarrow \mathbf{e}'\mathbf{b}_i^\top = \mathbf{T}_i(\mathbf{e}''\mathbf{e}''^T - I) \Rightarrow \mathbf{b}_i = (\mathbf{e}''\mathbf{e}''^T - \mathbf{I})\mathbf{T}_i^\top\mathbf{e}'$$

Por ello, finalmente se obtiene $\mathbf{P}''$,

$$\mathbf{P}'' = [(e''e''^T - \mathbf{I})[\mathbf{T}_1^\top, \mathbf{T}_2^\top, \mathbf{T}_3^\top]\mathbf{e}'|\mathbf{e}'']$$

Datos asociados a la proyección

Para una proyección en forma canónica $\pi_{\mathbf{C}} : \mathbb{P}^3 \to \mathbb{P}^2$ la matriz de proyección está dada por $(\mathbf{I}_3 \mid \mathbf{O})$. Por ello, el núcleo (centro de proyección) $\mathbf{C}$ tiene como coordenadas $[0:0:0:1]^\top$ y se le llama el *centro óptico*. En el plano de imagen con coordenadas homogéneas $[x_1 : x_2 : x_3]$ y afines $(x, y, 1)$ (siendo $x = x_1/x_3$ e $y = x_2/x_3$), la línea del infinito está dada por $x_3 = 0$ que es imagen de un plano proyectivo en $\mathbb{P}^3$ al que se llama *plano focal*.

En presencia de varias cámaras o de una cámara móvil, es necesario considerar la proyección proporcionada por una cámara arbitraria tiene un modelo proyectivo lineal dado que se representa mediante una 3×4-matriz $\mathbf{M}_\pi$. Escribimos vectorialmente el resultado de esta proyección sobre cualquier punto $\mathbf{P}$ con coordenadas afines $\mathbf{X}^a = (X, Y, Z, 1) = \tilde{\mathbf{X}}$ como

$$\mathbf{M}_\pi\mathbf{X} = \begin{pmatrix} \mathbf{m}_1^\top & m_{14} \\ \mathbf{m}_2^\top & m_{24} \\ \mathbf{m}_3^\top & m_{34} \end{pmatrix}\mathbf{X} = \begin{pmatrix} \mathbf{m}_1^\top\tilde{\mathbf{X}} + m_{14} \\ \mathbf{m}_2^\top\tilde{\mathbf{X}} + m_{24} \\ \mathbf{m}_3^\top\tilde{\mathbf{X}} + m_{34} \end{pmatrix}$$

Con esta notación, el *centro óptico* $\mathbf{C} = (\tilde{\mathbf{c}}, 1)$ o centro de proyección está dado por el núcleo de la aplicación anterior, es decir,

$$\mathbf{C} \in Ker(M_\pi) \;\Rightarrow\; \tilde{\mathbf{c}} = -\tilde{\mathbf{M}}_\pi\mathbf{m}_4$$

donde $\tilde{\mathbf{M}}_\pi$ es la 3×3-matriz correspondiente a las tres primeras columnas de la matriz de proyección [13] y $\mathbf{m}_4$ es la última columna de la matriz de proyección.

Dado cualquier punto $\mathbf{p} \in V$ de la vista, como $V \simeq \mathbb{P}^2$, el *rayo óptico* asociado a $\mathbf{p}$ está dado por el conjunto de puntos $\mathbf{P} \in \mathbb{P}^3$ que se proyectan sobre $\mathbf{p}$, es decir, tales que $M_\pi(\mathbf{X}) = \mathbf{x}$. Representamos en forma paramétrica a dicho conjunto como

$$\{\mathbf{P} = \mathbf{C} + \lambda\tilde{\mathbf{M}}_\pi\tilde{\mathbf{p}} \;\mid\; \lambda \neq 0\}$$

La representación proyectiva de los rayos ópticos es importante no sólo para cuestiones geométricas vinculadas a la proyección, sino también para una representación intrínseca de las propiedades radiométricas vinculadas a renderización (Ray Casting y Ray Tracing). Asimismo, es la clave para relacionar diferentes modelos de proyección (central, omnidireccional, múltiples cámaras sincronizadas).

1.1.2. Métodos efectivos para la puesta en correspondencia

Los métodos generales para identificar elementos homólogos son de tipo estadístico y afectan a procedimientos de búsqueda por barrido de elementos geométricos a lo largo de líneas ó, alternativamente, a la localización de "distribuciones de hechos" con características similares sobre la escena. En ambos casos de trata de elementos $0D$ que se pretende sean tan "densos" como sea posible.

Las estrategias básicas de búsqueda para puntos homólogos se basan en

- *Maximizar la correlación* $C_{\ell,r} = \sum I_\ell I_r$, donde la función I_α se refiere a las 3 componentes de una representación $3D$ del color (en lugar de la intensidad en la escala de grises).
- *Minimizar alguna distancia* $L^i(I_\ell(\mathbf{p}_\ell), I_r(\mathbf{p_r}))$ para distribuciones (clusters) de puntos comunes a las vistas, donde I_ℓ, I_r es la función de intensidad en la escala de grises de las imágenes izquierda y derecha respectivamente.

En ambos casos, se toma la suma correspondiente a la "disparidad" para *todos* los candidatos a puntos homólogos (modulo un umbral de búsqueda) situados a lo largo de líneas que se corresponden en pares de imágenes. Este proceso es elemental sobre vistas rectificadas, pues el desplazamiento se produce sobre líneas horizontales de barrido (como en una aproximación lineal a la visión humana). En este caso una homografía permite restringir el rango de búsqueda de puntos homólogos a una línea horizontal. En el caso general la búsqueda se realiza sobre la correspondiente línea epipolar asociada a una línea de barrido en una imagen [Loo99].

En genera, debido a oclusiones parciales, no es posible realizar una correspondencia píxel a píxel entre todos los puntos candidatos a homólogos, aunque para vistas próximas (pequeña línea base) este procedimiento proporciona resultados muy satisfactorios.

[13] Razonad por qué es no-degenerada

Para identificar hechos salientes y sus homólogos en un par de imágenes, es conveniente identificar características de la distribución asociada a cada cluster de puntos; en este caso se utilizan criterios de "pegado local" sobre trozos de superficies (parches) que se van propagando teniendo en cuenta la "semejanza" entre las distribuciones de datos en imágenes. El procedimiento anterior (correlación) requiere que las vistas sean muy próximas y una elección apropiada del tamaño de la ventana en función de la línea base b y de la escena; por ello, es apropiado para cabezales estéreo montados sobre una plataforma móvil con variabilidad muy controlada para la disparidad,. El procedimiento basado en detección de hechos es más tolerante con cambios en la línea base, en la disparidad y en las condiciones de iluminación; por ello es más flexible (es aplicable a casos en los que el método de correlaciones no funciona) aunque también más propenso a errores.

Cualquiera que sea la estrategia utilizada para la puesta en correspondencia, el primer output de este proceso es un *mapa denso de disparidades* que es la clave inicial para la Visión Estéreo [14].

Maximizando la correlación

para empezar se fija una ventana cuadrada W_k de tamaño $(2k+1)\times(2k+1)$ centrada en un "hecho aislado" $0D$ $\mathbf{p}$ (correspondientes a vértices, junturas, o máximos de intensidad). Supongamos que $\mathbf{p}_i$ y $\mathbf{p}'_i$ son puntos homólogos, es decir, tales que

$$\pi_r(\mathbf{P}_i) = \mathbf{p}_i \quad , \quad \pi_\ell(\mathbf{P}_i) = \mathbf{p}'_i$$

donde π_r es la proyección correspondiente a la cámara derecha y π_ℓ es la proyección correspondiente a la cámara izquierda. Entonces, el tamaño de la ventana a insertar depende de la *disparidad ordinaria* en Visión Estéreo Bicameral que se define como el máximo del par de distancias de la diferencia $\mathbf{x}_i - \mathbf{x}'_i$ medida sobre los ejes coordenados, es decir,

$$d \;:=\; max(x_i - x'_i \,,\, y_i - y'_i)$$

Obviamente, a priori se ignora la profundidad, por lo que el parámetro d debe ser estimado a priori (hay dispositivos de rango (infrarrojos, ultrasonido, láser) que lo proporcionan de forma inmediata) o bien introducido a mano y corregido sobre la marcha. [15]

El desplazamiento de la ventana W_k a lo largo de líneas sobre la imagen permite acotar el rango de búsqueda y propagar la búsqueda reforzando las hipótesis obtenidas en fases anteriores del proceso. El tamaño de la ventana condiciona fuertemente el mapa de disparidades; para una ventana pequeña se obtiene un mapa con alta definición pero elevado coste de procesamiento, mientras que para una ventana grande se obtiene un mapa con baja definición, pero bajo coste de procesamiento. La elección de la ventana depende de los requerimientos de la aplicación y de las características de la escena. De la misma forma que el filtro de medianas permite eliminar eficientemente los efectos de sal y pimienta, los "agujeros" que presenta la puesta en correspondencia de píxeles homólogos en Visión Estéreo deben ser "rellenados" mediante la aplicación de un filtro de medianas.

El método de búsqueda basado en correlación utiliza ventanas a lo largo de líneas de escaneo (método introducido originalmente en [Kan94]). La restricción epipolar proporciona un modelo estructural para estimar dichas líneas; de momento supongamos que ya son conocidas. La *Correlación* se establece originalmente a partir del producto de las intensidades

[14] La versión dinámica de este proceso es crucial para la generación de actores virtuales 3D y se desarrolla en el módulo 5

[15] Actualmente, se están desarrollando dispositivos láser de tiempo de vuelo que proporcionan información relativa a la profundidad en tiempo real para una colección dispersa de puntos correspondientes a objetos en movimiento; los progresos en esta línea son cruciales para estabilizar la información asociada a la Visión Estéreo Dinámica que se aborda en el módulo 5

$$\psi[I_\ell(x,y)\ ,\ I_r(x+d,y)]\ =\ I_\ell(x,y)\ .\ I_r(x-d,y)$$

la correlación máxima corresponde al mejor pegado; la *función de coste a minimizar* es la distancia entre las intensidades asociadas a los puntos situados en el interior de cada ventana a la que se llama *función de disparidad*. En el caso de la *distancia euclídea* L^2, la función de coste a minimizar es

$$SSD(\mathbf{p},d)\ :=\ \sum_{(u,v)\in W_k}[I_\ell(u,v)-I_r(u-d,v)]^2$$

Este enfoque presenta problemas típicos aproximación basada en mínimos cuadrados, es decir, la presencia de outliers da lugar a desviaciones muy significativas con respecto a los valores correctos esperados; en particular, la minimización según SSD penaliza las diferencias relevantes en los valores de la función de intensidad debida a factores no-controlados de la iluminación o al comportamiento no-lambertiano de las superficies que aparecen en imagen, p.e..

Una alternativa más barata desde el punto de vista computacional (evita multiplicaciones asociadas a mínimos cuadrados) consiste en tomar la Suma de Diferencias Absolutas

$$SAD(\mathbf{p},d)\ :=\ \sum_{(u,v)\in W_k}|\ I_\ell(u,v)-I_r(u-d,v)\ |$$

Aún así, esta solución da lugar a la presencia de un gran número de falsos positivos y de falsos negativos. Para tratar de evitar estos problemas es conveniente trabajar con *datos normalizados*, lo cual afecta a los valores de intensidad en cada ventana (imponiendo la condición de tener media 0) o bien al re-escalado de las intensidades de media nula para que tengan un rango de variación similar, es decir, varianza 1. Esta normalización se consigue haciendo

$$I\ :=\ \frac{I-\bar{I}}{\sigma_I}$$

e introduciendo la función de *correlación normalizada*

$$\psi[I_\ell(x,y),I_r(x+d,y)]=\frac{I_\ell(x,y).I_r(x-d,y)-\overline{I}_\ell\overline{I}_r}{\sigma_\ell\sigma_r(d)}$$

que tiene un comportamiento similar a la correlación ordinaria, pero "centrando los valores" mediante una sustracción de la intensidad media de la ventana de la correlación y normalizando por la desviación estándar de la intensidad en la ventana (no la imagen). Tiene la ventaja de que está acotada entre -1 y $+1$, dándose el mejor pegado para $+1$. Presenta el inconveniente de un mayor coste computacional.

Valoración: Los procedimientos descritos proporcionan resultados robustos para pares estéreo (habituales en Fotogrametría) o bien para pares capturados por un cabezal estéreo rígido con una pequeña línea base en Robótica Móvil. A pesar de las mejoras introducidas por la función de correlación normalizada, estos procedimientos son propensos a error en presencia de "amplia línea base" b o bien en presencia de saltos bruscos en la profundidad. Por ello, es conveniente disponer de estrategias alternativas que permitan reajustar la puesta en correspondencia para casos más generales. La figura 1.1.2 muestra las dificultades para la puesta en correspondencia en presencia de un área de solapamiento pequeña (amplia línea base).

Figura 1.4: Puesta en correspondencia de dos imágenes con amplia línea base

Minimizando la distancia entre distribuciones de hechos

El procedimiento basado en detección de hechos sigue una estrategia complementaria basada en minimizar la distancia entre distribuciones $2D$ de puntos (habitualmente vértices o junturas), sin imponer restricciones adicionales sobre el tamaño de la ventana que debemos desplazar sobre líneas de búsqueda (el desplazamiento de ventanas sobre líneas requiere alguna forma de geometría epipolar y falla en presencia de saltos de profundidad a lo largo de líneas epipolares). El procedimiento basado en detección de hechos es compatible con "línea base amplia" b y es asimismo compatible con distribuciones de puntos que puedan presentar saltos en la profundidad.

La mayor dificultad para este procedimiento radica en describir la "proximidad" entre distribuciones de puntos e introducir los criterios de semejanza entre dichas distribuciones. De una manera intuitiva, aunque el espacio de distribuciones de datos no sea una "variedad riemanniana" (M, ds^2) se puede razonar por analogía introduciendo diferentes "criterios métricos" que juegan un papel similar [16].

Algunos de los algoritmos más significativos son SIFT (Scale Invariant Feature Transform) y SURF (Speeded-Up Robust Feature). Estos algoritmos han sido desarrollados originalmente en relación con cuestiones de Reconocimiento (Módulo 4) y recientemente extendidos para estimación de movimiento (Módulo 3). La integración de ambos se lleva a cabo en el módulo 5.

1.1.3. Modelos proyectivos y Reconstrucción Dispersa

La *Geometría Proyectiva Lineal* afecta a *entidades lineales* (dadas por la anulación de expresiones algebraicas de grado 1 en las coordenadas de cada espacio), a las *operaciones* (unión, intersección representadas por suma y productos) y a las *transformaciones regulares* proyectivas (homografías) definidas sobre dichas entidades lineales y sus operaciones.

La *Reconstrucción Proyectiva* consiste en la generación de modelos 3D y nuevas vistas 2D a partir de dos o más imágenes módulo el grupo de las transformaciones proyectivas (homografías) o, con más generalidad, alguno de los subgrupos más significativos (afín o euclídeo). Una diferencia importante con respecto al enfoque presentado en los capítulos anteriores consiste en que las nuevas vistas no se

[16] El desarrollo de esta idea requiere elementos adicionales de Reconocimiento que se presentan con más detalle en el módulo 4

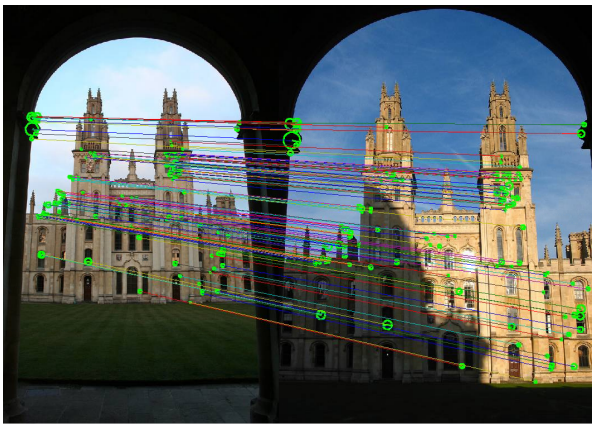

Figura 1.5: Búsqueda de punto homólogos mediante el algoritmo SIFT

obtienen necesariamente como proyección de un modelo 3D, sino que pueden ser generadas a partir de correspondencias entre vistas consideradas como planos proyectivos.

En esta subsección se revisan algunos aspectos de la Geometría Proyectiva Lineal que son significativos para la Reconstrucción Proyectiva; la significación procede de la utilización de haces Λ_α de líneas $\ell_{\alpha,i}$ parando un punto similares los haces de líneas de perspectiva; la diferencia fundamental consiste en que este punto no es un punto de fuga de la escena, sino la proyección del centro de una cámara desde el centro de otra cámara.

La Reconstrucción proyectiva trata no sólo de la recuperación de las coordenadas de "posición relativa" (con respecto a una referencia proyectiva) de una colección significativa (dispersa o densa) de "hechos" geométricos (puntos, líneas, cónicas), sino también del modelado e implementación de las transformaciones regulares (homografías) que actúan sobre dichos "hechos geométricos" y, por consiguiente, de la conservación de las relaciones de incidencias entre los elementos lineales. Los tópicos más significativos que se abordan en el resto de la sección son los siguientes:

- Homografías entre planos para la Reconstrucción: utiliza información minimal y sólo proporciona un esqueleto muy esquemático de la escena.
- Reconstrucción "densa" usando una colección redundante de puntos; más significativa pero carente de estructura.
- Especificar relaciones de incidencia en términos de restricciones epipolares (sección 2 de este capítulo)
- Algoritmo DLT para estimar homografías [17]

El modelo geométrico más simple para la representación perspectiva está dado por una proyección (de una porción) del espacio tridimensional sobre el plano de imagen. Usando coordenadas proyectivas (o euclídeas ampliadas), está representado por un 3×4-matriz. La generación de una nueva vista se lleva a cabo idealmente mediante una transformación regular (homografía) en el plano de imagen modelado como un plano proyectivo.

Ejercicio.- Verificad que bastan 4 puntos y sus imágenes para identificar una transformación proyectiva del plano $\mathbb{P}^2$ (*Indicación.*- Cada par de puntos homólogos impone dos restricciones. Una ho-

[17] Ver sección 4 del capítulo 2 para detalles

mografía del plano está determinada por una 3×3-matriz salvo factor de proporcionalidad, es decir, 8 parámetros salvo factor de proporcionalidad).

Cualquier cámara es un dispositivo de perspectiva pero presenta diferentes tipos de distorsiones (radiales y tangenciales) debidas a las características de la lente y la apertura del diafragma (ver capítulo.11 de OpenCV para un modelo usado frecuentemente utilizado en la implementación, p.e.). Por ello, para una cámara calibrada el único punto de la imagen que no tiene distorsión alguna es el punto central correspondiente a la proyección del foco sobre el plano de la cámara. Sin embargo,

- Para una proyección central tipo pinhole el campo de visión es muy reducido (diafragma casi cerrado en la lente), por lo que se requeriría una cantidad demasiado elevada de vistas para la reconstrucción;
- el punto central de una cámara no se percibe de la misma forma por la otra cámara de un cabezal estéreo y la utilización del "ojo ciclópeo" no es más que una solución particular que "falsea" la reconstrucción;
- una representación en perspectiva sólo facilita una visualización escenarios generados por la acción humana;
- la proyección asociada a un modelo de perspectiva sólo permite realizar una navegación interactiva para generar un modelo $3D$ para la parte visible (ver lección siguiente);
- salvo que se añada información adicional un modelo de perspectiva no proporciona coordenadas $3D$ para los puntos.

La identificación de (una colección redundante de) $N \geq 4$ pares de puntos homólogos $(\mathbf{p}_i, \mathbf{p}'_i)$ para $1 \leq i \leq N$ en diferentes vistas es el primer problema a resolver. Para modelos de perspectiva una primera aproximación al problema consiste en identificar vértices (esquinas o junturas) en imágenes "rectificadas"; una primera aproximación a la "rectificación" de la imagen se realiza con respecto a un "plano dominante" (ortogonal a la línea de visión que se desea privilegiar). Sin embargo, tras la "rectificación" de cada imagen con respecto a dicho plano y la corrección de la distorsión afín para los elementos más próximos, los puntos homólogos situados en otros planos (a diferentes profundidades) están situados en diferentes localizaciones y el pegado resulta fallido. Para obtener la localización (posición y orientación) de las cámaras y las coordenadas de los puntos $3D$ es imprescindible introducir restricciones adicionales para la puesta en correspondencia de una forma automática.

En algunas representaciones básicas para escenas muy sencillas, la estrategia descrita para la rectificación usando un único plano dominante da resultados "aceptables" a bajo nivel. Este enfoque es compatible con la corrección de las distorsiones radial y tangencial presentes en cada vista que se ha mostrado en el capítulo anterior. Una vez realizada, es necesario disponer de una estimación tosca de la distorsión aparente producida por la representación perspectiva asociada a la proyección central correspondiente a cada cámara. Un ejemplo típico es la navegación automática en escenas de pasillo (fácilmente extensible a escenarios tipo Manhattan) tal y como se muestra en la figura 1.1.3 del pasillo de Informática (trabajo conjunto con Margarita Gonzalo)

Sin embargo, esta solución no es satisfactoria en presencia de escenarios más complejos con saltos en profundidad (como ocurre con escenas de Laboratorio) ni siquiera bajo hipótesis relacionadas con modelos de perspectiva simplificados. Una de las causas es el carácter no-lineal de los modelos de perspectiva. La aproximación lineal a modelos de perspectiva (usando paraperspectiva o perspectiva débil) permite introducir una colección finita de planos de profundidad paralelos (perpendiculares a la línea de visión) que facilita la gestión computacional de la información contenida en la escena. Esta solución es un "parche" pues sólo proporciona soluciones aproximadas cuando la dirección del

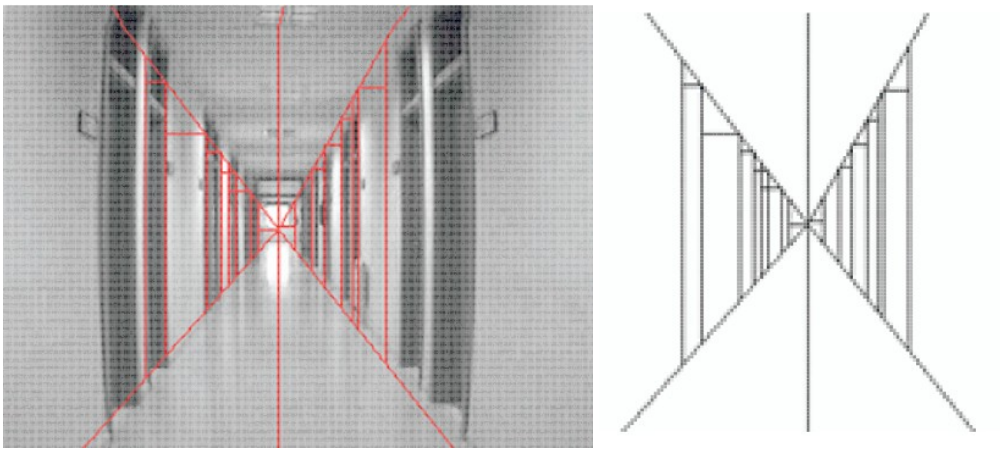

Figura 1.6: Ejemplo de escena tipo Manhattan

desplazamiento es perpendicular a la colección de planos. No obstante, en el apartado siguiente se presentan materiales relacionados pues proporcionan una solución aproximada de bajo coste en escenarios tipo Manhattan con resultados en tiempo real.

Homografías en Reconstrucción Proyectiva

Las homografías permiten realizar transformaciones sobre toda la imagen. Interesa evaluar el comportamiento de los elementos característicos de una escena en perspectiva, así como describir su relación con las proyecciones asociadas a las vistas.

De cara a la reconstrucción $3D$ es importante identificar el plano del infinito. Para una cámara con parámetros intrínsecos constantes, la homografía infinita debe ser conjugada con respecto a una matriz de rotación [Pol97a], condición que se etiqueta como restricción del módulo. La identificación del plano del infinito es crucial para estimar la cónica absoluta que proporciona el nexo entre las reconstrucciones euclídea (capítulo anterior) y afín (el presente capítulo). Cuando los parámetros intrínsecos varían, es necesario realizar una estimación previa de una traslación pura (además de la rotación anterior), con objeto de desacoplar la estimación de la calibración; Heyden et al desarrollaron en la segunda mitad de los noventa varios métodos basados en rectas para estimar la traslación pura y facilitar así la inicialización del proceso de calibración. En los apartados siguientes se revisa el papel que desempeñan las homografías en relación con los modelos de perspectiva y la reconstrucción.

Una *homografía de un espacio proyectivo* real n-dimensional $\mathbb{P}^n = \mathbb{P}(V^{n+1})$ (proyectivizado de un espacio vectorial $n+1$-dimensional) es una $(n+1)\times(n+1)$ matriz regular (es decir, con determinante no-nulo), definida salvo factor de escala. El conjunto de las homografías con la operación producto de matrices es un grupo al que se denota mediante $\mathbb{P}GL(n+1;\mathbb{R}) := GL(n+1;\mathbb{R})/\mathbb{R}^*$ donde $GL(n+1;\mathbb{R})$ es el grupo lineal general (transformaciones regulares o inversibles) y $\mathbb{R}^*$ el subgrupo 1-dimensional de las homotecias representadas por las matrices diagonales λI_{n+1} siendo I_{n+1} la matriz identidad y $\lambda \in \mathbb{R} - \{0\}$.

Lema.- Fijado el plano de proyección Π, la matriz M_π de la proyección central $\pi_{\mathbf{C}}$ da lugar a una homografía del plano proyectivo imagen.

En efecto, eligiendo coordenadas podemos suponer que el plano de proyección Π corresponde al plano $Z = 0$ (como en los grabados de Durero). Por ello, la matriz de la proyección general se escribe ahora como

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{14} \\ m_{21} & m_{22} & m_{24} \\ m_{31} & m_{32} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}$$

con $rk(M^{124}) = 3$, por lo que se trata de una transformación proyectiva (de hecho, la forma general de una homografía del plano imagen).

Recíprocamente, cualquier homografía se puede elevar a una cantidad infinita de proyecciones, dependiendo del plano que fijemos como plano de proyección. A esta "elevación" le llamamos *proyección inversa* [18]. Esta simple observación es clave para comparar los datos contenidos en dos vistas (planos de imagen) correspondientes a elementos de la escena que son coplanarios (método de homografías locales); esta idea se desarrolla en la sección 2 de este capítulo.

Ejercicio.- Describir la elevación de una homografía a una proyección sobre cada uno de los planos coordenados $X_i = 0$

Proposición.- La relación entre los elementos homólogos que son imagen de puntos situados en un mismo plano de la escena (pertenecientes a dos modelos de perspectiva, p.e.) está dada por una transformación proyectiva entre dos planos proyectivos $\mathbb{P}^2$.

Demostración: En efecto, supongamos que $\mathbf{p}_1 \in \Pi_1$, $\mathbf{p}_2 \in \Pi_2$ son homólogos, es decir, existe $\mathbf{P} \in \mathbb{P}^3$ tal que $\pi_1(\mathbf{P}) = \mathbf{p}$ y $\pi_2(\mathbf{P}) = \mathbf{p}_2$. Si $\mathbf{P} \in \Pi$ (plano de la escena), entonces la aplicación proyección π_1 se eleva a una homografía $\mathbf{H}_1$ y la aplicación proyección π_2 se eleva a una homografía $\mathbf{H}_2$ (ver Lema más arriba) de modo que

$$\mathbf{p}_1 = H_1\mathbf{P} \quad \text{y} \quad \mathbf{p}_2 = H_2\mathbf{P}$$

por lo que sustituyendo el valor de la primera en la segunda se obtiene

$$\mathbf{p}_2 = H_2\mathbf{P} = H_2H_1^{-1}\mathbf{p}_1 = H\mathbf{p}_1$$

siendo $H := H_2H_1^{-1}$ la homografía buscada. Obviamente, la misma homografía sirve para todos los puntos que pertenecen al mismo plano que $\mathbf{P}$, por lo que bastan 4 puntos pare determinarla.

Nótese que cada plano proyectivo completa el plano de imagen Π_j añadiendo las rectas del infinito obtenidas conectando por partes 3 puntos de fuga linealmente independientes. La aplicación de las homografías $H = H_2H_1^{-1}$ a los elementos pertenecientes incluso al mismo plano genera distorsiones que es preciso corregir. A mayores, es necesario introducir restricciones estructurales que no dependan de ningún modelo de perspectiva previo, sino tan sólo de la localización relativa de los elementos homólogos. Esta cuestión se resuelve en la sección siguiente.

Los elementos de perspectiva (puntos de fuga y líneas de horizonte) son relativamente fáciles de estimar en escenas arquitectónicas. En particular, la implementación de las transformaciones basadas en homografías permite generar nuevas vistas diferentes de las iniciales, lo cual facilita una visualización de interés para diferentes aplicaciones. La visualización depende del marco geométrico elegido con la jerarquía descrita en el capítulo 1: El marco más débil pero amplio es el proyectivo (incluye a los demás); el más preciso es el euclídeo, pero requiere una verificación de parámetros con mayor coste y presenta un mantenimiento más difícil. El marco afín proporciona una solución intermedia de compromiso, facilitando la conexión entre diferentes perspectivas de una misma escena e incorporando la métrica mediante la "deformación" aparente de una circunferencia.

[18] En ocasiones también se etiqueta como reproyección, pero esta terminología es confusa y no será utilizada

Figura 1.7: Ejemplo de homografía 3D aplicada sobre una escena

Homografías entre planos para la Reconstrucción

Una *homografía* ó *colineación* del plano proyectivo está dada por una transformación regular lineal $\mathbb{P}^2 \to \mathbb{P}^2$ de la forma $\mathbf{y} = A\mathbf{x}$, donde $A = (a_{ij})_{1\leq i,j\leq 3} \in \mathbb{P}GL(3,\mathbf{R})$ es una matriz inversible (es decir, con determinante no-nulo). Esta expresión vectorial se reformula como tres ecuaciones escalares de la forma

$$y_i \;=\; a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 \;=\; \mathbf{a}_i\mathbf{x}$$

para $1 \leq i\ leq3$. Si suponemos que $y_3 \neq 0$, la colineación en $D_+(y_3)$ está dada por las expresiones racionales:

$$\frac{y_1}{y_3} = \frac{\mathbf{a}_1\mathbf{x}}{\mathbf{a}_3\mathbf{x}} \quad , \quad \frac{y_2}{y_3} = \frac{\mathbf{a}_2\mathbf{x}}{\mathbf{a}_3\mathbf{x}}$$

que se pueden reescribir en términos de las coordenadas usuales no-homogéneas como

$$X = \frac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}y + a_{33}} \quad , \quad Y = \frac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}y + a_{33}} \,.$$

Por ello, cada punto y su homólogo determinan dos ecuaciones. Como una colineación está determinada por 8 parámetros salvo factor de proporcionalidad, es necesario considerar cuatro puntos $\{\mathbf{p}_i \mid 1 \leq i \leq 4\}$ en posición general y sus transformadas $\{\mathbf{p}'_i = A(\mathbf{p}_i) \mid 1 \leq i \leq 4\}$ para determinar una homografia ó colineación entre los dos planos de imagen.

De una forma puramente teórica, el carácter homogéneo de $\mathbb{P}^2$ permite seleccionar la referencia estándar dada por $\mathbf{p}_1 = [1:0:0]$, $\mathbf{p}_2 = [0:1:0]$, $\mathbf{p}_3 = [0:0:1]$ y $\mathbf{u} = [1:1:1]$ en el espacio de partida y calcular sus imágenes en el plano proyectivo $\mathbb{P}^2$ de llegada; esta elección simplifica los cálculos.

En la práctica, para automatizar el proceso con una imagen dada es necesario seleccionar puntos (vértices ó esquinas, típicamente), estimar sus coordenadas píxel, convertirlas a coordenadas afines, fijar radio de búsqueda para puntos homólogos, estimar los puntos homólogos, construir la transformación (por el método descrito más arriba) y validarla para una colección lo más densa posible de puntos significativos (vértices ó junturas).

Una extensión al caso 3D

La reconstrucción proyectiva consiste en obtener un conjunto de puntos $3D$ que denotamos mediante $\mathbb{X}_i \in \mathbb{P}^3$ y un conjunto de matrices P_j para las proyecciones $\pi_j : \mathbb{P}^3 \to \mathbb{P}^2$ compatible con los datos contenidos en las imágenes I_j. Por construcción, la reconstrucción proyectiva está determinada salvo homografías $H \in \mathbb{P}GL(4, \mathbb{R})$, es decir, se requieren 15 parámetros para su estimación. Esta condición de invariancia se traduce en varias propiedades:

- La reconstrucción $\{(H\mathbf{X}_i), (P_j H^{-1})\}$ debe ser compatible con los mismos inputs para cualquier transformación proyectiva $H \in \mathbb{P}GL(4, \mathbb{R})$
- *Robustez lineal:* Conservación de relaciones de incidencia: $\mathbf{x}_i \in \ell_{ij} \subset \Pi_{ijk}$
- *Robustez curvada:* Conservación de relaciones de incidencia: $p \in \mathcal{C} \subset \mathcal{S}$

La homografía infinita

El objetivo de la homografía infinita es la estimación del plano del infinito como paso previo para estimar la cónica absoluta que facilita la conexión entre las reconstrucciones euclídea y proyectiva.

Ejercicio (avanzado).- Mostrar un método para estimar dicha homografía.

1.1.4. Reconstrucción densa

La *reconstrucción densa* afecta a la identificación de al menos varios cientos de puntos $3D$ obtenidos a partir del correcto emparejamiento de candidatos a puntos homólogos en dos vistas. La reconstrucción densa requiere localizaciones próximas para la cámara, es decir,

$$d(\mathbf{C}), \mathbf{C}') < \tau_d \quad , \quad \|or(\mathbf{C})) - or(\mathbf{C}')\| < \tau_{or}$$

donde or denota la orientación de la cámara, siendo τ_d y τ_{or} los umbrales (thresholds) fijados por el usuario; con ello, se pretende maximizar la cobertura del objeto o la escena y acotar el número de auto-oclusiones. Las dos formas más frecuentes de expresar la orientación están dadas por los ángulos de Euler o bien los cuaterniones. A la distancia entre localizaciones (posición y orientación) se le llama la *línea base* y se le denota mediante b; cuando las localizaciones correspondientes a las imágenes que se capturan están "próximas", se dice que la línea base es "pequeña". Un ejemplo típico de pequeña línea base viene dado por un muestreo frecuente de las imágenes proporcionadas por una secuencia continua de vídeo.

La reconstrucción densa es opuesta a la reconstrucción minimal que solo afecta al número mínimo de puntos que se requieren para obtener la "estructura" geométrica de la escena (típica en los modelos de perspectiva, p.e.). Asimismo es diferente de la *reconstrucción dispersa* en la que sólo se pretende reconstruir partes del objeto o de la escena que son visibles desde localizaciones alejadas de las cámaras. El pegado y la gestión de oclusiones son más complicados en el caso de amplia línea base que en el caso de pequeña línea base, si bien proporcionan una información más amplia de la escena; como las deformaciones aparentes y las discontinuidades en la disparidad pueden ser mayores es conveniente adoptar una estrategia afín local para la reconstrucción $3D$. En escenas que presentan una gran complejidad desde el punto de vista volumétrico, puede ser útil realizar una reconstrucción dispersa para cada una de las regiones significativas (invariantes desde el punto de vista afín) y proceder en una segunda fase al "pegado" y refinamiento de las reconstrucciones dispersas. En la mayor

parte de esta sección centramos la atención en la reconstrucción densa; un desarrollo más detallado de la reconstrucción densa que combina el enfoque radiométrico y el punto de vista invariante afín se puede ver en [Tuy04]

La Reconstrucción densa es compatible con cualquiera de los marcos geométricos con precisión creciente (proyectivo, afín, euclídeo). En cualquier caso, no se pretende el emparejamiento de "todos" los puntos con homólogo identificable (hay muchos que no tienen homólogo debido a oclusiones, p.e.), sino de una cantidad "suficiente". La nube de puntos $3D$ permite construir mallas (típicamente triangulares) sobre las que se realizar un proceso de "rellenado" (algoritmos de propagación) que completan la información a la resolución deseada. Dependiendo del nivel de detalle requerido, la reconstrucción puede requerir desde el punto dde vista computacional una cantidad prohibitiva de iteraciones o recursos muy elevados para "mover¡¡ de forma coherente la nube con la malla asociada; en consecuencia, es necesario fijar diferentes niveles de detalle.

El *objetivo a medio nivel* es la visualización en términos de imagen o de modelos tridimensionales; para ello es necesario utilizar una cantidad redundante de elementos robustos (métodos probabilísticos). La información contenida en imágenes digitales reales presenta oclusiones parciales, estructuras complicadas para escenas u objetos eventualmente curvados, detalles eventualmente distorsionados y un largo etcétera que requieren un tratamiento denso de la información. Para desarrollar este tipo de reconstrucción se sigue una estrategia de complejidad creciente con diferentes articulaciones entre aspectos locales y globales, incluyendo la conservación de elementos estructurales de carácter afín (elementos del infinito) o métrico (cónica o cuádrica absoluta).

Una estrategia básica ideal

Una estrategia básica ideal para la reconstrucción dispersa consiste en los pasos siguientes:

1. Identificar "hechos geométricos relevantes" para la reconstrucción proyectiva candidatos a homólogos tales como puntos $\{\mathbf{x}_i \mid 1 \leq i \leq N_1\}$, líneas $\{\ell_j \mid 1 \leq j \leq N_2\}$ de $\mathbb{P}^2$ o una combinación de ambas $\{\mathbf{x}_i, \ell_j \mid 1 \leq i \leq N_1, 1 \leq j \leq N_2\}$.

2. *Validación:* Verificar la condición de ser homólogos utilizando restricciones estructurales definidas por matrices (fundamental o esencial) o tensores invariantes más generales.

3. *Visualización estática:* Implementar la acción de homografías $\mathbf{h} \in \mathbb{P}(GL(3;\mathbb{R}))$ sobre líneas ℓ_i: Representar mediante la inversa $\mathbf{h}^{-1}$ de $\mathbf{h} \Rightarrow \mathbf{x}' = \mathbf{h}\mathbf{x} \Rightarrow \ell' = \mathbf{h}^{-1}\ell$

4. *Extensión proyectiva:* Gestionar la información mediante transformaciones definidas sobre haces de rectas, evaluando invariantes, tanto para los asociados a modelos de perspectiva, como para los haces de líneas epipolares.

5. *Refinamiento métrico del modelo:* Introducir información métrica asociada a la acción de homografías extendiendo la acción sobre elementos lineales a una *acción sobre cónicas*: $\mathbf{x}^\top \mathbf{Q}\mathbf{x} = 0$ y *cónicas duales* $\ell^\top \mathbf{Q}^* \ell = 0$ donde la cónica dual $\mathbf{Q}^*$ de una cónica regular $\mathbf{Q}$ está representada por la inversa de la matriz de la cónica $M_{\mathbf{Q}}^{-1}$ [19]

6. Dotar de *Robustez* al modelo tanto en los aspectos relativos al *muestreo* (variantes de Ransac, típicamente), como a la conservación de relaciones de *incidencia*: $\mathbf{x}^\top \ell = 0$. Nótese que cualquier cónica $\mathbf{Q}$ induce una dualidad entre puntos $\mathbf{x}$ y líneas $\ell = \mathbf{Q}\mathbf{x}$ lo cual implica que $\mathbf{Q}' = \mathbf{h}^{-T}\mathbf{Q}\mathbf{h}^{-1}$ y, por tanto, $\mathbf{Q}*' = \mathbf{h}\mathbf{Q}^*\mathbf{h}^\top$

[19] ¿Qué ocurre si Q es una cónica degenerada asociada a dos rectas secantes o una recta doble?

Todos los elementos anteriores deben ser estimados sabiendo que están contaminados por ruido, distorsiones o artefactos inducidos por diferentes herramientas de procesamiento y análisis. Por ello, es necesario identificar el tipo de error, corregirlo y optimizar los procesos que permiten "organizar" la información correspondiente a una cantidad densa de puntos homólogos o su empaquetamiento en términos de primitivas lineales sencillas (dadas por configuraciones de líneas o planos).

Sistemas Coordenados Normalizados

La generación de una nueva vista es el primer paso de la visualización y se lleva a cabo construyendo una matriz $M(\pi'')$ de proyección π. Una primera aproximación al problema consiste en generar vistas intermedias que, idealmente, corresponden a una interpolación continua entre la matrices de proyección de referencia. Para dar forma a esa idea, es conveniente identificar formas canónicas para las matrices de proyección de referencia. Denotemos mediante

$$\tilde{\mathbf{K}} := \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_3^\top & 1 \end{pmatrix}$$

a la (4×4)-matriz correspondiente a un cambio de coordenadas en el complementario del hiperplano del infinito, donde $\mathbf{R}$ y $\mathbf{t}$ representan la posición y la orientación de la cámara respectivamente (parámetros extrínsecos). La transformación $\tilde{\mathbf{R}}$ representa una colineación que conserva el plano del infinito y la cónica absoluta.

Elijamos ahora un sistema coordenado de modo que la matriz de proyección se escriba $M(\pi_j) = (\mathbf{I}_3\ \mathbf{0})$. En el sistema normalizado $\mathbf{C}_1 xyz$, la matriz $M(\pi_2)$ correspondiente a la segunda proyección se escribe

$$M(\pi_1)\tilde{\mathbf{K}}^{-1} \ = \ (\mathbf{R}^\top - \mathbf{R}^\top \mathbf{t})$$

lo cual permite expresar los *epipolos* en este nuevo sistema coordenado como

$$\tilde{\mathbf{e}}_{12} \ = \ M(\pi_1)\begin{pmatrix} \mathbf{C}_1\mathbf{C}_2 \\ 1 \end{pmatrix} \ = \ M(\pi_1)\begin{pmatrix} \mathbf{t} \\ 1 \end{pmatrix} = \ \mathbf{t} \quad , \quad \tilde{\mathbf{e}}_{21} \ = \ \tilde{\mathbf{P}}_2\begin{pmatrix} \mathbf{0}_3 \\ 1 \end{pmatrix} \ = \ -\mathbf{R}^\top \mathbf{t}$$

y la *línea epipolar* correspondiente a $\mathbf{m}_1 = (u_1, v_1, 1)^\top)$ como $\mathbf{m}_2 = (u_2, v_2, 1)^\top)$, por lo que

$$\tilde{\mathbf{e}}_{21} \wedge M(\pi_2)M(\pi_1)^{-1}\tilde{\mathbf{m}}_1 \ = \ -\mathbf{R}^\top \mathbf{t} \wedge \mathbf{R}^\top \tilde{\mathbf{m}}_1$$

que es proyectivamente equivalente a $\mathbf{R}^\top(\mathbf{t}\wedge\tilde{\mathbf{m}}_1)$. Análogamente, la línea epipolar correspondiente a $\mathbf{m}_2 = (u_2, v_2, 1)^\top)$ está dada por

$$\tilde{\mathbf{e}}_{12} \wedge M(\pi_1)M(\pi_2)^{-1}\tilde{\mathbf{m}}_2 \ = \ \mathbf{t} \wedge \mathbf{R}\tilde{\mathbf{m}}_2$$

Discretización del movimiento para la simulación

La simulación del movimiento de cámara generada por un movimiento de ratón se puede modelar como una interpolación continua entre vistas o como una navegación interactiva en torno a un modelo $3D$ del objeto o de la escena. La segunda aproximación requiere herramientas más avanzadas de modelado que se abordan más adelante y con más extensión en el Curso sobre Visualización. Por ello, en este apartado se adopta el primer punto de vista. La interpolación entre vistas se puede formular asimismo de dos formas diferentes, según que atendamos a la información contenida en vistas (como

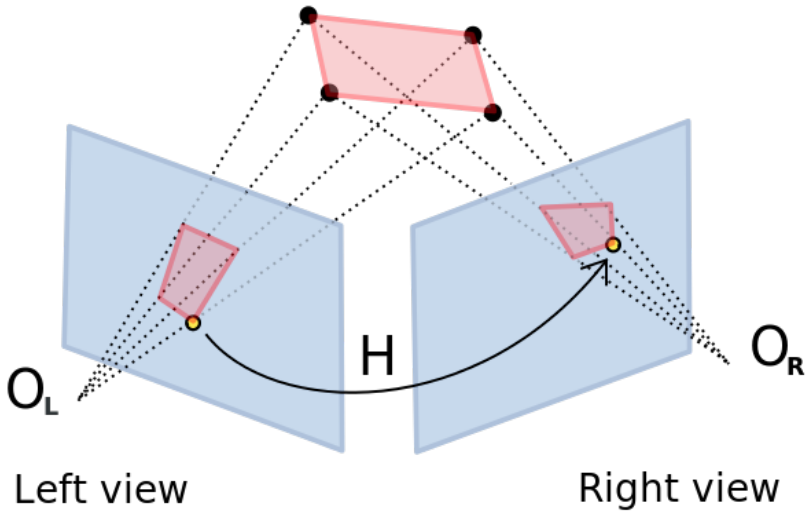

Figura 1.8: Un objeto visto desde dos cámaras diferentes se ve afectado por unas transformaciones que llevan sus vértices de una vista a otra (homografía)

imágenes de una proyección) o bien a la información contenida en las aplicaciones proyección como (3×4)-matrices; la segunda aproximación se desarrolla en el capítulo siguiente. La restricción epipolar proporciona la clave para generar nuevas vistas a partir de dos imágenes "próximas" sin necesidad de disponer de un modelo $3D$.

Por consiguiente, la restricción epipolar proporciona una aproximación a la discretización del movimiento que permite conectar dos vistas iniciales; en el módulo 3 (movimiento) se muestra cómo la restricción epipolar proporciona una restricción estructural para recuperar la estructura (SFM: Structure from Motion) o la forma (sfm: shape from motion) a partir del movimiento. De una forma simplificada, el movimiento se puede entender de forma literal en términos de una cámara de vídeo móvil o bien en términos de una colección de imágenes capturadas por posiblemente diferentes cámaras. El paso del primer enfoque al segundo se lleva a cabo en términos de una discretización del movimiento; tiene la ventaja de la invariancia de los parámetros intrínsecos de la cámara, por lo que basta calcular el cambio en la localización, aunque la iluminación pueda cambiar. Este apartado está centrado en el segundo enfoque.

El cambio en la localización (posición y orientación) de la cámara se modela en términos de una rotación y una traslación en el espacio. La composición de estas transformaciones da lugar a una distorsión aparente sobre los elementos comunes en la imagen. Si se selecciona una región planar de la imagen, el modelo más simple para dicha distorsión, está dado por una transformación del plano proyectivo que se asocia a cada cámara. Obviamente, la transformación depende de la región seleccionada.

Una *discretización lineal* del camino $\gamma(\lambda) : (1-\lambda)I + \lambda H$ que conecta la imagen original (correspondiente a la transformación dada por la matriz identidad I) con la rectificada (asociada a la homografía H) proporciona una aproximación lineal de primer orden a una simulación de las transformaciones en la imagen. La discretización se realiza introduciendo una partición $\cup_{i=0}^{N}[\lambda_i, \lambda_{i+1}]$ con $\lambda_0 = 0$ y $\lambda_N = 1$. Salvo que H esté muy próxima a I, este método no da buenos resultados porque el camino $Im(\gamma(\lambda))$ no tiene por que estar contenido en el grupo de transformaciones; por consiguiente, los elementos $\gamma(\lambda_i)$ no corresponden necesariamente a movimientos de cámara representados por un grupo

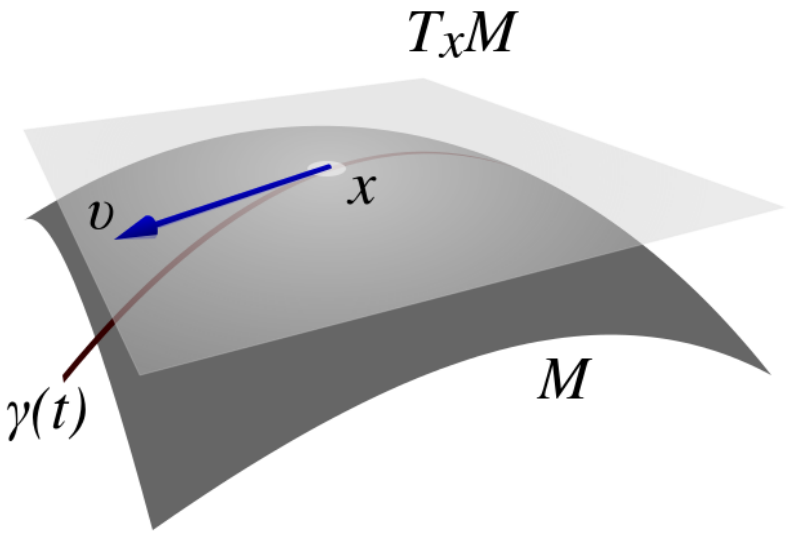

Figura 1.9: Representación del espacio tangente a una variedad en un punto

g de transformaciones.

Linealización en el espacio tangente

Para resolver un problema de optimización sobre un grupo de transformaciones es necesario encontrar el camino γ que minimiza un funcional (distancia, energía, etc); como el ambiente es no-lineal debemos pasar al espacio tangente de la variedad. En este caso, la variedad es el grupo clásico G (especial ortogonal, afín, proyectivo) y su linealización en cualquier punto es el "trasladado" del espacio tangente $\mathfrak{g} := T_I G$ al grupo G en la matriz identidad I(elemento neutro para la multiplicación). La elevación del camino óptimo γ al espacio tangente(álgebra de Lie) proporciona una dirección que ahora sí representa "desplazamientos infinitesimales" de cámara. En la figura 1.1.4 se muestra una representación ideal del espacio tangente a una variedad en un punto y del vector que representa el espacio tangente $\mathbf{v} = t_{\mathbf{p}}\gamma \subset T_{\mathbf{p}}M$ a una curva γ (que representa el camino óptimo buscado) en el punto $\mathbf{p} \in M$

Para fijar ideas, consideramos en primer lugar el *primer ejemplo* más sencillo para cuestiones de reconstrucción que corresponde a las rotaciones globales de (un cuadrilátero contenido en) una imagen. Dichas rotaciones están representadas por matrices

$$\begin{pmatrix} cos\,\theta & -sen\,\theta \\ sen\theta & cos\,\theta \end{pmatrix}$$

rotación plana de ángulo θ que se representa en forma compleja mediante el producto por $z = e^{i\theta}$. La diferencial de la aplicación asociada a la multiplicación modifica la función original en (la adición de) un ángulo θ. Dicho de otra forma: la multiplicación de matrices en el grupo original se traduce en una adición de parámetros sobre el espacio tangente.

Un *segundo ejemplo* menos trivial está asociado a las rotaciones de cámara en cada uno de los

planos coordenado $z = 0$, $y = 0$ o $x = 0$ están representadas por matrices

$$\begin{pmatrix} c_\theta & -s_\theta & 0 \\ s_\theta & c_\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} c_\theta & 0 & -s_\theta \\ 0 & 1 & 0 \\ s_\theta & 0 & c_\theta \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_\theta & -s_\theta \\ 1 & s_\theta & c_\theta \end{pmatrix}$$

donde $c_\theta = cos\ theta$ y $s_\theta = sen\ theta$; el espacio tangente en la matriz identidad I_3 es la matriz

$$\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

(evaluar el desarrollo de Taylor de cada función trigonométrica para $\theta = 0$). Por ello, la rotación infinitesimal con respecto a cualquiera de los ejes coordenados está representada por una matriz antisimétrica X; su exponencial

$$A_t := exp(tX) = I + tX + \frac{t^2}{2!}X^2 + \ldots + \frac{t^n}{n!}X^n + \ldots$$

proporciona una rotación ordinaria que se puede representar como un subgrupo uniparamétrico (dependiente de t). Nótese que el ejemplo anterior es un caso particular de este último (razonadlo como ejercicio).

Simulando movimientos de cámara

En particular, la simulación de movimientos rotacionales de cámara se puede realizar en los pasos siguientes:

1. Calcular el espacio tangente $T_I SO(3;\mathbb{R})$ en la identidad (álgebra de Lie) al grupo de las rotaciones y verificar que está formado por matrices antisimétricas.
2. Resolver la representación de la descomposición en el espacio tangente $\mathfrak{g}$
3. Discretizar el resultado en términos de una poligonal contenida en el espacio tangente.
4. Integrar el resultado (paso del álgebra $\mathfrak{g}$ al grupo G mediante la aplicación exponencial $exp : \mathfrak{g} \rightarrow$ dada por $X \mapsto exp(tX)$

A pesar de la aparente complicación, este procedimiento es más simple que otros, pues cada paso depende de un único parámetro. El desacoplamiento entre las diferentes rotaciones se lleva a cabo mediante un análisis de componentes principales sobre el espacio tangente al grupo $SO(3;\mathbb{R})$ de las rotaciones del espacio ordinario $\mathbb{R}^3$

Ejercicio.- Calcula la exponencial de la matriz X_i que representa cada transformación descrita en los 4 pasos anteriores y asociada a cada uno de los ejes coordenados.

En particular, la *interpolación lineal* entre vistas debe realizarse *siempre* en el espacio tangente al grupo (álgebra de Lie $\mathfrak{g}$ del grupo G) integrando el resultado para su visualización en el grupo; esta estrategia se desarrolla más adelante con mayor generalidad en relación con cuestiones la recuperación de la estructura o de la forma a partir del movimiento y el análisis computacional de movimientos (capítulo 5 del módulo 2 con desarrollos adicionales en los módulos 3 y 5 del CEViC).

La composición de caminos en el grupo realizada utilizando la exponencial de elementos del álgebra debe tener en cuenta la falta de conmutatividad que se expresa en términos de la *fórmula de Baker-Campbell-Hausdorff* y que escribimos infinitesimalmente como

$$exp(tX).exp(tY) = exp(t(X+Y) + o(t^2)$$

(para más detalles ver p.e. Sattinger y Weaver). Por ello, la aplicación reiterada de dicha composición puede dar lugar a errores que degradan el pegado de datos homólogos [20]

[20] Esta situación es especialmente crítica para la composición de movimientos de cámara asociados a la previsualización de un rodaje, p.e.

1.2. Geometría Epipolar

La *Geometría Epipolar* extiende la Geometría Proyectiva utilizada en los dos capítulos anteriores, proporciona el marco para una visualización que extiende los modelos de perspectiva $2{,}5D$ (capítulo 1) y facilita la generación o síntesis de nuevas vistas de forma interactiva mediante movimientos de ratón. La restricción epipolar en el caso afín (resp. euclídeo) está dada por la *Matriz Fundamental* **F** (resp. la *Matriz Esencial* **E**). La generación de nuevas vistas se lleva a cabo construyendo elementos homólogos en planos de imagen "sintéticos" que verifican relaciones de incidencia y paralelismo (versión afín) o de ortogonalidad (versión euclídea) entre puntos, líneas y cónicas. Una vez especificado el modelo de la restricción, el problema más importante es la *estimación robusta* de la matriz (fundamental o esencial).

El problema de generar nuevas vistas a partir de dos o más disponibles debe resolver previamente la puesta en correspondencia entre elementos homólogos. La aproximación (bi)lineal a este problema se expresa mediante una matriz (fundamental o esencial). Esta matriz representa condiciones de incidencia entre elementos homólogos que se expresan algebraicamente mediante una forma bilineal con restricciones adicionales. El conjunto de dichas relaciones bilineales con restricciones adicionales es la variedad fundamental $\mathcal{F}$ (para el caso afín) o la variedad esencial $\mathcal{E}$ (para el caso euclídeo). Necesitamos obtener nuevas vistas de una forma tan rápida y robusta como sea posible; esto significa que el problema de la generación de nuevas vistas se convierte en un problema de optimización sobre $\mathcal{F}$ (cuyos elementos son matrices fundamentales **F**) o sobre $\mathcal{E}$ (cuyos elementos son matrices esenciales **E**). Una vez identificadas las propiedades de $\mathcal{F}$ o de $\mathcal{E}$, el problema más importante es describir procedimientos eficientes para la *optimización* sobre $\mathcal{E}$.

La *Geometría Epipolar* es el estudio de las propiedades geométricas invariantes asociada a los elementos comunes de un par de vistas próximas proporcionadas por dos cámaras idénticas o la discretización del movimiento de una cámara no necesariamente calibrada. Para ello, utiliza relaciones estructurales (independientes de clase de cámaras) entre pares de puntos homólogos contenidos en las vistas.

En la sección precedente se han presentado modelos y algoritmos para imágenes en las que se dispone de primitivas básicas correspondientes a puntos y líneas procedentes del análisis de imagen. Sin embargo, hay situaciones (escenarios naturales, p.e.) en los que no se dispone de líneas largas que permitan "organizar" la escena a partir de varias imágenes (exploración espacial, p.e.). Para llevar a cabo una reconstrucción $3D$ de estos casos, es necesario realizar una puesta en correspondencia entre elementos homólogos con "escasa" o "poco organizada" información geométrica. El cálculo de los puntos $3D$ proporciona información sobre las matrices de proyección. Por último, la generación de nuevas vistas debe ser compatible con la restricción epipolar, lo cual suministra criterios de optimización sobre el conjunto de matrices que soportan esta información.

De forma natural, se plantea la cuestión siguiente: ¿Cómo identificar el homólogo de cada punto "significativo"? ¿Es posible realizar una aproximación similar a la correspondiente a mapas de perspectiva para objetos "naturales" a partir de datos que no contienen elementos generados por el hombre?. La respuesta procede de la Geometría Proyectiva que proporciona el marco para "pegar" diferentes vistas de un objeto o de una escena mediante la detección y puesta en correspondencia de una "cantidad redundante" (es decir, no-minimal) de "hechos salientes" (junturas, p.e.).

Los elementos más relevantes que proporcionan el hilo conductor para las subsecciones de esta sección son los siguientes:

1. *Rectas epipolares:* A un punto $\mathbf{p} \in \Pi$ de una imagen le corresponde una línea $\ell' \subset \Pi'$ en la imagen a la que se llama su recta epipolar.

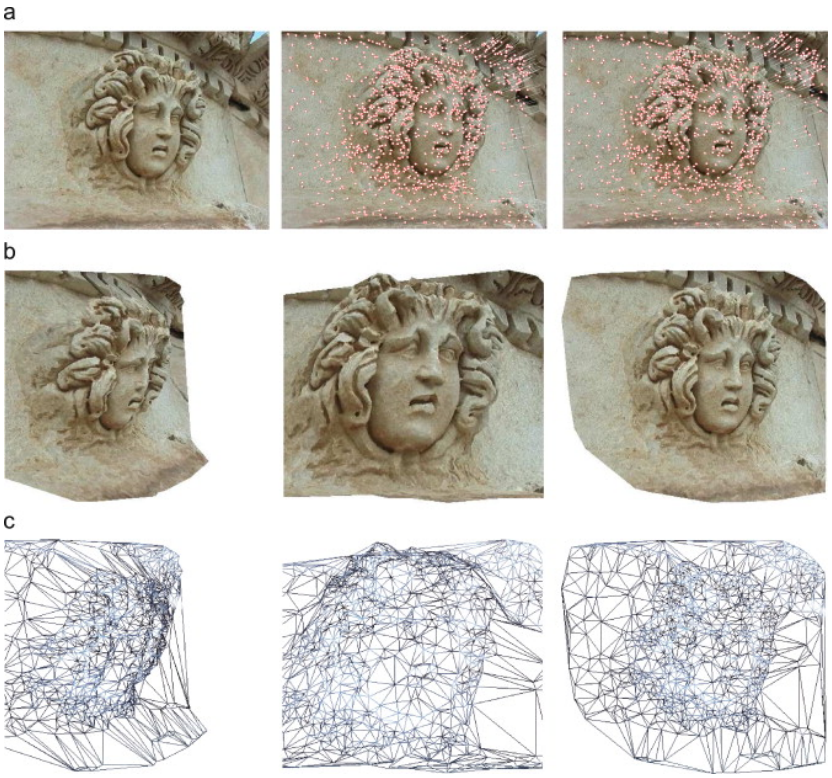

Figura 1.10: Resultados de reconstrucción 3D de la cara de Medusa a partir de dos vistas basados en modelos de proyección en cuasi-perspectiva

2. *Matriz Fundamental* **F**: Proporciona la restricción estructural $\mathbf{x}^\top \mathbf{F}\mathbf{x}' = 0$ entre puntos homólogos para la generación de nuevas vistas y la Reconstrucción *Afín o Proyectiva*, donde $\mathbf{x}$ son las coordenadas de $\mathbf{p}$ y $\ell' := \mathbf{F}\mathbf{x}'$ su recta epipolar.

3. Matriz Esencial **E**: Proporciona la restricción estructural $\mathbf{x}^\top \mathbf{E}\mathbf{x}' = 0$ para la generación de nuevas vistas y la Reconstrucción *Euclídea*.

La estrategia que se desarrolla en esta sección está basada en explotar propiedades de la *restricción epipolar*; esta restricción se expresa mediante una relación bilineal $\mathbf{p}^\top \mathbf{F}\mathbf{p}' = 0$ entre puntos homólogos, donde **F** es la matriz fundamental. Esta estrategia (desarrollada inicialmente por Longet-Higgins, 1981) permite no sólo obtener las coordenadas $3D$ de puntos salientes, sino también el movimiento de cámara (rotación y traslación) y los parámetros intrínsecos de la cámara.

En toda la sección y el resto del capítulo, se supone que se tiene una *hipótesis de rigidez* sobre objetos y escena, es decir, lo único que cambia es la localización (posición y orientación) de la cámara. En esta sección se supone asimismo que la cámara es siempre la misma y que sus parámetros intrínsecos permanecen constantes.

1.2.1. Nociones básicas

El emparejamiento de puntos o de segmentos se lleva a cabo en el marco proyectivo marco sobre líneas epipolares. En esta subsección se presentan los conceptos básicos. Aun a riesgo de ser redun-

dantes empezamos recordando notación para pasar a continuación a describir las propiedades de incidencia en las que se apoya too el capítulo.

Consideremos un punto $\mathbf{P} \in \mathbb{P}^3$ que se proyecta sobre sobre $\mathbf{p}_i = \pi_j(\mathbf{P}) \in \Pi_j$ donde Π_j es el plano de la j-ésima cámara y $\pi_j : \mathbb{P}^3 \to \Pi_j$ es la proyección sobre el plano de la cámara desde el centro $\mathbf{C}_j$ de la cámara.

Dadas dos imágenes procedentes con sus proyecciones correspondientes π y π' llamamos *epipolo en* el plano Π (respectivamente Π' , y lo denotamos mediante **e**. a la proyección $\mathbf{e} := \pi(\mathbf{C}') \in \Pi$ (respectivamente $\mathbf{e}' := \pi'(\mathbf{C}) \in \Pi'$) del centro óptico sobre el plano del otro centro óptico.

Cuando se tiene una colección de n vistas con sus correspondientes aplicaciones proyección π_i, se obtiene una colección de epipolos que denotamos mediante $\mathbf{e}_{ij} := \pi_i(\mathbf{C}_j) \in \Pi_i$ donde Π_i es el plano de la i-ésima proyección y $\mathbf{C}_j$ el centro de la j-ésima proyección π_j. De momento, nos restringimos al caso de sólo dos cámaras, aunque se hará uso de la notación general cuando sea preciso [21]

Por construcción la recta $\mathbf{e} \times \mathbf{e}' = \mathbf{C} \times \mathbf{C}'$ (que contiene a la "línea base" $b = d(\mathbf{C}, \mathbf{C}')$) corta al plano Π en el punto **e** y al plano Π' en el punto $\mathbf{e}'$. Por ello, $\mathbf{e} \times \mathbf{e}' = \mathbf{C} \times \mathbf{C}'$.

El plano $< \mathbf{C}, \mathbf{C}', \mathbf{P} >$ corta al plano Π a lo largo de la línea $\ell = \mathbf{e} \times \mathbf{p} \subset \Pi$ donde $\mathbf{p} = \pi(\mathbf{P})$ y $\mathbf{e} = \pi(\mathbf{C}')$. Análogamente, el plano $< \mathbf{C}, \mathbf{C}', \mathbf{P} >$ corta al plano Π' a lo largo de la línea $\ell' = \mathbf{e}' \times \mathbf{p}' \subset \Pi'$ donde $\mathbf{p}' = \pi'(\mathbf{P})$ y $\mathbf{e}' = \pi'(\mathbf{C})$.

Cuando **P** varía en el espacio tridimensional, el plano $< \mathbf{C}, \mathbf{C}', \mathbf{P} >$ describe un haz Λ con eje la recta $< \mathbf{C}, \mathbf{C}' >$. La intersección del haz Λ con el plano Π (resp. Π') proporciona un haz de rectas $\lambda = \{\ell_i := \mathbf{e} \times \mathbf{p}'_i\}_{i \in I} \subset \Pi$ (resp. $\lambda' = \{\ell_i := \mathbf{e}' \times \mathbf{p}'_i\}_{i \in I} \subset \Pi'$) que pasan por el epipolo **e** (resp. el epipolo $\mathbf{e}'$). Si el epipolo **e** está a distancia infinita (en este caso, las rectas son paralelas) se obtiene un modelo de cámara afín.

Ejercicio.- Expresar vectorialmente y analíticamente todas las condiciones de incidencia descritas en los tres párrafos anteriores.

Ejercicio.- Verificad que a un punto de la imagen izquierda le corresponde una línea de la imagen derecha (o viceversa) utilizando un argumento sintético. (*Indicación:* Usar el plano epipolar determinado por la línea base y el punto **P**; todos los puntos de la línea $\mathbf{C} \times \mathbf{P}$ se proyectan sobre $\mathbf{p} \in \Pi$. La proyección de esta línea sobre Π' desde $\mathbf{C}'$ proporciona la recta epipolar buscada. Este argumento se extiende a la segunda recta y la primera cámara). Más abajo se muestra una demostración analítica.

Invirtiendo la matriz de proyección

Si la matriz $\mathbf{M}_j$ es conocida, las coordenadas $3D$ de cada centro óptico $\mathbf{C}_j$ se obtienen resolviendo el sistema homogéneo $\mathbf{P}_i\mathbf{M} = 0$; en particular, si conocemos las características de la cámara (caso calibrado), como $rango(\mathbf{M}_j) = 3$, el centro óptico está dado como

$$\mathbf{C}_j = -\pi_j^{-1}\mathbf{p}_i$$

donde la inversión de la proyección se expresa en la forma vectorial presentada al final de la primera subsección (ver §1,1,4 para detalles). Las coordenadas de $\mathbf{C}_j$ permiten calcular cada epipolo $\mathbf{e}_{ij} := \Pi_i(\mathbf{C}_j)$.

En el caso bicameral a cada punto $\mathbf{m}_1 \in \Pi_1$ le corresponde una línea epipolar que se obtiene conectando el epipolo $\mathbf{e}_{21} \in \Pi_2$ con el punto de intersección $\mathbf{m}_2$ de la línea $< \mathbf{e}_{12}, \mathbf{m} >$ con el plano Π_2. Como

[21] El apéndice a este capítulo aborda el caso correspondiente a tres o más vistas capturadas por dispositivos no calibrados; en el capítulo 5 se desarrolla estrategias para recuperar la estructura a partir de una cámara en movimiento

$$\mathbf{e}_{21} = \tilde{\mathbf{P}}_2 \begin{pmatrix} -\mathbf{P}_1^{-1}\tilde{\mathbf{p}}_1 \\ 1 \end{pmatrix} \quad , \quad \tilde{\mathbf{m}}_2 = \mathbf{P}_2\mathbf{P}_1^{-1}\tilde{\mathbf{m}}_1$$

lo cual permite *calcular las líneas epipolares* en términos del producto vectorial $\mathbf{e}_{21} \wedge \tilde{\mathbf{m}}_2$. Este producto vectorial se representa también como

$$\mathbf{F}\tilde{\mathbf{m}}_1 \quad \text{donde} \quad \mathbf{F} = \tilde{\mathbf{E}}_2\mathbf{P}_2\mathbf{P}_1^{-1} \ ,$$

siendo $\tilde{\mathbf{E}}_2$ la matriz antisimétrica que representa el producto vectorial con $\mathbf{e}_{21}$, es decir, $\tilde{\mathbf{E}}_2\mathbf{x} := \mathbf{e}_{21} \wedge \mathbf{x}$. De este modo, se obtiene la *ecuación de Longet-Higgins:*

$$\tilde{\mathbf{m}}_2^\top \mathbf{F}\tilde{\mathbf{m}}_1 \ = \ 0$$

que relaciona las coordenadas de un punto de un plano óptico (p.e. $\tilde{\mathbf{m}}_1$) con los del vector director $\mathbf{F}^\top\tilde{\mathbf{m}}_2$ de la línea epipolar que le corresponde.

Ejercicio.- Adaptad la notación de §1,1,4 para calcular epipolos y líneas epipolares. *Indicación.-* El segundo epipolo está dado por

$$\tilde{\mathbf{e}}_2 \ = \ \mathbf{M}_{\pi_2} \begin{pmatrix} \mathbf{C}_1 \\ 1 \end{pmatrix}$$

por lo que las ecuaciones paramétricas de la línea epipolar del punto $\mathbf{p}_1 \in V_1$ se escriben en forma paramétrica como

$$\mathbf{p}_2^\top \ = \ \tilde{\mathbf{e}}_2 + \lambda\tilde{M}_{\pi_2}\tilde{M}_{\pi_1}^{-1}\tilde{\mathbf{p}}_1$$

donde $\tilde{M}_{\pi_i}$ denota la 3×3 matriz correspondiente a la primera caja de la matriz de proyección y $\tilde{p}$ a las 3 primeras componentes de la representación afín de cada punto. Si ahora hacemos $\tilde{v} := \tilde{M}_{\pi_2}\tilde{M}_{\pi_1}^{-1}\tilde{\mathbf{p}}_1$, la igualdad vectorial anterior se escribe en términos de coordenadas afines (u, v) de la imagen correspondiente a la segunda vista como

$$u = \frac{[\tilde{e}_2]_1 + \lambda[\tilde{v}]_1}{[\tilde{e}_2]_3 + \lambda[\tilde{v}]_3} \quad , \quad v = \frac{[\tilde{e}_2]_2 + \lambda[\tilde{v}]_2}{[\tilde{e}_2]_3 + \lambda[\tilde{v}]_3}$$

donde $[]_i$ denota la i-ésima componente en la representación afín del punto proyectivo.

Restricción Epipolar y Orientación

La geometría epipolar sólo depende de datos contenidos en imagen y no de la estructura de la escena (esta última es clave para los modelos de perspectiva, p.e.). Por ello se aplica a cualquier par de imágenes con "suficiente" área de solapamiento. Par ala aplicación de este modelo es crucial que que la línea base sea "pequeña", incluyendo la condición de proximidad entre las orientaciones; de este modo se pueden incorporar procedimientos de búsqueda basados en ventanas de búsqueda sobre líneas epipolares (criterios de proximidad) y en en criterios de ordenación (restricción de orden) con los correspondiente algoritmos descritos en el módulo 1.

La restricción epipolar es compatible con la geometría asociada a los haces de rectas λ y λ' que pasan por cada epipolo $\mathbf{e}$ y $\mathbf{e}'$. Un criterio más débil que la restricción epipolar utiliza transformaciones entre "paquetes" de hechos geométricos, donde es necesario estimar características de la distribución de estos hechos.

La recuperación de la estructura $3D$ de la escena se puede realizar modulo una transformación afín (propuesto inicialmente en [Koe91]), o proyectiva (propuesto inicialmente en [Fau92], [RF93],

entre otros) cuando conocemos la geometría epipolar, aunque desconozcamos los parámetros intrínsecos (caso no-calibrado) o de orientación de la cámara (caso débilmente calibrado).

En un sistema binocular, la orientación extrínseca de la cámara está restringida: tiene esencialmente sólo tres grados de libertad representados por tres ángulos (etiquetados en ocasiones como vergencia, mirada y elevación; ver más abajo) que varían de forma continua cuando el sistema cambia su punto de fijación en el campo visual. Por ello, estos parámetros deben ser evaluados de forma dinámica e independiente usando la información proporcionada por las imágenes, es decir, no pueden ser determinados (ni siquiera de forma aproximada) cuando conocemos la calibración de la cámara. La identificación de puntos de referencia en la escena es crucial para la estimación del posicionamiento (posición-orientación) del sistema bicameral.

Ejercicio.- Calcular la orientación de la línea epipolar en términos vectoriales y paramétricos. (*Indicación.-* Calcular las derivadas (u_λ, v_λ) con respecto a λ de las expresiones vectoriales y paramétricas obtenidas en el ejercicio del apartado anterior). Verificad que si $[\tilde{\mathbf{e}}_2]_3 = 0$, entonces el epipolo $\mathbf{e}_2$ se va al infinito y las líneas epipolares son paralelas (*Indicación:* Comprobad que

$$(u_\lambda, v_\lambda) \;=\; -([\frac{\tilde{\mathbf{e}}_2]_1}{[\tilde{\mathbf{v}}]_3}[\frac{\tilde{\mathbf{e}}_2]_2}{[\tilde{\mathbf{v}}]_3}) \;=\; (\tilde{\mathbf{e}}_2]_1, \tilde{\mathbf{e}}_2]_2)$$

que no depende de $\tilde{\mathbf{v}}$). Este proceso se aplica igualmente a la primera imagen, por lo que las líneas epipolares se pueden convertir siempre en líneas paralelas; este proceso recibe el nombre de *rectificación simultánea* y se presenta con más detalle más abajo (ver §2,3 de este capítulo para más detalles).

Descripción de la Geometría Epipolar

Para minimizar el uso de subíndices denotamos mediante Π y Π' a los planos de proyección de cada una de las cámaras $\mathbf{C}$ y $\mathbf{C}'$ [22]. Definimos:

- El *epipolo* $\mathbf{e}$ (resp. $\mathbf{e}'$) es el punto de intersección de la recta $\mathbf{C}\times\mathbf{C}'$ (línea base) que une los dos centros de cámara y el plano de la cámara $\Pi_{\mathbf{C}}$ (respectivamente, $\Pi_{\mathbf{C}'}$. Se puede interpretara como la proyección de $\mathbf{C}$ desde $\mathbf{C}'$ (resp. proyección de $\mathbf{C}'$ desde $\mathbf{C}$)
- Un *plano epipolar* en un plano que contiene la línea base dada por la recta $\mathbf{C}\times\mathbf{C}'$ que une los dos centros de cámara a la que se denota mediante b. Los planos epipolares forman un haz de planos Λ_b que es isomorfo a una recta proyectiva $\mathbb{P}^1$.
- Cada plano epipolar se determina por la línea base $\mathbf{C}\times\mathbf{C}'$ y un tercer punto no colineal $\mathbf{P}_i$. La proyección de $\mathbf{C}\times\mathbf{P}_i$ desde $\mathbf{C}'$ da la recta $\mathbf{e}\times\mathbf{p}_i$. La proyección de $\mathbf{C}'\times\mathbf{P}_i$ desde $\mathbf{C}$ da la recta $\mathbf{e}'\times\mathbf{p}'_i$.
- Una *recta epipolar* ℓ_i (resp. ℓ'_i es la intersección de un plano epipolar Π_α el plano de cámara Π (resp Π'). Cuando Π_α varía en el haz de planos epipolares, su intersección sobre cada plano de cámara da lugar a un haz de rectas epipolares que denotaremos mediante $\lambda_\alpha := (\ell_i)_{i\in\alpha}$ y análogamente para el otro haz $\lambda'_\alpha := (\ell'_i)_{i\in\alpha}$ de líneas epipolares.
- Las rectas epipolares $\mathbf{e}\times\mathbf{p}_i$ y $\mathbf{e}'\times\mathbf{p}'_i$ se cortan en un punto que pertenece a la intersección $\Pi\cap\Pi'$ de los dos planos de las cámaras que calculamos asimismo como $\Pi\times\Pi'$.

La última condición muestra que la relación entre $\mathbf{p}_i$ y su homólogo $\mathbf{p}'_i$ es una relación bilineal. La expresión de esta relación bilineal es la *restricción epipolar*. Esta restricción se expresa de diferentes formas dependiendo del contexto afín o euclídeo elegido para la visualización.

[22] En presencia de varias vistas V_k para $k = 1,\ldots,n$ volveremos a utilizar superíndices para denotar los puntos o las rectas que pertenecen a cada vista

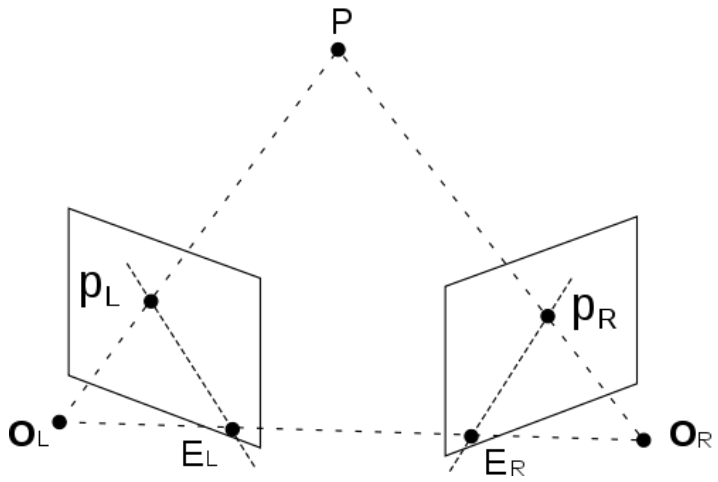

Figura 1.11: Esquema clásico que ilustra los elementos de la geometría epipolar

Cálculo vectorial para líneas epipolares

A continuación se esboza el *cálculo de la Geometría Epipolar* a partir de dos vistas. La clave geométrica para representar la restricción epipolar consiste en utilizar la condición de coplanariedad entre los elementos descritos en el apartado anterior

Denotaremos mediante $\mathbf{P}_i$ a cada punto en el espacio tridimensional con coordenadas $\mathbf{X}_i = [X_{i1} : X_{i2} : X_{i3} : X_{i4}]^\top$; mediante $\pi_{\mathbf{C}} : \mathbb{P}^3 \to \Pi \simeq \mathbb{P}^2$ a la primera proyección sobre el primer plano, mediante $\pi_{\mathbf{C}'} : \mathbb{P}^3 \to \Pi' \simeq \mathbb{P}^2$ a la segunda proyección. Sea $\mathbf{p}_i = \pi_{\mathbf{C}}(\mathbf{P}_i)$ con coordenadas $\mathbf{x}_i = [x_{i1} : x_{i2} : x_{i3}]^\top$ y $\mathbf{p}'_i = \pi_{\mathbf{C}'}(\mathbf{P}_i)$ con coordenadas $\mathbf{x}'_i = [y_{i1} : y_{i2} : y_{i3}]^\top$.

Los *epipolos* se definen como

$$\mathbf{e} = \pi_{\mathbf{C}'}(\mathbf{C}) \quad , \quad \mathbf{e}' = \pi_{\mathbf{C}}(\mathbf{C}')$$

y corresponden a los puntos de intersección de la línea base b dada por $\mathbf{C} \times \mathbf{C}'$ con los planos de imagen de cada cámara:

$$\mathbf{e} = b \cap \Pi \quad , \quad \mathbf{e}' = b \cap \Pi'$$

Las *líneas epipolares* $\ell_i = \mathbf{e} \times \mathbf{p}_i$ del primer haz de líneas se obtienen conectando el epipolo $\mathbf{e}_i \in \Pi_i$ con cualquier "punto saliente" $\mathbf{p}_i \in \Pi$ (vértice o juntura habitualmente, pero también máximos de intensidad) del primer plano. Las *líneas epipolares* $\ell'_i = \mathbf{e}' \times \mathbf{p}'_i$ del segundo haz de líneas se obtienen conectando el epipolo $\mathbf{e}'_i \in \Pi'_i$ con cualquier "punto saliente" $\mathbf{p}'_i \in \Pi'$ del segundo plano.

Denotemos mediante $\mathbf{t}$ al vector traslación $\mathbf{C} - \mathbf{C}'$ y mediante $\mathbf{R}$ a la rotación del rayo $\mathbf{P}_i\mathbf{C}$ en el rayo $\mathbf{P}_i\mathbf{C}'$. Los dos rayos y la recta determinada por el vector traslación $\mathbf{t}$ son coplanarios. Expresamos esta condición mediante

$$\mathbf{x}_i'^T(\mathbf{t} \times \mathbf{R}\mathbf{x}_i) \; = =$$

Se tiene una *correspondencia* $\mathbf{x} \mapsto \ell$ dada por

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} yw - zv \\ zu - xw \\ xv - yu \end{pmatrix} = \begin{pmatrix} 0 & w & -z \\ -w & 0 & u \\ z & -u & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

La correspondencia anterior representa una colineación $\mathbf{A}$ de $(\mathbb{P}^2)^\nu$.

Las correspondencias entre 3 líneas $\ell_i \mapsto \ell_i'$ de cada haz epipolar $(\mathbb{P}^1)^\nu$ imponen $5 = 2+2+1$ condiciones que se expresan mediante una restricción sobre $\mathbf{A}$ (nótese que ℓ_3 sólo impone una restricción que corresponde a la condición de pasar por el punto de intersección $\ell_1 \cap \ell_2$).

Ejercicio avanzado (Rectas epipolares para tres vistas) Extended el argumento presentado en el apartado anterior para construir las rectas epipolares correspondientes al caso de tres vistas. Para ello empezamos adaptando la notación de una forma muy esquemática

1. Sea $\mathbf{p}$ un punto en la primera vista con coordenadas $\mathbf{x}$, consideramos ℓ' la rectas epipolares en la segunda y tercera vista respectivamente, entonces

$$\ell'^T(\sum_i \mathbf{x}^i T_i) = 0^\top \;\; y \;\; (\sum_i \mathbf{x}^i T_i)\ell'' = 0$$

   $\Rightarrow$ Las rectas epipolares ℓ' y ℓ'' respecto de $\mathbf{x}$ se pueden calcular como el núcleo a izquierda y derecha de la matriz $\sum_i \mathbf{x}^i T_i$.

2. *Cálculo de epipolos* El epipolo $\mathbf{e}'$ en la segunda imagen es la intersección de todos los vectores del núcleo a la izquierda de las matrices $T_i, i = 1,2,3$. De manera análoga se hace para calcular $\mathbf{e}''$ con los vectores del núcleo de los T_i a la derecha.

Para extender al caso de tres vistas los argumentos presentados más arriba, se sugiere desarrollar el *esquema siguiente:*

1. Sea $\mathbf{p}$ un punto en la primera vista con coordenadas $\mathbf{x}$, consideramos ℓ' la rectas epipolares en la segunda y tercera vista respectivamente, entonces

$$\ell'^T(\sum_i \mathbf{x}^i T_i) = 0^\top \;\; y \;\; (\sum_i \mathbf{x}^i T_i)\ell'' = 0$$

   $\Rightarrow$ Las rectas epipolares ℓ' y ℓ'' respecto de $\mathbf{x}$ se pueden calcular como el núcleo a izquierda y derecha de la matriz $\sum_i \mathbf{x}^i T_i$.

2. *Cálculo de epipolos* El epipolo $\mathbf{e}'$ en la segunda imagen es la intersección de todos los vectores del núcleo a la izquierda de las matrices $T_i, i = 1,2,3$. De manera análoga se hace para calcular $\mathbf{e}''$ con los vectores del núcleo de los T_i a la derecha.

1.2.2. Cálculo de las relaciones estructurales

La Geometría Epipolar asociada a un par de cámaras es una reformulación de condiciones de incidencia para la intersección de haces de planos que pasan por la línea base con los planos de imagen correspondientes a cada una de las cámaras. En esta subsección se describen los modelos significativos para la estimación, remitiendo a la sección siguiente para una discusión más detallada de las estrategias generales (DLT), el análisis de su invariancia ("normalización" de datos) y del carácter robusto (RanSaC) de los estimadores estadísticos utilizados. Se sigue una estrategia de complejidad creciente que comienza con una descripción de propiedades relevantes de la matriz fundamental, se ilustra el efecto de movimientos rígidos sobre esta matriz y, por último, se aborda el caso general.

Explotando la restricción de coplanariedad

Denotemos mediante $\mathbf{p}$ y $\mathbf{p}'$ puntos homólogos con coordenadas $\mathbf{x}$ y $\mathbf{x}'$ en los planos correspondientes a las imágenes izquierda y derecha, respectivamente, que escribimos con coordenadas homogéneas normalizadas en el espacio ambiente (en la práctica la última coordenada no debería ser 1, sino la longitud focal para cada cámara que suponemos igual). En la representación 3D de la escena, los vectores $\mathbf{Cp}$, $\mathbf{Cp}'$ y $\mathbf{t}$ (correspondiente a la traslación entre los centros ópticos $\mathbf{C}$ y $\mathbf{C}'$) son coplanarios. El cambio de orientación de las cámaras $\mathbf{C}$ y $\mathbf{C}'$ se representa mediante una rotación $\mathbf{R}$ en el espacio cartesiano. Por ello, podemos reescribir la *restricción de coplanariedad* en forma vectorial como sigue:

$$\mathbf{x}_{\mathbf{C}}^{\top\prime}(\mathbf{t}\times\mathbf{R}\mathbf{x}_{\mathbf{C}}) \;=\; 0$$

donde el subíndice de las coordenadas se refiere al caso ideal en el que la calibración interna está dada por la matriz identidad. Reescribimos ahora el producto vectorial $\times$ en términos del producto por la matriz antisimétrica que corresponde al vector de traslación $\mathbf{t} = (t_x, t_y, t_z)^\top$, obteniendo

$$\mathbf{x}_{\mathbf{C}}^{\top\prime}[\mathbf{t}]_\times\mathbf{R}\mathbf{x}_{\mathbf{C}} \;=\; 0$$

donde

$$\mathbf{t}]_\times \;:=\; \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix}$$

La *matriz esencial* es por definición $\mathbf{E} := \mathbf{R}\mathbf{x}_{\mathbf{C}}$ por lo que la *restricción epipolar* para el caso euclídeo se escribe como una relación bilineal entre puntos homólogos:

$$\mathbf{x}_{\mathbf{C}}^{\prime\top}\mathbf{E}\mathbf{x}_{\mathbf{C}} \;=\; 0$$

Esta formulación no tiene en cuenta las distorsiones procedentes de la cámara, es decir, sólo tiene en cuenta la información asociada a los parámetros extrínsecos que corresponden al marco euclídeo ideal. Para recuperar una versión más realista que incluya la posible distorsión asociada a parámetros intrínsecos de cada cámara, es necesario introducir las 3×3-matrices de calibración y la relación entre las coordenadas

$$\mathbf{x} = \mathbf{K}\mathbf{x}_{\mathbf{C}} \quad \text{y} \quad \mathbf{x}' = \mathbf{K}'\mathbf{x}'_{\mathbf{C}'}$$

Sustituyendo en la relación bilineal obtenida más arriba se tiene otra relación bilineal

$$\mathbf{x}'^{\top}\mathbf{K}'^{-\top}\mathbf{E}\mathbf{K}^{-1}\mathbf{x} \;=:\; \mathbf{x}'^{\top}\mathbf{F}\mathbf{x} \;=\; 0$$

donde

$$\mathbf{F} \;:=\; \mathbf{K}'^{-\top}\mathbf{E}\mathbf{K}^{-1} \;=\; \mathbf{K}'^{-\top}[\mathbf{t}]_\times\mathbf{R}\mathbf{K}^{-1}$$

es la *matriz fundamental* que proporciona la relación estructural más importante entre puntos homólogos de pares de vistas.

Tal y como se ha comentado más arriba, la matriz fundamental lleva un punto $\mathbf{p}$ a una línea $\ell := \mathbf{F}\mathbf{x}$, es decir, existe una ambigüedad en la correspondencia epipolar definida por la matriz fundamental. En otras palabras cualquier punto $\mathbf{p} \in \ell$ se aplica sobre la misma línea ℓ' contenida en la vista derecha. Esta ambigüedad se traduce en que $det(\mathbf{F}) = 0$ que es una hipersuperficie cúbica en

el espacio proyectivo 8-dimensional de los coeficientes; en particular, la inversa de la matriz no está bien definida.

Ejercicio.- Describir la relación estructural asociada a la matriz fundamental para rectas homólogas contenidas en pares de vistas tomadas desde localizaciones próximas de cámaras (*Indicación:* Utilizad la dualidad inducida por una cónica no-degenerada en el plano. Nótese que la correspondencia entre líneas sigue siendo ambigua como aplicación entre planos proyectivos duales. ¿Qué puedes decir si la cónica considerada para la dualidad es la cónica absoluta?)

Propiedades de la matriz fundamental

A efectos de utilización posterior, recordamos aquí algunas propiedades que han sido mencionadas más arriba:

- Si $\mathbf{F}$ es la matriz fundamental entre la cámara $\mathbf{C}$ y la cámara $\mathbf{C}'$, entonces $\mathbf{F}^\top$ es la matriz fundamental entre la cámara $\mathbf{C}'$ y la cámara $\mathbf{C}$.
- La línea epipolar de $\mathbf{x}$ es $\ell' := \mathbf{F}\mathbf{x}$. Análogamente, la línea epipolar de $\mathbf{x}'$ es $\ell := \mathbf{F}^\top\mathbf{x}'$
- Los epipolos se calculan como el núcleo a la derecha o a la izquierda de la matriz fundamental. Así, p.e.
$$\ell^\top \mathbf{e} \;=\; \mathbf{x}'^T\mathbf{F}\mathbf{e} = 0 \quad \forall \mathbf{x}' \quad \Rightarrow \mathbf{F}\mathbf{e} = 0$$
y análogamente para $\mathbf{e}'$

Para estimar la matriz fundamental se utiliza el método SVD (Descomposición en Valores Singulares):

$$\mathbf{F} \;=\; \mathbf{U}\,\mathrm{diag}(\sigma_1, \sigma_2, 0)\mathbf{V}^\top$$

donde

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{e}'] \quad \text{y} \quad \mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{e}]$$

La identificación de la columna en $\mathbf{V}$ que corresponde al valor propio 0 proporciona un método de cálculo sencillo de los epipolos de la matriz fundamental [23].

Ejercicio (avanzado).- Supongamos que los epipolos no están a distancia infinita (cámaras con "vergencia")como ocurre en los casos de perspectiva angular y oblicua, 'p.e.. Denotemos mediante $\mathbf{M}_\pi = \mathbf{K}[\mathbf{I} \mid \mathbf{O}]$ y $\mathbf{M}_{\pi'} = \mathbf{K}'[\mathbf{R} \mid \mathbf{t}]$. Verificad [Har03] que

$$\mathbf{F} \;=\; \mathbf{K}'^{-\top}[\mathbf{t}]_\times\mathbf{R}\mathbf{K}^{-1} \;=\; [\mathbf{K}'\mathbf{t}]_\times\mathbf{K}'\mathbf{R}\mathbf{K}^{-1} \;=\; \mathbf{K}'^{-\top}\mathbf{R}\mathbf{K}^\top[\mathbf{K}\mathbf{R}^\top\mathbf{t}]_\times$$

El caso de una traslación pura

Supongamos que los dos planos de las cámaras son paralelos. En este caso, la línea base es paralela a a los dos planos. Por tanto, el haz de planos epipolares (que pasan por la línea base) corta a cada uno de los planos paralelos a lo largo de un haz de líneas epipolares que, en este caso, son paralelas a la línea base. La búsqueda de elementos homólogos sobre líneas epipolares se lleva a sobre haces

[23] En la última sección de este capítulo se presentan métodos de estimación más robustos

de líneas paralelas, Si la discretización del haz de planos epipolares está regularmente espaciada, la distancia entre las líneas epipolares es mayor en los extremos de la imagen que en la parte central. Esta representación es similar a la de la visión humana.

Si las dos cámaras son idénticas $\mathbf{K} = \mathbf{K}'$, la matriz fundamental se escribe de forma muy simple como

$$\mathbf{F} = [\mathbf{Kt}]_\times = [\mathbf{e}']_\times = \begin{pmatrix} 0 & -e'_z & e'_y \\ e'_z & 0 & -e'_x \\ -e'_y & e'_x & 0 \end{pmatrix}$$

y los epipolos ocupan la misma posición en ambas imágenes. Si la traslación es paralela al plano de imagen, entonces los epipolos se obtienen como puntos de corte de líneas epipolares horizontales paralelas; analíticamente: $e_z = e'_z = 0$. En este caso, $\mathbf{e}' = [1, 0, 0]'$. Por consiguiente, la matriz fundamental es de la forma

$$\mathbf{F} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

que es precisamente la matriz de la rotación infinitesimal con respecto al eje Ox (uno de los tres generadores del álgebra de Lie del grupo de rotaciones), lo cual facilita una reinterpretación en términos dinámicos de utilidad para cámaras móviles.

Ejercicio.- Bajo las condiciones de este apartado, verificad que la restricción epipolar general $\mathbf{x}'^\top \mathbf{F}\mathbf{x} = 0$ se re-escribe $y = y'$

Nota.- El ejercicio precedente permite reducir el rango de búsqueda sobre líneas paralelas al eje Ox tal y como se realiza en la percepción humana. Por consiguiente, la puesta en correspondencia de elementos homólogos correspondientes a una traslación perpendicular al eje visual muestra el principio que se aplica para la rectificación simultánea de pares de vistas estéreo en Fotogrametría; este caso se presenta con más detalle en la subsección 3.

Vistas separadas por una rotación pura

Sólo se facilita un esquema general dejando los detalles al lector como ejercicio. Si las dos cámaras son idénticas, $\mathbf{K} = \mathbf{K}'$. Por otro lado, como el vector de traslación $\mathbf{t}$ entre los centros $\mathbf{C}$ y $\mathbf{C}'$ de las cámaras es nulo, se tiene que

$$\mathbf{x} = \mathbf{K}[\mathbf{I} \mid \mathbf{O}]\mathbf{X} \quad , \quad \mathbf{x}' = \mathbf{K}[\mathbf{R} \mid \mathbf{O}]\mathbf{X}$$

por lo que

$$\mathbf{x}' = \mathbf{KRK}^{-1}\mathbf{x} = \mathbf{H}\mathbf{x}$$

donde $\mathbf{H}$ es una homografía que aplica las coordenadas de puntos en imagen en coordenadas "normalizadas" (centradas en el punto principal de la cámara a una distancia unidad desde la cámara). El efecto que tiene la rotación $\mathbf{R}$ y la multiplicación por $\mathbf{K}$ sobre los puntos de la imagen inicial es la generación de las coordenadas en el plano focal de la segunda cámara una vez realizada la rotación.

El caso general

El modelo teórico para la estimación de la matriz fundamental utiliza inicialmente el método de coeficientes indeterminados. En otras palabras, suponiendo que conocemos una "cantidad suficiente" de pares de elementos homólogos $(\mathbf{p}_i, \mathbf{p}'_i)$ pertenecientes a las vistas izquierda y derecha, se impone la condición $\mathbf{x}'^{\top}\mathbf{F}\mathbf{x} = 0$ y se resuelve en las entradas (f_{ij}) de $\mathbf{F}$. Para facilitar la resolución se adopta una notación afín para representar las coordenadas $[x_i, y_i, 1]^{\top}$ para ($\mathbf{p}_i$ y $[x'_i, y'_i, 1]^{\top}$ para ($\mathbf{p}'_i$ donde $i = 1\ldots, n$ y se reescribe la restricción epipolar como

$$x'xf_{11} + x'yf_{12} + x'f_{13} + yx'f_{21} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33} = 0$$

(omitiendo subíndices para aligerar la notación) [24] Como cada par de puntos homólogos da lugar a una ecuación, para una colección de (al menos) 8 pares de puntos homólogos se obtiene un sistema que escribimos en forma matricial como

$$\begin{pmatrix} x'_1x_1 & x'_1y_1 & x'_1 & y_1x'_1 & y'_1y_1 & y'_1 & x_1 & y_1 & 1 \\ x'_2x_2 & x'_2y_2 & x'_2 & y_2x'_2 & y'_2y_2 & y'_2 & x_2 & y_2 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 1 \\ x'_nx_n & x'_ny_n & x'_n & y_nx'_n & y'_ny_n & y'_n & x_n & y_n & 1 \end{pmatrix} \begin{pmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{pmatrix} = \mathbf{O}$$

o en forma vectorial más abreviada como

$$\mathbf{A}\mathbf{f} = \mathbf{O}$$

donde $\mathbf{A}$ representa la matriz de los coeficientes (un punto de $Im(s_{2,2})$), y $\mathbf{f}$ el vector asociado a las componentes de la matriz fundamental. Nuevamente, este sistema se resuelve por el método SVD. Esta presentación debe utilizar al menos $n = 8$ pares de puntos homólogos en "posición general". La solución obtenida debe verificar la condición $det(\mathbf{F}) = 0$, lo cual pone de manifiesto que en realidad bastan 7 puntos para determinar $\mathbf{F}$. Sin embargo, esta descripción tiene serios problemas para garantizar

- La elección apropiada de pares de puntos homólogos en "posición general": Si los pares de puntos homólogos no imponen condiciones independientes, el sistema está indeterminado.
- La estabilidad de las soluciones: pequeñas modificaciones en las entradas pueden alterar el comportamiento de las soluciones debido, entre otras cosas, al carácter no-lineal de la restricción $det(\mathbf{F}) = 0$

A la vista de estas dificultades es necesario utilizar una "cantidad redundante" de $n > 8$ puntos como candidatos a homólogos. La estrategia de resolución en el caso lineal (algoritmo de 8 puntos) consiste en los pasos siguientes:

1. Seleccionar una 8-upla de pares de puntos homólogos (vértices o máximos de intensidad, p.e.)

[24] Como es habitual, la versión homogénea utiliza el embebimiento de Segre $s_{2,2} : \mathbb{P}^2 \times \mathbb{P}^2 \hookrightarrow \mathbb{P}^8$ que en este caso no consideramos pues este enfoque sólo es válido cuando las localizaciones de las cámaras están suficientemente próximas. Por ello, podemos restringirnos al caso de una copia afín.

2. Calcular la matriz fundamental **F** correspondiente a dicha elección
3. Propagar a otras 8 uplas de pares de puntos homólogos y verificar si el "pegado" es apropiado (el error está por debajo de un umbral)
4. Si el pegado es correcto para al menos un 60% de pares de homólogos candidatos, aceptar la matriz **F** como válida
5. Si el pegado es incorrecto, elegir otra 8.upla y repetir el proceso.

Esta descripción es un caso particular de las estrategias RanSaC (Random Sample Consensus) que se exponen en la última sección de este capítulo con más detalle.

1.2.3. Rectificación de un par de vistas

La *rectificación simultánea de un par de vistas* $V_{r0}, V_{\ell 0}$ es la transformación que las lleva en un par de vistas $V_{r1}, V_{\ell 1}$ de modo que las líneas epipolares son colineales y paralelas a uno de los ejes de la imagen (habitualmente el eje Ox). En el espacio euclídeo se interpreta como una rotación de las cámaras originales en torno al centro óptico de la cámara. En el espacio afín se puede interpretar como un "abatimiento" de cada plano de imagen sobre un plano paralelo a la línea base b de modo que las líneas epipolares sean líneas paralelas a dicha línea base. Esta transformación modifica las (matrices de las) proyecciones de perspectiva que es necesario recalcular de acuerdo con la conservación del centro óptico y la restricción de paralelismo (nuevos planos de imagen paralelos a b).

Una vez realizada la rectificación el problema de búsqueda $2D$ de puntos homólogos (puesta en correspondencia) se reduce a un problema $1D$ sobre líneas epipolares paralelas en las imágenes rectificadas. Este procedimiento se llevaba a cabo tradicionalmente en la Fotogrametría Terrestre de forma manual mediante un restituidor; el procedimiento manual requería una elevada experiencia y generaba una fatiga ocular considerable. En primer lugar a se aborda el caso calibrado en el que se suponen conocidas previamente las matrices de proyección proyectiva (PPM) correspondientes a cada cámara. A continuación se comenta el caso no-calibrado.

El caso calibrado

En este apartado se siguen las notas de Fusiello con algunos cambios de notación. Denotemos mediante M_{01} y M_{02} a las 3×4-matrices iniciales que representan las proyecciones sobre ls planos de las cámaras y mediante

$$M_{k1} = \begin{pmatrix} \mathbf{a}_1^\top & a_{14} \\ \mathbf{a}_2^\top & a_{24} \\ \mathbf{a}_3^\top & a_{34} \end{pmatrix} \quad \text{y} \quad M_{k2} = \begin{pmatrix} \mathbf{b}_1^\top & b_{14} \\ \mathbf{b}_2^\top & b_{24} \\ \mathbf{b}_3^\top & b_{34} \end{pmatrix}$$

a las matrices de proyección obtenidas tras k iteraciones que corresponden a las condiciones requeridas (conservación del centro óptico y líneas epipolares horizontales sobre planos paralelos a la línea base b. Bajo estas condiciones, las matrices de proyección perspectiva (PPM) tienen el mismo plano focal si y sólo si

$$\mathbf{a}_3 = \mathbf{b}_3 \quad \text{y} \quad a_{34} = b_{34}$$

Los centros de proyección $\mathbf{C}_{k1}$ y $\mathbf{C}_{k2}$ para las proyecciones con matrices M_{k1} y M_{k2} están dadas por el núcleo de cada aplicación, es decir,

Figura 1.12: Rectificación de imágenes con haces de líneas epipolares paralelas al eje Ox

$$M_{k1}\begin{pmatrix} \tilde{\mathbf{c}}_{k1} \\ 1 \end{pmatrix} = \mathbf{O} \quad \text{y} \quad M_{k2}\begin{pmatrix} \tilde{\mathbf{c}}_{k2} \\ 1 \end{pmatrix} = \mathbf{O}$$

donde los centros ópticos iniciales se calculan vectorialmente mediante

$$\tilde{\mathbf{c}}_{01} = -M_{01}^{-1}\tilde{\mathbf{p}}_{01} \quad \text{y} \quad \tilde{\mathbf{c}}_{02} = -M_{02}^{-1}\tilde{\mathbf{p}}_{02}$$

En términos analíticos hay que resolver el sistema correspondiente a 6 ecuaciones lineales

$$\mathbf{a}_i^\top \mathbf{c}_{01} + a_{i4} = 0\ ,\ \mathbf{b}_i^\top \mathbf{c}_{01} + b_{i4} = 0 \quad i = 1,2,3$$

que corresponden a la expresión vectorial en términos de componentes.

Para que las líneas epipolares en los dos planos paralelos (correspondientes a la rectificación es necesario que puedan alinearse de forma automática; es decir, es necesario no sólo que las líneas epipolares sean paralelas, sino que la componente vertical sea la misma en ambos planos, es decir,

$$\frac{\mathbf{a}_2\tilde{\mathbf{x}} + a_{24}}{\mathbf{a}_3\tilde{\mathbf{x}} + a_{34}} = \frac{\mathbf{b}_2\tilde{\mathbf{x}} + b_{24}}{\mathbf{b}_3\tilde{\mathbf{x}} + b_{34}}$$

Utilizando ahora las relaciones $\mathbf{a}_3 = \mathbf{b}_3$ y $a_{34} = b_{34}$ se obtienen las relaciones adicionales

$$\mathbf{a}_2 = \mathbf{b}_2 \quad \text{y} \quad a_{24} = b_{24}$$

La *orientación del plano retinal rectificante* se elige de modo que los planos focales rectificantes sean paralelos a la intersección de los dos planos focales originales, es decir

$$\mathbf{a}_3^\top(\mathbf{f}_{13} \wedge \mathbf{f}_{23})$$

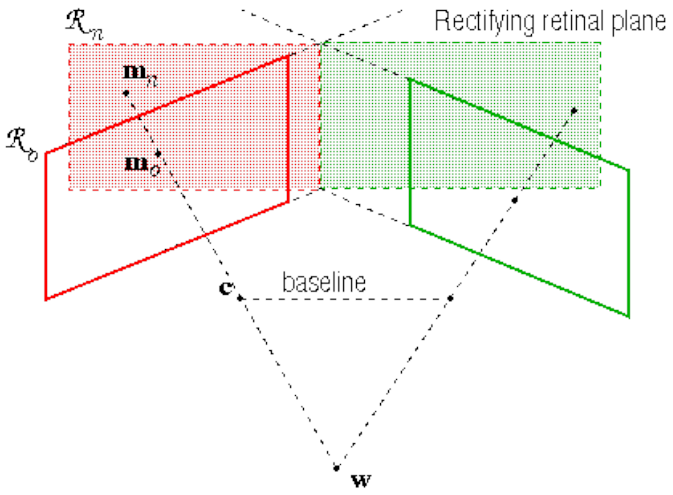

Figura 1.13: Esquema que ilustra los elementos que intervienen en la rectificación de imágenes en estéreo sobre el plano de la retina

donde $\mathbf{f}_{13}$ y $\mathbf{f}_{23}$ son las terceras filas de las matrices iniciales de proyección $\mathbf{M}_{01}$ y $\mathbf{M}_{02}$. En virtud de las relaciones anteriores la relación similar para $\mathbf{b}_3$ es redundante.

Las intersecciones del plano retinal con los planos $\mathbf{a}_1^\top \tilde{\mathbf{x}} + a_{14} = 0$ y $\mathbf{a}_2^\top \tilde{\mathbf{x}} + a_{24} = 0$ corresponde a los ejes u y v, respectivamente, de la referencia retinal. Para que esta *referencia* sea *ortogonal* los planos deben ser perpendiculares, es decir,

$$\mathbf{a}_1^\top \mathbf{a}_2 = 0 \quad \text{y} \quad \mathbf{b}_1^\top \mathbf{a}_2 = 0$$

donde se ha utilizado la restricción $\mathbf{a}_2 = \mathbf{b}_2$ mostrada más arriba. El *punto principal* (u_0, v_0) se obtiene como

$$u_0 = \mathbf{a}_1^\top \mathbf{a}_3 \quad , \quad v_0 = \mathbf{a}_2^\top \mathbf{a}_3$$

Para una cámara calibrada se tiene

$$\mathbf{a}_1^\top \mathbf{a}_3 = 0 \quad , \quad \mathbf{a}_2^\top \mathbf{a}_3 = 0 \quad , \quad \mathbf{b}_1^\top \mathbf{a}_3 = 0$$

Las *deformaciones oblicuas* en las longitudes focales a lo largo de las direcciones horizontal y vertical están dadas por

$$\alpha_u = \|\mathbf{a}_1 \wedge \mathbf{a}_3\| \quad , \quad \alpha_v = \|\mathbf{a}_2 \wedge \mathbf{a}_3\|$$

Como las matrices de proyección perspectiva (PPM) están definidas salvo factor de escala, una elección habitual es $\|\mathbf{a}_3\| = 1$ y $\|\mathbf{b}_3\| = 1$.

La resolución efectiva de las PPM utiliza una colección de 4 sistemas de ecuaciones más las restricciones lineales presentadas más arriba[25]

[25] http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/FUSIELLO/node16.html

Rectificación sin calibración previa

La estrategia que se sigue para el caso no-calibrado es similar a la presentada en el apartado anterior, pero con un uso más sistemático de la restricción epipolar presentada en el §2,2. Por ello, sólo se muestra una descripción cualitativa, dejando los detalles como ejercicio.

1.3. Reconstrucción 3D

El *objetivo* de la reconstrucción 3D basada en dos o más vistas es la generación automática de un modelo tridimensional de escenas o de objetos volumétricos que pueda ser navegado de forma interactiva por parte del usuario [26]. Para ello, es necesario incorporar diferentes tipos de restricciones de tipo topológico (ordenación, p.e.), algebraico (unicidad, p.e.) y geométrico (epipolar, p.e.).

Para simplificar, en esta sección nos restringimos al caso de vistas próximas tomadas desde diferentes posiciones por cámaras estáticas [27]. La Reconstrucción 3D depende del marco geométrico elegido (proyectivo, afín, euclídeo), según un orden de complejidad creciente. Esta distinción básica da lugar a la subdivisión de la sección en subsecciones donde se revisan los modelos y resultados más significativos con objeto de facilitar una solución más efectiva para la reconstrucción 3D.

Una diferencia importante con respecto a secciones anteriores consiste en que a lo largo de toda esta sección se aborda la reconstrucción densa, una vez resuelto el "modelo estructural" asociado a información dispersa. El carácter denso de la reconstrucción implica que es necesario "rellenar" los datos geométricos pegados mediante técnicas de "pegado" que afectan a "parches" o trozos de superficies que se "pegan" sobre los elementos estructurales [28].

El "pegado denso" proporciona un modelo que sigue siendo discreto y que, por consiguiente, es necesario completar "rellenando" el espacio entre los puntos significativos pegados. La interpolación suave tiene un coste computacional demasiado elevado; por ello se recurre a modelos lineales a trozos dados por mallas espaciales que sean compatibles con las restricciones mencionadas anteriormente. La triangulación de una imagen atendiendo a hechos 0-dimensionales es muy simple y se resuelve de forma óptima mediante triangulaciones de Delaunay.

Sin embargo, las triangulaciones asociadas a los mismos hechos $0D$ para diferentes vistas no es la misma, pues las orientaciones con respecto a la línea de visión de cada superficie es variable. Para resolver este problema es necesario introducir restricciones adicionales para puntos $3D$ obtenidos mediante reproyección y que afectan a la profundidad y a la orientación relativa de los triángulos asociados a las ternas de puntos significativos $3D$ más próximos obtenidos tras aplicar la restricción epipolar.

Un método para completar información correspondiente a agujeros espúreos consiste en detectar el borde de los objetos y propagar en espiral hacia el interior, utilizando la información correspondiente a las "células básicas" (triángulos ó cuadriláteros) de la malla asociada [29]

1.3.1. Reconstrucción proyectiva basada en dos imágenes

En esta subsección se describen elementos significativos para la Geometría Epipolar y se presentan las restricciones estructurales que deben verificar elementos homólogos en dos vistas. Está organizada en las siguientes apartados:

1. Principio básico ideal para la reconstrucción
2. Rectas epipolares. Cálculo algebraico
3. Elementos de cálculo vectorial para líneas epipolares

[26] La reconstrucción 3D a partir de una sola vista se desarrolla en el capítulo 4 de este módulo

[27] La incorporación de restricciones vinculadas al movimiento se aborda en el capítulo 5 de este módulo

[28] Es deseable que la Reconstrucción 3D de un objeto no tenga agujeros en el objeto reconstruido. Una resolución eficiente de esta cuestión es un problema difícil que se aborda en el módulo 5 utilizando modelos de propagación sobre trozos de superficie $S^\alpha = \partial B^\alpha$ que acotan los objetos volumétricos B^α

[29] Esta estrategia se detalla en el módulo sobre Visión Estéreo Dinámica

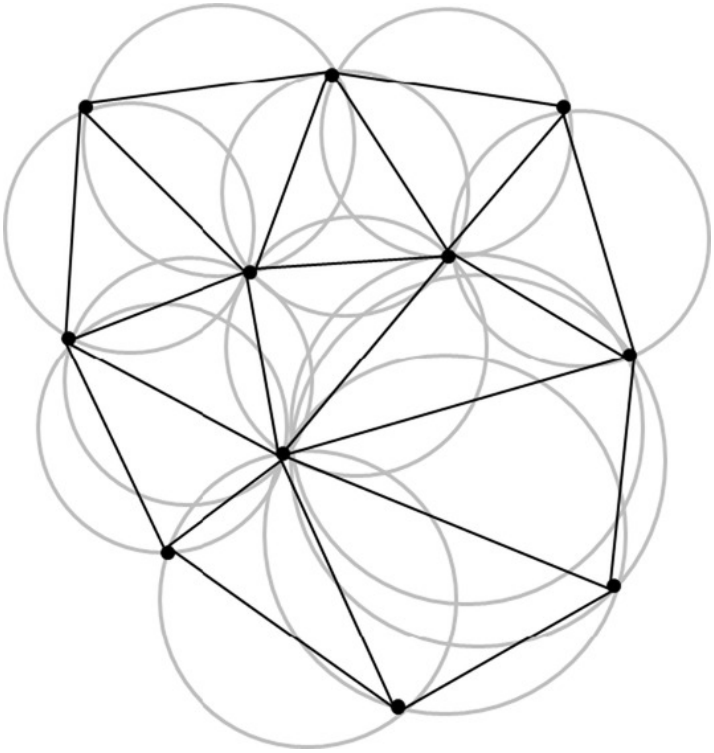

Figura 1.14: Ilustración de la triangulación de Delaunay sobre un conjunto de vértices

Principio básico ideal para la reconstrucción

Problema.- Se desea reconstruir un punto $\mathbf{P}_i$ en el espacio tridimensional a partir de las proyecciones $\mathbf{p}_{i\alpha}$ sobre cada plano imagen Π_α para $\alpha = 1, 2$

El *principio básico ideal* para la reconstrucción consiste en que la intersección $r_{i1} \cap r_{i2}$ de los rayos $r_{i1} := \overline{\mathbf{F}_1\mathbf{p}_{i1}}$, $r_{i2} := \overline{\mathbf{F}_2\mathbf{p}_{i2}}$ que pasan por puntos homólogos $\mathbf{p}_{i1}$, $\mathbf{p}_{i2}$ determina de una forma teórica el punto $\mathbf{P}_i$. En la práctica dichos rayos no se cortan, sino que se cruzan en el espacio. Para resolver este problema hay *diferentes aproximaciones al problema*:

- *Métrica:* Minimizar distancia entre rayos sobre una recta perpendicular común que denotamos L_i.
- *Probabilista:* Se supone que la distribución del ruido es Gaussiana y se toma el punto medio del segmento que determina la recta L_i al cortar a r_{i1} y r_{i2} y minimizar.
- *Mínimos cuadrados para puntos próximos:* Buscar un par alternativo de puntos homólogos $\hat{\mathbf{p}}_{i1}$, $\hat{\mathbf{p}}_{i2}$ en las imágenes tales que la suma de los cuadrados de las distancias a los puntos originales $\mathbf{p}_{i1}$, $\mathbf{p}_{i2}$ sea mínima y se satisfaga la restricción epipolar.
- *Geométrica:* Encontrar "líneas epipolares" ℓ_{i1}, ℓ_{i2} (homólogas) que minimicen $d(\ell_{i1}, \ell_{i2})$ y proporcionen el punto original de la correspondencia.
- *Parametrización y resolución algebraica:* Parametrizar el haz de líneas epipolares, llevar a cabo una reformulación global el problema y encontrar raíces reales del polinomio de grado 6 (Hartley & Sturm'97).

Figura 1.15: Búsqueda de puntos homólogos a lo largo de líneas epipolares

1.3.2. Reconstrucción afín

La Reconstrucción afín proporciona representaciones que son invariantes módulo transformaciones afines. Recordemos que $\mathbb{P}^n = \mathbb{A}^n \cup H_\infty$. Por ello, una transformación afín del espacio ordinario está dada por una *homografía del espacio proyectivo* representada por una 4×4-matriz regular (definida salvo factor de proporcionalidad) que deja invariante el plano del infinito H_∞; para una adecuada elección de coordenadas, podemos suponer que H_∞ está dada por $X_4 = 0$; bajo esta hipótesis la última fila de las transformaciones afines es (0001).

Para facilitar la visualización en el plano de imagen de la transformación proyectiva, se recurre a una representación en el plano de imagen sobre la que se proyecta la transformación realizada en el espacio. Para estimar dicha transformación se utiliza nuevamente la restricción epipolar que está representada en nuestro caso por la matriz fundamental **F** introducida por Longuet-Higgins (1980). A continuación se muestra la estrategia desarrollada por R.Hartley (algoritmo lineal de 8 puntos) para la *estimación de* **F**.

La elevada cantidad de candidatos y los problemas vinculados a una asignación incorrecta de elementos homólogos, sugiere validar la elección de 8 pares de candidatos a homólogos extendiendo la reconstrucción a oras 8-uplas de pares de puntos: si el porcentaje de outliers está por debajo de un umbral, se da por válida la reconstrucción; en caso contrario hay que seleccionar otra 8-upla y repetir el proceso. Esta estrategia es una variante del procedimiento RANSAC (Random Sampling Consensus) adaptada por P.Torr en su tesis (1996), realizada bajo la dirección de A.Zisserman. La subsección concluye presentando algunas consecuencias de la Reconstrucción afín que son de utilidad para la producción de contenidos digitales.

Descripción de la matriz fundamental

Una homografía o colineación arbitraria $\mathbf{H}$ de $\mathbb{P}^2$ requiere 8 parámetros. Sin embargo la *matriz fundamental* $\mathbf{F} := \mathbf{AH}$ queda unívocamente determinada por las restricciones siguientes:

1. $\mathbf{F}\mathbf{e}_{12} = 0$, es decir, $\mathbf{e}_{12} \in Ker(\mathbf{F})$. Por consiguiente $rango(\mathbf{F}) = 2$
2. $\mathbf{e}_{21}^\top \ell' = 0$: Todas las líneas ℓ' del pencil derecho pasan por $\mathbf{e}_{21}^\top$
3. $\mathbf{e}_{21}^\top \mathbf{F} = 0$ (recíproca de la primera condición)
4. $\mathbf{x}' \in \ell' \Rightarrow (\mathbf{x}')^\top . \ell' = 0$

Con esta notación, si $\mathbf{F}$ la *matriz fundamental* la *Restricción epipolar* se escribe

$$(\mathbf{x}')^\top \mathbf{F}\mathbf{x} = 0$$

Esta igualdad se puede interpretar de dos formas:

- Haciendo $\ell' := \mathbf{F}\mathbf{x}$ (recta epipolar) se tiene la igualdad obvia $(\mathbf{x}')^\top \ell' = 0$, es decir, $(\mathbf{x}')^\top \in \ell'$ (en otras palabras $\mathbf{x}'$ pertenece a su recta epipolar)
- Haciendo $\ell = (\mathbf{F}\mathbf{x}')^\top$ se tiene la igualdad obvia $\mathbf{x}\ell = 0$, es decir, $(\mathbf{x})^\top \in \ell$ (en otras palabras $\mathbf{x}$ pertenece a su recta epipolar)

Los elementos homólogos asociados a dos proyecciones deben ser equivalentes módulo una transformación proyectiva. Como cada matriz de proyección depende de 11 parámetros (12 salvo escala) y una transformación proyectiva depende de 15 parámetros (16 salvo escala), se tiene que $\mathbf{F}$ depende de 7 parámetros (salvo escala).

Alternativamente, $\mathbf{F}$ es una 3×3-matriz que depende de 8 parámetros (9 salvo escala); la condición $rang(\mathbf{F}) = 2$ se traduce en una relación cúbica $det(\mathbf{F}) = 0$ entre los coeficientes de $\mathbf{F}$, por lo que $\mathbf{F}$ depende de 7 parámetros. Esta presentación subraya el carácter no lineal de la estimación.

Métodos de optimización Local

El *objetivo* de este tipo de métodos es la búsqueda local e identificación de elementos comunes. La *idea básica* para los métodos locales consiste en restringir la "ventana de búsqueda". Para ello, es necesario desarrollar herramientas estadísticas que permitan comparar datos contenidos en ventanas de búsqueda. La *correlación* proporciona una metodología general para diferentes tipos de criterios que pueden corresponder a

- Algún tipo de *distancia:* La distancia L_2 es más simple, pero la L_1 es más robusta.
- *M-estimación* con un funcional cuadrático truncado (Black y Anandan, 1993) → correlación robusta (Rousseeuw & Leroy'87).
- *Mínimas medianas cuadradas* → correlación robusta no-lineal: Muy eficiente cuando se combina con color (A.Viloria y JF)

Ejercicio (avanzado).- Describid explícitamente y realizad una implementación computacional para cada uno de los tres métodos descritos.

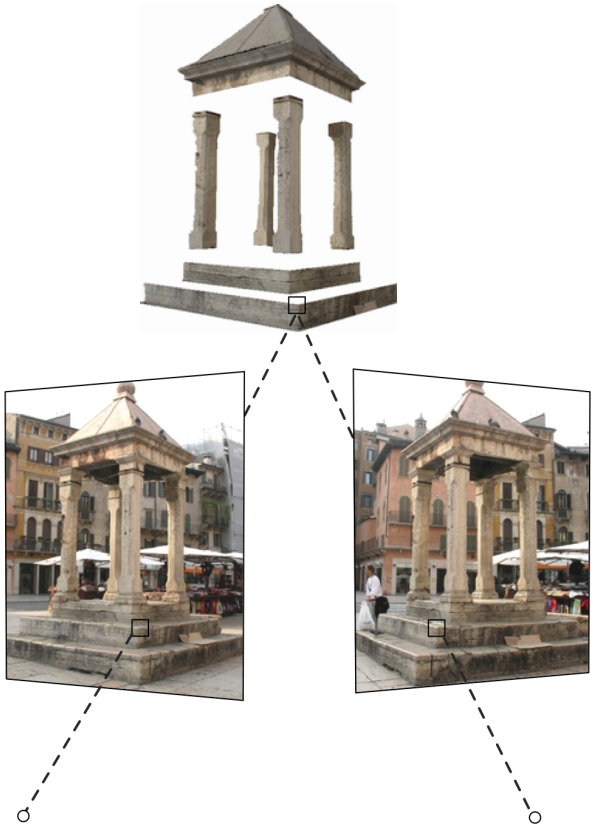

Figura 1.16: Dos vistas en perspectiva de la misma escena con los puntos homólogos resaltados

Estimando la matriz fundamental: Caso no calibrado

En el caso no-calibrado, la matriz fundamental es una 3×3 -matriz con determinante nulo. Dicha condición está representada por una hipersuperficie cúbica en el espacio de 9 parámetros f_{ij} de la matriz **F**. Por ello, bastan 7 parámetros (8 salvo factor de proporcionalidad) para identificar **F**. En consecuencia, el número mínimo de pares de puntos homólogos para alcanzar una solución única es 7 (cada par de puntos homólogos "genéricos" impone una condición l.i.). La optimización correspondiente es no-lineal, lo cual da lugar a soluciones inestables y con una convergencia lenta. Longet-Higgins (1981) propuso utilizar un algoritmo basado en 8 puntos que es lineal, aunque no sea minimal. Este algoritmo consiste en resolver el sistema lineal

$$\mathbf{Af} = \mathbf{0}$$

correspondiente a "imponer las condiciones" de ser homólogos para cada uno de los 8 pares de puntos homólogos, siendo $\mathbf{f} = (f_{11}, \ldots, f_{33})^\top$ el array que describe las entradas de **F** como vector (incógnitas a determinar) y **A** un 8×9-matriz cuyas filas representan la condición para que $\mathbf{p}_i$ y $\mathbf{p}'_i$ sean (candidatos a) homólogos con $1 \leq i \leq 8$.

Para obtener una matriz **A** bien condicionada, Hartley (1997) propone tomar coordenadas normalizadas para los pares candidatos a homólogos, De este modo, se obtiene un vector **f** de norma unidad en $Ker(A)$. Sin embargo, esta elección puede generar matrices **A** que no tienen rango 2, dando lugar a interpretaciones inconsistentes de **F** (las líneas epipolares de cada plano no pasan necesariamente

por el epipolo). La solución clásica [Bou95] para este problema consiste en realizar una SVD para **F** forzando que uno de los valores propios sea nulo (otro problema de optimización, nuevamente).

El *algoritmo de 7 puntos* consiste en resolver el sistema formado por 7 pares de puntos homólogos. El núcleo $Ker(\mathbf{A})$ del sistema $\mathbf{Af} = 0$ tiene dimensión 2. Para evitar el problema mencionado más arriba, se eligen dos soluciones particulares $\mathbf{F}_1$ y $\mathbf{F}_2$ del sistema y se impone la condición de tener rango 2 a la recta $(1-\lambda)\mathbf{F}_1 + \lambda\mathbf{F}_2$; la condición $det[(1-\lambda)\mathbf{F}_1 + \lambda\mathbf{F}_2] = 0$ es una cúbica que tiene 1 o 3 soluciones reales.

La aproximación al problema correspondiente al algoritmo de 7 (pares de) puntos tiene la ventaja de que las soluciones obtenidas corresponden a matrices fundamentales que verifican la restricción epipolar. Sin embargo, tiene el inconveniente del carácter no-lineal del problema de optimización asociado (para una discusión más detallada ver la defensa del algoritmo de 8 puntos por R.Hartley). No obstante, la aproximación basada en 7 puntos se utiliza en relación con los algoritmos robustos basados en variantes de Ransac (ver más abajo).

Ambigüedad en la reconstrucción

En la introducción a la sección se ha comentado la unicidad como una restricción clave para generar un modelo 3D de forma automática. Lamentablemente, existe una ambigüedad en la posición del plano de la cámara con respecto al centro óptico, lo cual da lugar a dos reconstrucciones equivalentes que son compatibles con los datos de partida. Ambas reconstrucciones están relacionadas mediante una simetría con respecto al centro óptico (modelo del plano proyectivo basado en identificar puntos antipodales de una esfera). Esta ambigüedad se evito inicialmente introduciendo la noción de orientación sobre el modelo proyectivo de la cámara (según [LF96]) ó, casi simultáneamente, mediante la introducción de una tercera vista (enfoque que se desarrolla en el apéndice a este capítulo). La introducción de una orientación sobre el modelo proyectivo se aplica a cuestiones relacionadas con diferentes aspectos cuyo desarrollo se deja como ejercicio.

- *Reconstrucciones posibles e imposibles en estéreo* basada en la compartimentación del plano en las 4 regiones que son la traza de las zonas en las que los 2 planos focales dividen al espacio ambiente. La reconstrucción posible corresponde a que los puntos estén situados todos por delante de ambos planos, por lo que debemos eliminar los falsos pegados correspondientes a las regiones restantes. El criterio de orientación positiva elegido corresponde a fijar como sentido positivo sobre la línea epipolar el que va desde el epipolo al punto situado enfrente.
- *Supresión de superficies ocultas*.- Si dos puntos diferentes de la escena se proyectan sobre el mismo punto en una imagen, no es posible eliminar uno de ellos sin recurrir a visión estéreo, pues ambos pertenecen a la misma recta epipolar.

1.3.3. Reconstrucción Euclídea. Matriz Esencial

Esta sección es complementaria a la reconstrucción euclidea usando la calibración de la cámara descrita en el capitulo anterior de este mismo modulo, adoptando la misma metodología que en el caso no calibrado.

Estimando la matriz fundamental: Caso calibrado

Recordemos que la matriz fundamental **F** se descompone en producto de tres matrices:

$$\mathbf{F} = \mathbf{K}'\mathbf{E}\mathbf{K}^{-1}$$

donde $\mathbf{K}$ es la matriz de calibración de la primera cámara, $\mathbf{K}'$ es la matriz de calibración de la segunda cámara y

$$\mathbf{E} = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & t_x \\ -t_y & t_x & 0 \end{pmatrix} \mathbf{R}$$

es la matriz esencial (composición de una rotación y una traslación) que, en ocasiones se representa simbólicamente como

$$\mathbf{E} = \mathbf{t} \times \mathbf{R}$$

Supongamos que los parámetros intrínsecos de la cámara son conocidos. Si se toman coordenadas normalizadas (ver más arriba), es posible reducir en una unidad el número de parámetros. Habitualmente se normaliza el vector de traslación (algo que desde el punto de vista geométrico carece de justificación) y se supone que $\|\mathbf{t}\| = 1$, por lo que $\mathbf{E}$ sólo requiere 5 parámetros (2 para $\mathbf{t}$ normalizado y 3 para $\mathbf{R}$).

Cada par de puntos homólogos $(\mathbf{p}, \mathbf{p}')$ impone una restricción (verificadlo como ejercicio escribiendo las ecuaciones); en consecuencia, de forma teórica bastaría tomar 5 pares de puntos homólogos. Sin embargo, la resolución que se obtiene al tomar 5 pares es inestable y en general proporciona 10 soluciones que son compatibles con los datos. Por ello, hay que añadir más pares de puntos homólogos, lo cual da lugar a un sistema redundante que debe resolverse mediante técnicas de optimización. Estas cuestiones se comentan en varios lugares en relación con la Reconstrucción Euclídea (capítulo 2) y los procedimientos de optimización (última sección de este capítulo).

En el marco proyectivo de este capítulo, la reconstrucción euclídea es similar a las reconstrucciones proyectiva y afín descritas más arriba; en este último caso, los datos son invariantes por las transformaciones del grupo proyectivo $G(\mathbb{P}^n) = \mathbb{P}GL(n+1;\mathbb{R})$ o del grupo afín $G(\mathbb{A}^n)$ (transformaciones proyectivas que dejan invariante un hiperplano del infinito). En el caso euclídeo el grupo estructural está dado por el grupo euclídeo $G(\mathbb{E}^n)$ las transformaciones rígidas (rotaciones o traslaciones) salvo escala, es decir, por el grupo de semejanzas; este grupo. La reconstrucción euclídea requiere incorporar *restricciones métricas* sobre el modelo proyectivo. Por ello, la jerarquía natural entre los grupos $G(\mathbb{E}^n \hookrightarrow G(\mathbb{A}^n) \subset G(\mathbb{P}^n)$ induce una jerarquía natural entre las reconstrucciones euclídea, afín y proyectiva.

Para una cámara fija o dos cámaras idénticas con diferente localización, la *matriz esencial* $\mathbf{E}$ en el caso euclídeo corresponde al movimiento rígido en el espacio cartesiano ordinario (dotado de la métrica euclídea). Por ello, es composición de una rotación $\mathbf{R}$ y una traslación $\mathbf{t} = (t_x, t_y, t_z)^\top$ representada por el producto de matrices:

$$\mathbf{E} = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & t_x \\ -t_y & t_x & 0 \end{pmatrix} \mathbf{R}$$

que, en ocasiones se representa simbólicamente como

$$\mathbf{E} = \mathbf{t} \times \mathbf{R}$$

La matriz esencial juega el mismo papel que la matriz fundamental $\mathbf{F}$ para la puesta en correspondencia de datos homólogos. Por los requerimientos de precisión, es claro que el emparejamiento

de datos homólogos para el caso euclídeo debe incorporar necesariamente las matrices de calibración interna **K** y **K**′ de las cámaras. Las matrices fundamental y esencial están relacionadas por

$$\mathbf{F} = \mathbf{K}'\mathbf{E}\mathbf{K}$$

es decir, son las clases de conjugación por la acción izquierda-derecha de las matrices de calibración interna. La matriz esencial tiene un autovalor nulo (de la misma forma que la matriz fundamental), pero además los otros dos valores propios son iguales. El estudio de la geometría de la variedad esencial (matrices que son equivalentes a una matriz con autovalores $(o, \lambda, lambda)$) tiene interés para acelerar procedimientos de optimización.

En el capítulo 2 se ha mostrado cómo la información métrica relacionada con la matriz de calibración **K** está "codificada" en la cónica absoluta $\omega_\infty = \Omega_\infty \cap H_\infty$ que es intersección de la cuádrica absoluta Ω_∞ con el plano del infinito H_∞. Este resultado proporciona la clave para la integración en el marco proyectivo que vertebra todo el capítulo. Por ello, de una forma ideal, basta con realizar una estimación de dicha cónica en imagen y "corregir" le deformación generada por una homografía. Si se adopta el enfoque jerarquizado que parte de la reconstrucción proyectiva, es necesario introducir restricciones métricas adicionales sobre la restricción epipolar representada por la matriz fundamental. En el caso euclídeo, la *matriz esencial* desempeña el papel estructural para el caso euclídeo. Por ello, es necesario estudiar las propiedades de la matriz esencial, introducir estimadores robustos y diseñar / implementar algoritmos que permitan realizar dichas tareas de forma automática.

Para reforzar el emparejamiento proyectivo que utilizar restricciones bilineales, se introducen *restricciones euclídeas* relativas a la conservación de las longitudes de los segmentos a comparar modulo un cierto umbral de tolerancia que depende de la disparidad del sistema bicameral (en el caso estático) y de la estimación del movimiento (si las cámaras se han desplazado; ver módulo 5 para más detalles).

La longitud de segmentos homólogos puede variar entre dos vistas fuera del umbral de tolerancia permitido, debido a una oclusión parcial del segmento elegido para la comparación. Además, los dos procedimientos comentados más arriba (añadir segmentos putativos o bien puntos intermedios) fallan en presencia de oclusiones parciales. Por ello, esta situación debe ser identificada simultáneamente en términos de la coherencia global de la puesta en correspondencia y desechada para evitar errores en los procesos de realimentación del sistema.

Para minimizar la distancia entre la elevación de elementos (puntos o segmentos) candidatos homólogos interesa desarrollar procedimientos de optimización muy rápidos. Esta cuestión afecta a la función de coste propuesta y a la posibilidad de métodos estándar rápidos de descomposición o factorización de las matrices resultantes.

Norma euclídea y restricción epipolar

El *primer candidato* viene dado por la minimización de la *métrica euclídea* evaluada sobre las proyecciones de cada punto $\mathbf{p}_1, \ldots, \mathbf{p}_r$ en cada imagen

$$\mathcal{C} := (\sum_i d_i)^2) := (\sum_{i=1}^{r} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]$$

que relaciona los valores esperados en las coordenadas (x_i, y_i) de $\mathbf{p}_i$ con sus valores actuales $(\hat{x}_i, \hat{y}_i)$ en cada imagen. La minimización de esta función se hace sometida a la *restricción epipolar*

$$\hat{\mathbf{x}}_2^\top F_{12} \hat{\mathbf{x}}_1 = 0$$

donde F_{12} es la matriz fundamental que relaciona la primera y a segunda imagen. La función de coste asociada $\mathcal{C}\sum_i d_i)^2)$ presenta un mínimo para la cual se obtiene la relación buscada (matriz fundamental, homografía o transformación cuadrática) entre puntos homólogos.

Ajuste basado en perturbaciones de datos

En el caso más general correspondiente a múltiples vistas, debido a diferentes problemas en la captura, procesamiento y análisis, los puntos candidatos a homólogos están lejos de satisfacer el sistema de ecuaciones que acabamos de mostrar. Por ello, es preciso implementar métodos más flexibles (de tipo probabilístico) que en el límite proporcionen la comparación euclídea que acabamos de mostrar. Una adaptación del criterio que acabamos de mostrar en términos de máxima verosimilitud (MLE, según [TZ97]) conduce a *minimizar una función de densidad de datos perturbados* según una distribución normal asociada a la métrica euclídea

$$\mathcal{P} := \Pi_i \frac{1}{2\pi\sigma^2} exp\frac{-\sum_i d_i)^2}{2\sigma^2} = \Pi_i \frac{1}{2\pi\sigma^2} exp\frac{(-\sum_{i=1}^{r}[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2])}{2\sigma^2}$$

El marco general para la interpretación geométrica de este proceso de optimización (en términos de mínimos cuadrados o de su versión probabilista) esta dado por una copia afín de la imagen $Im(s_{2,2})$ producto de Segre de dos copias del plano proyectivo $\mathbf{P}^2$. Este producto es una variedad 4-dimensional en $\mathbf{P}^7$, por lo que la ecuación epipolar $\hat{\mathbf{x}}_2^\top F_{12}\hat{\mathbf{x}}_1 = 0$ representa una sección lineal de la variedad de Segre. Así vemos que las matrices fundamentales parametrizan una subvariedad 3-dimensional de la variedad de Segre, mientras que las proyectividades y las transformaciones cuadráticas corresponden a variedades 2-dimensionales (razonadlo como ejercicio). Nótese que como $s_{2,2}$ es un embebimiento, se puede identificar $\mathbf{P}^2\times\mathbf{P}^2$ con su imagen en $\mathbf{P}^7$, por lo que no se hace distinción alguna en la notación.

La condición para que un par de puntos $(\mathbf{p},\mathbf{p}') \in \mathbf{P}^2\times\mathbf{P}^2$ sean homólogos (verifiquen la restricción epipolar) corresponde a que la proyección sobre el punto más próximo sea ortogonal al plano tangente a la variedad de Segre en el punto representado por la 4-upla $(\hat{\mathbf{p}},\hat{\mathbf{p}}') = (\hat{x},\hat{y},\hat{x}',\hat{y}')$. La distancia ortogonal medida en la copia afín es equivalente al error de reproyección del punto proyectivo 3D proyectado, que es la distancia d_i minimizada con la función distancia euclídea usual.

Distancia ortogonal y matriz fundamental

Hartley y Sturm (1994) han demostrado que dada la matriz fundamental $\mathbf{F}$, es posible recuperar la distancia ortogonal, $\hat{\mathbf{x}}^{(1)}$ y $\hat{\mathbf{x}}^{(2)}$ como soluciones de un polinomio de grado 6 en una variable. Sampson y Taubin han desarrollado un método aproximado para calcular las soluciones de este polinomio (hay una variante desarrollada por Torr y Zisserman en 1997).

1.3.4. Matriz esencial para la reconstrucción euclídea

La Reconstrucción $3D$ utiliza una condición de coplanariedad entre las rectas que conectan cada punto $\mathbf{P}$ con sus proyecciones $\mathbf{p}$ y $\mathbf{p}'$, y con la recta que pasa por los centro de proyección $\mathbf{C}$ y $\mathbf{C}'$. En el caso euclídeo, la condición de coplanariedad se expresa como una anulación del producto mixto

$$\overline{\mathbf{Cp}}[\overline{\mathbf{CC}'}\times\mathbf{C}'\mathbf{p}'] = 0$$

En e caso calibrado se pueden elegir coordenadas para que las matrices de proyección están dadas por

$$M_\pi = (Id \mid \underline{0}) \quad \text{y} \quad M_{\pi'} = (\mathbf{R}^\top \mid --\mathbf{R}^\top \mathbf{t})$$

donde $\mathbf{t}$ representa la traslación que lleva $\mathbf{C}$ en $\mathbf{C}'$ y $\mathbf{R}$ la rotación tridimensional asociada al cambio en la orientación. Con esta notación, la condición de coplanariedad presentada más arriba se reescribe en el caso calibrado como

$$\mathbf{p}[\mathbf{t} \times \mathbf{R}\mathbf{p}'] = 0 \;\Rightarrow\; \mathbf{p}^\top \mathbf{E}\mathbf{p}' = 0\,,$$

donde $\mathbf{E} := \mathbf{t} \times \mathbf{R}$ es la *matriz esencial* (Longet-Higgins, 1981) y el producto $\mathbf{t} \times \mathbf{R}$ se entiende como el producto matricial de la matriz antisimétrica asociada al vector traslación $\mathbf{t}$ por la rotación $\mathbf{R}$.

El enfoque geométrico para la Reconstrucción 3D en el caso euclídeo basado en condiciones de coplanariedad y un conocimiento previo de calibración es ideal. En la práctica los datos están corrompidos por el ruido, la manipulación de imagen, los cálculos inexactos (es necesario truncar desarrollos en serie), las oclusiones parciales o bien las asignaciones erróneas entre elementos homólogos.

Una primera estrategia para el caso afín basado en condiciones de incidencia se esboza en el apartado siguiente. Grosso modo, consiste en añadir más información relativa a una PL-estructura generada a partir de puntos 3D, optimizar y suavizar. El mayor problema inicial concierne a expresar dicha condición de coplanariedad, sin utilizar información métrica asociada a cámaras calibradas; este problema reaparece en múltiples formas a lo largo del capítulo.

La identificación de la transformación más apropiada para visualizar una escena desde una localización inicial, requiere construir un "camino" sobre el conjunto de transformaciones geométricas asociadas al marco geométrico elegido. En el caso euclídeo, las transformaciones son rotaciones $\mathbf{R} \in SO(3;\mathbb{R})$ y traslaciones $\mathbf{t} \in \mathbb{R}^3$ en el espacio euclídeo ordinario $\mathbb{E}^3$ (espacio cartesiano dotado con la métrica euclídea).

Las transformaciones de semejanza se componen multiplicando las matrices que las representan; con esta operación el conjunto de las transformaciones euclídeas forman el grupo euclídeo al que denotamos mediante $E(3;\mathbb{R})$ [30]. El cociente del grupo euclídeo por el grupo de las homotecias $\mathbb{R}^* := \{\lambda I_3 \mid \lambda \in \mathbb{R}^*\}$ (constantes no-nulas) corresponde a las transformaciones euclídeas salvo factor de escala a las que se llama *transformaciones de semejanza* y denotaremos mediante $SE(3;\mathbb{R})$.

Superponiendo información geométrica

El modelado basado en varias vistas utiliza la información geométrica (puntos y líneas) contenida o generada a partir del análisis de las imágenes digitales (módulo 1); en el caso no-calibrado, esta utilización se realiza de forma independiente con respecto a los dispositivos utilizados para la captura de información o de información previa sobre la escena. En particular y para minimizar los problemas de pegado, ello implica que las imágenes han sido "corregidas" (rectificación y corrección de las distorsiones radial y tangencial); de este modo se obtienen inputs robustos para el pegado geométrico y cualquiera de las reconstrucciones que se consideran en este capítulo están bien definidas *salvo escala*. La corrección de distorsiones se realiza de forma independiente para cada imagen; la rectificación se realiza para facilitar el escaneo simultáneo para cada par de imágenes (pero no se puede hacer de forma simultánea para todas las vistas).

[30] Formalmente, el grupo euclídeo es el producto semidirecto del grupo especial ortogonal $SO(n;\mathbb{R}) := O(n;\mathbb{R}) \cap SL(3;\mathbb{R})$ y el grupo de traslaciones $\mathbb{R}^n$, donde $O(n;\mathbb{R}) := \{AinGL(n;\mathbb{R}) \mid {}^\top AA = I_n\}$ es el grupo ortogonal y $SL(n;\mathbb{R}) := \{AinGL(n;\mathbb{R}) \mid det(A) = 1\}$ es el grupo especial lineal

Los enfoques basados en PL-estructuras (PL: Piecewise Linear o lineales a trozos, como las mallas triangulares o cuadrangulares) o basadas en propiedades radiométricas (asociadas al color u otros funcionales asociados a la luz) se abordan en capítulos posteriores y son apropiadas para el modelado tridimensional, pero no para la reconstrucción 3D que es previa al modelado. En el capítulo siguiente se aborda el caso correspondiente a la visualización avanzada (incluyendo aspectos radiométricos) de objetos o de esencias; para ello, se utilizan diferentes tipos de proyección, incluyendo proyecciones esféricas o cilíndricas con algunas aplicaciones básicas.

Hacia un modelo más completo

El *segundo objetivo* de la Reconstrucción 3*D* es la generación automática de un modelo 3*D* tan completo como sea posible a partir de las imágenes capturadas de forma secuencial o simultánea. Este objetivo extiende el primer objetivo mencionado más arriba; en efecto, la introducción de una cámara virtual en la localización (posición y orientación) deseada permite obtener una imagen como proyección asociada a dicha localización.

En este capítulo sólo se considera la recuperación de la información geométrica planar (relativa a la imagen) o volumétrica (relativa a la escena) para "la estructura lineal" que concierne a la parte visible de objetos B_α o de la escena tridimensional $\mathcal{E}$. La "estructura lineal" se entiende en un sentido inicialmente restringido a una colección de puntos, líneas y planos "dispersos" que ayudan a organizar la información de forma invariante con respecto a la localización del observador; esta estructura se "densifica" dependiendo de los recursos disponibles y de las requerimientos de la aplicación. Así, p.e. una aplicación que deba dar respuesta en tiempo real sólo puede procesar información tosca relativa a la estructura global de la escena.

A menudo se presentan auto-oclusiones o problemas de navegación que impiden obtener una cobertura completa para el objeto o la escena. La reconstrucción que sólo muestra la parte visible se etiqueta como *reconstrucción pseudo-volumétrica* y es suficiente para un gran número de aplicaciones multimedia (incluyendo visualización avanzada, entornos inmersivos, videojuegos, etc).

Estructura diferenciable del grupo de semejanzas

Para cualquier grupo "clásico" de matrices, la operación interna de multiplicación entre matrices y el cálculo del inverso de una matriz son aplicaciones "suaves", es decir, de clase C^∞; en este caso, las operaciones usuales sobre el grupo inducen una estructura de variedad diferenciable. Se dice que G es un *grupo de Lie* si es un grupo y tiene estructura de variedad diferenciable.

Todos los grupos clásicos mencionados (y también los correspondientes a las geometrías, afín, proyectiva, simpléctica, hermítica, etc) son grupos de Lie. Salvo el caso del grupo de traslaciones, dicho grupo nunca es globalmente un espacio cartesiano ordinario, aunque localmente sí lo sea (la equivalencia local esta dada por aplicaciones suaves compatibles entre sí sobre los "cambios de carta").

En particular, el grupo de las semejanzas adquiere una estructura de variedad diferenciable a partir de las operaciones usuales (interna y de paso al inverso). El conocimiento de esta estructura permite resolver problemas de interpolación o de optimización sobre la variedad soporte; un ejemplo típico es el cálculo de caminos de longitud mínima o *geodésicas* que se definen con respecto a una métrica G-invariante (es decir, invariante por la acción de G, la misma en todos los elementos de G) donde G es en este caso $SE(3;\mathbb{R})$; el argumento es válido para cualquier otro grupo clásico, si bien la métrica es específica para cada grupo.

Figura 1.17: Resultados de reconstruir una escena a partir de dos vistas con oclusiones

Una *estrategia general* para resolver este tipo de problemas consiste en linealizar el problema, es decir, pasar al espacio tangente del grupo en el punto que es isomorfo al álgebra de Lie $\mathfrak{g} := T_I G$ de G; así p.e. para el grupo ortogonal $O(n;\mathbb{R})$ su álgebra de Lie $\mathfrak{o}(n) := T_I O(n;\mathbb{R})$ es el espacio de las matrices anti-simétricas que es un espacio vectorial con generadores mucho más fáciles de calcular; la exponencial de la solución en el espacio tangente proporciona la solución en el grupo original.

Supongamos una escena simplificada con un objeto rígido sencillo en primer plano para el cual se realiza una "pequeña" transformación euclídea elemental (rotación o traslación). De forma ideal, para determinar una transformación euclídea en el plano euclídeo $\mathbb{E}^2$ (resp. espacio euclídeo $\mathbb{E}^3$), es necesario conocer un par de puntos y sus homólogos; en la práctica o bien se considera un mayor número de k-uplas de puntos y sus homólogos y se optimiza o bien se propagan los resultados asociados a cada par de puntos y sus homólogos en puntos próximos. En cualquier caso, se obtienen "nubes" de puntos en el grupo de transformaciones para las cuales es necesario identificar un "valor promedio".

Ejercicio.- Implementad el algoritmo correspondiente a medias móviles para pares de vértices homólogos de la unión de dos paralelepípedos y evaluar la transformación rígida asociada a una rotación, una traslación y una composición de ambas.

En general, la situación no es tan simple como la del ejercicio anterior. Por ello, es necesario diseñar procedimientos robustos que permitan realizar una estimación en condiciones más realistas. El apartado siguiente muestra una primera aproximación al problema

Estimación basada en PCA

El *Análisis de Componentes Principales* (PCA en lo sucesivo) es un método de estimación de tipo lineal que es una adaptación de la descomposición de valores propios (SVD) a un contexto probabilista. En [Fle83] se presenta una adaptación del algoritmo PCA para la estimación de transformaciones

ortogonales y de semejanza.

La estrategia para calcular el camino de longitud mínima (geodésica) consiste en linealizar el problema, reemplazando el camino buscado por una poligonal cuyos vértices determinan segmentos en $\mathfrak{g}$ de modo que la poligonal en $\mathfrak{g}$ tenga longitud mínima; la exponencial de cada uno de estos segmentos proporciona una curva que, al ser enlazada con las adyacentes (asociadas a los vértices más próximos), da una geodésica suave a trozos que aproxima la solución óptima. Esta estrategia se aplica tanto al movimiento del objeto como al movimiento de la cámara y es válido para cualquiera de los grupos clásicos descritos más arriba.

Para automatizar estos procesos es necesario desarrollar estrategias de aprendizaje basadas en variedades asociadas a los grupos de Lie. Cualquier cámara genera efectos de perspectiva debido a la propia naturaleza de la proyección; por ello, siempre se tiene un efecto de deformación aparente asociada a efectos de perspectiva procedentes de la captura. El modelado habitual de los efectos de perspectiva se realiza en términos de transformaciones afines; por ello, es necesario desarrollar métodos de *aprendizaje sobre grupos afines* para el modelado de los flujos visuales relacionados con el movimiento de la cámara.

El estudio de grupos afines es bien conocido en Geometría Algebraica, pero su aplicación a problemas de Reconstrucción requiere desarrollar herramientas de estimación adicionales; a la vista de la incertidumbre sobre la correcta estimación de parámetros, es conveniente utilizar métodos robustos tipo Ransac [Fis81], adaptados en este caso a variedades ambiente dadas por grupos de Lie.

Ejercicio (avanzado).- Esbozad métodos de aprendizaje no-supervisado en grupos afines y modelado robusto tipo Ransac para estimar y propagar las transformaciones obtenidas en el proceso de aprendizaje.

1.4. Estimación y optimización

La Transformación Lineal Directa (DLT) en lo sucesivo es un algoritmo que permite resolver problemas relacionados con ecuaciones lineales salvo factor de escala para una cantidad redundante de puntos. Dos casos típicos relevantes para este capítulo corresponden a la estimación de la matriz de proyección asociada a una cámara tipo pinhole y a la estimación de homografías. Para acelerar la convergencia de este método y por cuestiones de invariancia con respecto a otras transformaciones de otros marcos geométricos es conveniente contar con datos normalizados, cuestión con la que se concluye la primera subsección.

La disponibilidad de una parametrización global permite discretizar de forma más eficiente el espacio ambiente y diseñar estrategias adaptativas para la resolución de problemas de estimación y optimización. cualquier parametrización está asociada a una representación curvilínea de sistemas de coordenados cartesianos del espacio ordinario. Los métodos efectivos relacionados con restricciones invariantes por la acción de un grupo requieren una parametrización local que pueda trasladarse globalmente al grupo; esta parametrización se lleva a cabo utilizando un sistema de generadores del álgebra de Lie $\mathfrak{g} = T_I G$ del grupo de Lie G; la acción definida en un entorno de la matriz identidad I se extiende por la acción (llamada adjunta) de $\mathfrak{g}$ sobre G. La existencia de una acción de grupos de Lie sobre variedades (asociadas a las restricciones epipolares representadas por **F** y **E**) permite trasladar la parametrización local del álgebra de Lie a la variedad fundamental o a la variedad esencial.

1.4.1. Una revisión de la DLT

Una estimación precisa de datos es crucial para llevar a cabo la *auto-calibración* (ver capítulo anterior) o la *calibración* on-line a partir de dos o más vistas. Este procedimiento se sigue un esquema jerarquizado que parte del caso proyectivo como el más general (también el más débil) e incorpora herramientas más finas para estimar elementos afines (planos del infinito como hiperplanos H_∞ para la reconstrucción afín) o métricos (como la cónica ω_∞ o la cuádrica absoluta Ω_∞ para la reconstrucción euclídea).

Como la reconstrucción $3D$ debe estar bien definida salvo transformaciones en el espacio, este enfoque conlleva la necesidad de realizar la invariancia no sólo de los objetos sino de las transformaciones utilizadas; esta condición es bastante más sutil e implica la introducción de *datos normalizados* para garantizar la invariancia de los resultados obtenidos por transformaciones realizadas para resolver el problema. En esta sección se desarrolla el modelo basado en la Transformación Lineal Directa o DLT (Direct Linear Transformation). En esta subsección se presentan métodos algebraicos relacionados con la estimación de datos geométricos de cara a mejorar el rendimiento de los operadores a construir.

Un ejemplo significativo

Convergencia en modelos lineales

La cuestión de la convergencia afecta sobre todo a la invariancia de las medidas efectuadas y a su estabilidad por otras medidas cuando se realizan transformaciones que pueden afecta a proyecciones o bien a homografías (en los espacios de partida y llegada). Idealmente, una proyección se desacopla en dos homografías (que verifican restricciones adicionales). Por ello, para fijar ideas, nos limitamos a tratar de identificar el comportamiento de las homografías **H** en relación con transformaciones **T** realizadas sobre la imagen. Uno de los métodos más frecuentes para estimar **H** es el método de la

transformada lineal directa (DLT) presentado más arriba. Deseamos ver si dicho método es invariante o no con respecto a las transformaciones $\mathbf{T}$ y $\mathbf{T}'$ realizados sobre los espacios de partid y llegada.

Supongamos que $\mathbf{p}$ y $\mathbf{p}'$ son puntos homólogos vía una homografía $\mathbf{H}$ que escribimos como

$$\mathbf{H}(\mathbf{x}) = \mathbf{x}'$$

donde $\mathbf{x}$ y $\mathbf{x}'$ denotan las coordenadas homogéneas de los puntos. Supongamos que se tienen transformaciones $\mathbf{T}$, $\mathbf{T}'$ que actúan sobre los espacios de partida y llegada que denotamos mediante

$$\tilde{\mathbf{x}} = \mathbf{T}\mathbf{x} \quad , \quad \tilde{\mathbf{x}}' = \mathbf{T}'\mathbf{x}'$$

induciendo una transformación $\tilde{\mathbf{H}}$ que representamos mediante

$$\tilde{\mathbf{x}}' = \tilde{\mathbf{H}}\tilde{\mathbf{x}} \;\Rightarrow\; \mathbf{T}'\mathbf{x}' = \tilde{\mathbf{H}}\mathbf{T}\mathbf{x} \;\Rightarrow\; \mathbf{x}' = (\mathbf{T}')^{-1}\tilde{\mathbf{H}}\mathbf{T}\mathbf{x}$$

Idealmente, desde el punto de vista algebraico, tendríamos que las transformadas de la matriz $\mathbf{H}$ que representa la homografía pertenecen a la misma clase de conjugación que $\mathbf{H}$ vía las transformaciones $\mathbf{T}$ y $\mathbf{T}'$ es decir

$$\mathbf{T} \;=\; (\mathbf{T}')^{-1}\tilde{\mathbf{H}}\mathbf{T}$$

1.4.2. Parametrización

Una parametrización apropiada facilita procedimientos de búsqueda, remuestreo y estabilidad en la convergencia de algoritmos orientados hacia estimación de elementos homólogos y transformaciones o proyecciones que les afectan. De este modo, es posible mejorar la eficiencia de los algoritmos relacionados con la gestión de la información. Por ello, esta subsección tiene un carácter complementario con respecto a la anterior.

Inicialmente se puede utilizar espacio paramétrico de muestras o, por el contrario, aproximación no-paramétrica; en "situaciones reales" la hipótesis relativa a que los modelos se ajusten a solapamientos de distribuciones previamente establecidas (como las normales) es poco verosímil.

En presencia de outliers es conveniente utilizar métodos robustos. El mejor candidato para métodos robustos es (alguna variante de) RanSaC o la estimación basada en máxima verosimilitud (MLE)

El objetivo último que se pretende en esta subsección es la automatización para la puesta en correspondencia relativa pares de puntos homólogos en homografías; la estrategia propuesta se adapta obviamente a la puesta en correspondencia para proyecciones, al interpretar estas últimas como una doble clase de conjugación para la matriz de proyección canónica.

1.4.3. Métodos robustos

El enfoque desarrollado en esta subsección afecta a la robustez de los métodos para estimar transformaciones que actúan sobre los datos. Las transformaciones típicas son homografías, proyecciones (que factorizamos en homografías) y, sobre todo, relaciones estructurales asociadas a elementos homólogos en diferentes vistas. Estas relaciones se expresan mediante operadores multilineales. El primer ejemplo es la restricción epipolar que facilita la puesta en correspondencia entre elementos homólogos contenidos en dos vistas tanto para el caso afín (matriz fundamental) como para el caso euclídeo (matriz esencial).

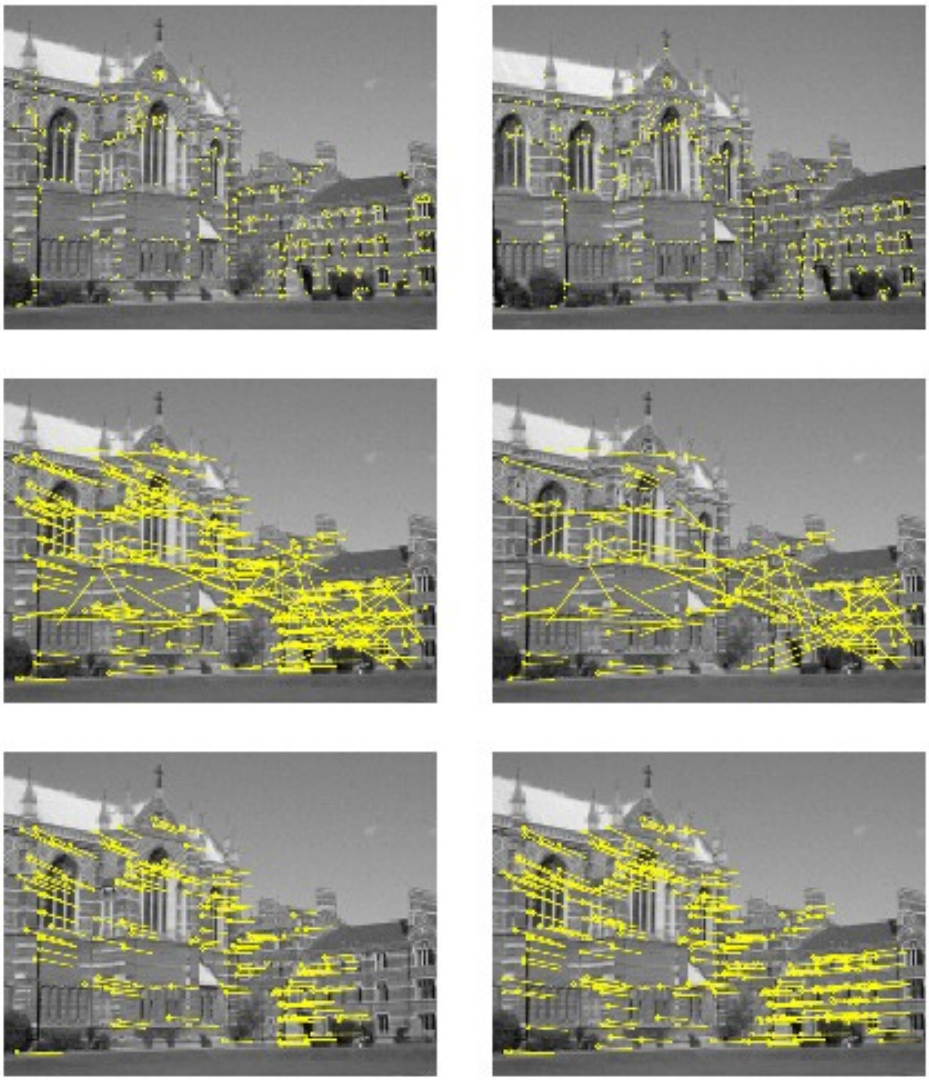

Figura 1.18: Algoritmo para la puesta en correspondencia de puntos homólogos [Pollefeys]

Automatizando la puesta en correspondencia

El objetivo de este apartado es mostrar cómo se puede generar una homografía de forma automática. Los dos procedimientos básicos de tipo local (correlación y detección de hechos) han sido presentados en la primera sección. En este apartado, se reutilizan las restricciones asociadas a la estimación de transformaciones para desarrollar una solución global del problema.

Extendiendo el marco habitual

El marco habitual para la puesta en correspondencia de datos homólogos presentado a lo largo de toda la sección supone siempre que la "línea base" b es "pequeña", es decir, el segmento que une las dos localizaciones próximas de la(s) cámara(s) tiene longitud por debajo de un umbral. De este modo, se maximiza el área de solapamiento entre las vistas y se restringe el rango de búsqueda, acelerando la convergencia del proceso. Esta hipótesis no siempre se cumple. Por ello, es importante identificar

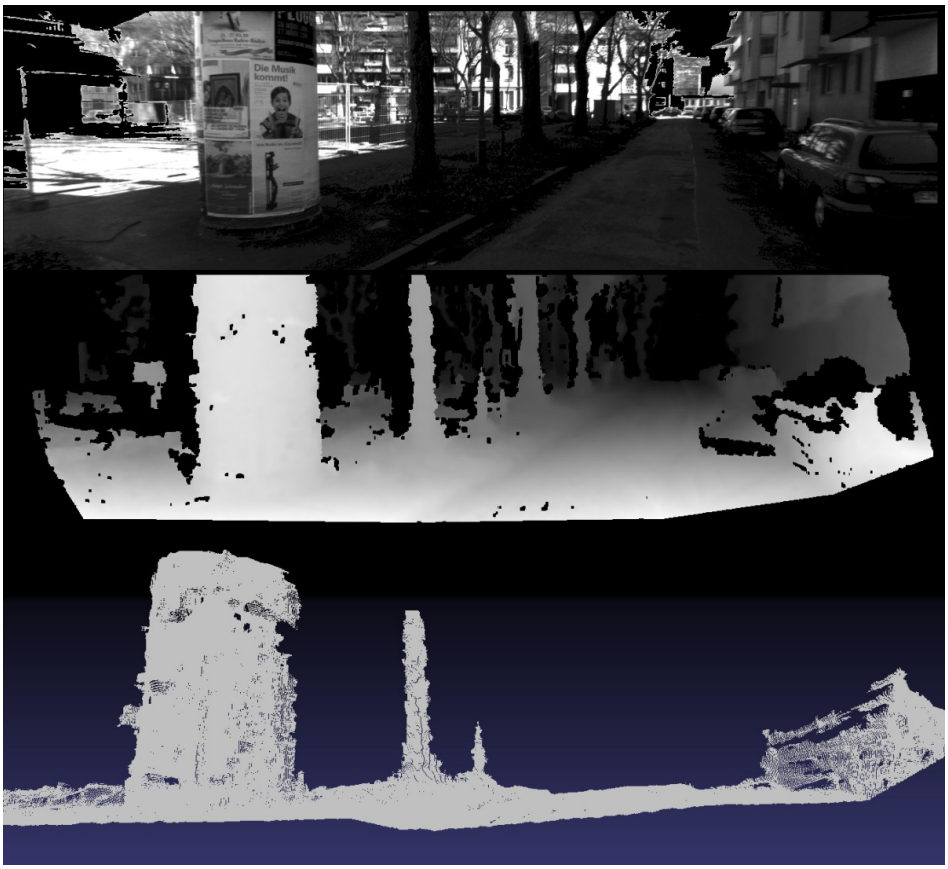

Figura 1.19: Reconstrucción 3D de un escenario urbano a partir de un par de cámaras con amplia linea base [www.cvlibs.net/software/libelas]

estrategias que permitan abordar el problema cuando b no cumple dicho requisito. Este problema se trata de forma detallada en [Zha95]

Desde un punto de vista más experimental, es conveniente usar algoritmos tipo SIFT, SURF o MSER, p.e.) para la extracción, el agrupamiento y el pegado de hechos locales.

Un análisis exhaustivo de estrategias para la detección y extracción de hechos locales se puede ver en [Tuy08]. El caso más complicado corresponde a la puesta en correspondencia para datos móviles capturados por una cámara móvil o webcam. En este caso interesa estimar la localización y el movimiento de forma simultánea (algoritmos tipo SLAM). Para una única cámara las estrategias SURF proporcionan resultados más estables en relación con SLAM. Este tópico se aborda con más detalle en el cap.5 de este módulo.