Capítulo 2: Calibración

Javier Finat y Fco Javier Delgado del Hoyo

8 de mayo de 2015

1 Calibración

En el capítulo anterior se ha mostrado la utilidad de los modelos de perspectiva para proporcionar una primera visualización de una escena a partir de un imagen. El modelo de proyección central es el más simple y representa la proyección del mundo 3D a una imagen 2D usando la intersección con el plano de imagen del haz de rayos que pasan por un centro ideal **C** de la cámara. Habitualmente, interesa que la representación sea lo más fiel posible, aunque el escenario pueda ser virtual. La precisión se formula en términos de restricciones métricas; es necesario desarrollar una metodología que permita

- validar la fidelidad en la representación, lo cual requiere "calibrar" las cámaras, es decir, corregir los errores asociados a este modelo ideal de proyección central;
- representar las posibles transformaciones a realizar sobre la escena, lo cual requiere precisar el marco geométrico para la representación.

La *calibración* consiste en realizar una estimación métrica de *características de los dispositivos* a partir de la información extraída de imágenes digitales:

- Las características pueden ser geométricas o radiométricas; a su vez, las características geométricas pueden ser internas (parámetros intrínsecos) vinculadas a la lenta ó externas (parámetros extrínsecos) vinculados a la localización de la cámara.
- La información a extraer está soportada habitualmente por "hechos salientes" (habitualmente puntos) observables en la imagen que pueden ser proyección de puntos georeferenciados (típico en Fotogrametría terrestre) ó bien "esquinas", vértices o máximos de intensidad.

Con la calibración se pretende obtener una mayor fiabilidad para estimar la información geométrica o radiométrica de la escena o los objetos presentes en la escena a partir del análisis de imagen. Los datos objetivos corresponden a objetos rígidos en la escena. La corrección de la distorsión de cada cámara permite no sólo incorporar información métrica a la escena, sino comparar datos correspondientes a diferentes dispositivos o una cámara en movimiento. Por consiguiente, la calibración ocupa un lugar central que ocupa la calibración en múltiples aspectos de la Visión [1].

Para una gran cantidad de aplicaciones es necesario disponer de medidas precisas relativas a objetos o escenas, incluyendo una acotación del error; la calibración permite identificar el sesgo y suprimir el efecto debido a distorsiones procedentes del propio equipo de medición. La estimación precisa de las características geométricas o radiométricas requiere un marco euclídeo para el espacio ambiente o para el espacio de color. Dentro de este marco es necesario corregir las distorsiones asociadas a los sensores utilizados, los dispositivos de captura (cámaras para espectro visible o no-visible) y los "artefactos" generados en las fases de procesamiento y análisis de la información.

En relación con las *distorsiones*, una distinción básica afecta a distorsiones de tipo radiométrico y geométrico, atendiendo a la distinción básica presentada en el módulo 1.

[1] De hecho, la calibración es "transversal" a todos los demás módulos del CEViC

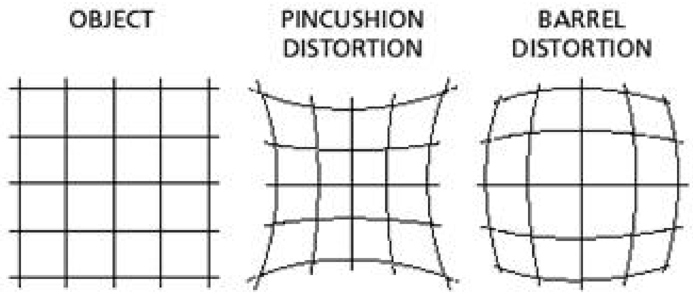

Figura 1.1: Tipos de distorsión geométrica

- Las *distorsiones geométricas* están relacionadas con deformaciones debidas a la naturaleza curvada de la(s) lente(s) contenida(s) en la cámara. Las distorsiones locales aparecen como píxeles no cuadrados. Las distorsiones globales más frecuentes son tipo tonel ó tipo cojín.

- Las *distorsiones radiométricas* están relacionadas con diferencias en la *respuesta del sensor*: Dos cámaras diferentes en la misma localización, con el mismo enfoque y bajo las mismas condiciones de iluminación dan lugar a dos imágenes con diferentes propiedades radiométricas.

Las distorsiones radiométricas proceden de una combinación de diferentes causas relacionadas con la respuesta de los sensores tales como:

- *ruido* relacionado con la energía térmica de la luz, p.e.;

- *transferencia ineficiente:* el valor almacenado es menor que el valor identificado por el sensor, p.e.;

- *no-linealidad del sensor:* los valores registrados no son proporcionales a la cantidad de luz

Actualmente estas cuestiones están siendo resueltas vía hardware de una manera cada vez más eficiente para la mayor parte de las aplicaciones. El sensor CCD o el CMOS es una parrilla rectangular de fotosensores con prestaciones crecientes; en particular, cada célula de un sensor CMOS puede ser reprogramada para obtener una respuesta selectiva y especifica a diferentes inputs.

Un inconveniente más serio es la variabilidad de respuesta radiométrica con respecto a diferentes condiciones ambientales que se aborda en la sección 4 de este capítulo.

La *estrategia general adoptada en este capítulo* consiste en resolver en primer lugar los problemas relativos a la calibración geométrica y, una vez "resueltos" estos problemas, superponer la información radiométrica (para la que se requiere asimismo una calibración). La primera sección está dedicada a exponer el modelo más sencillo de proyección que corresponde a una cámara tipo pinhole y algunas de sus variantes, incluyendo la estimación de homografías entre la parrilla de calibración y su proyección sobre la vista.

La *proyección central* desde un punto **C** (centro de la cámara) se representa inicialmente mediante un haz de rayos pasando por **C**. Esta representación se gestiona mediante diferentes modelos de perspectiva habitualmente angular o bien oblicua (dependiendo del tipo de lente utilizado); los elementos más significativos de estos tipos de proyección se abordan en la segunda subsección. A continuación se describen modelos de proyecciones más frecuentes o más realistas y se muestran métodos robustos

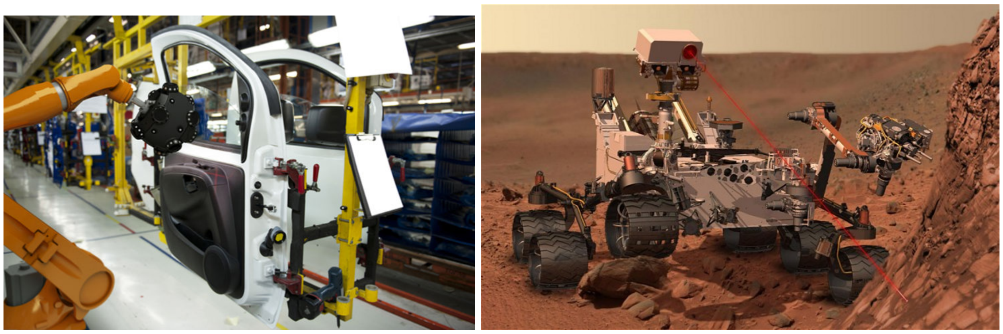

Figura 1.2: La calibración de cámaras permite desarrollar tareas precisas por robots en entornos controlados o por robots exploradores de la NASA. La diferencia esencial radica en el proceso de calibración: off-line vs on-line

para la estimación de dichas proyecciones. Para terminar se abordan algunos aspectos ópticos más avanzados asociados a la distorsión en regiones periféricas de la imagen.

Una vez aclarado el marco general, el *objetivo fundamental* de la *calibración de cámaras* es facilitar recursos (dispositivos físicos, sensores, modelos matemáticos y herramientas software) para poder realizar medidas de objetos a partir de información contenida en imágenes. De una forma más técnica, la calibración de cámaras es el proceso de *estimación de parámetros* de un sensor métrico pasivo que permite relacionar la información de la imagen y del mundo mediante la puesta en correspondencia entre puntos.

La *Fotogrametría terrestre* tradicional utiliza plantillas o parrillas de calibración para una estimación off-line de la calibración de una cámara. En ausencia de plantillas, la *auto-calibración* de una cámara se realiza con referencia a un objeto rígido que se reconstruye módulo transformaciones de semejanza (rotaciones, traslaciones y escala). En cualquier caso (Calibración Fotogramétrica off-line o auto-calibración on-line) el carácter preciso de la información a extraer requiere

- identificar modelos matemáticos “exactos” (basados en información minimal);
- diseñar una metodología (que maximice la bondad del ajuste y minimice los errores);
- realizar una implementación computacional eficiente que garantice control y minimización del error.

La condición de minimalidad relativa a los datos es la más apropiada para resolver de forma exacta un problem que esté bien planteado pues permite obtener una solución única. La solución única obtenida proporciona una inicialización en la resolución del problema que sirve para verificar si los otros elementos (puntos ó líneas) candidatos a homólogos satisfacen las restricciones de la reconstrucción realizada [2]

A menudo los datos están corrompidos por el ruido o bien los elementos (puntos o rectas) están próximos a configuraciones degeneradas. Por ello, es necesario trabajar con información redundante. La estrategia básica consiste en resolver el problema de forma exacta para una colección minimal de elementos homólogos y aplicar la solución obtenida a otros pares formados por k-tuplas de elementos homólogos.

[2] Ver http://cmp.felk.cvut.cz/minimal/index.php para una colección de métodos y algoritmos d egran utilidad para cuestiones de Reconstrucción 3D.

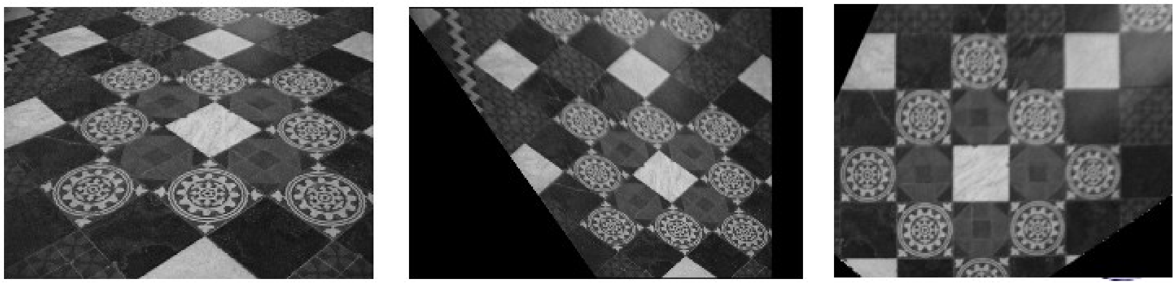

Figura 1.3: Secuencia de imágenes del suelo rectificado para reconstrucción 3D [ver Hartley & Zisserman o el curso de Pollefeys]

La implementación se basa en algoritmos para la (auto-)calibración; se sigue un orden de complejidad creciente, empezando por los que presentan menor complejidad computacional y mayor versatilidad en relación con la fusión de información procedente de diferentes fuentes. En Fotogrametría se supone disponible información georeferenciada sobre 3 puntos de la escena que se proyecta sobre cada plano de imagen (lo cual proporciona una referencia afín); el método RANSAC [3] permite calcular la POSE de la cámara. Una revisión y extensión de este método se puede ver en la tesis de P.Torr (Oxford, 1997

La información métrica proporcionada por la calibración facilita la interacción entre diferentes tipos de dispositivos (incluyendo aplicaciones a la Robótica asistida por Visión) en escenas estructuradas. Frecuentemente, es necesario incorporar información procedentes de imágenes estáticas, vídeo, imágenes aéreas, infrarrojos o diferentes tipos de láser. La fusión de esta información se realiza teniendo en cuenta la rigidez del objeto usando técnicas de auto-calibración. Esta fusión de información permite mostrar modelos geométricos precisos para la Reconstrucción Euclídea que permiten generar nuevas vistas sintéticas de un objeto con geometría rígida [4].

De acuerdo con las observaciones anteriores, el análisis de dispositivos, modelos y sus fundamentos teóricos se aborda en la primera sección. La calibración comparte objetivos y, por consiguiente, métodos con la Fotogrametría que se aborda en la sección 2 desde el punto de vista de la calibración. La extracción de información debe ser independiente de dispositivos y de la localización (posición y orientación) de la cámara, pues frecuentemente no se dispone de dicha información; este requisito conduce al desarrollo de métodos basados en la Geometría Proyectiva que muestran cómo obtener información métrica incluso a partir de información contenida sólo en imagen; estas cuestiones se abordan en la sección 3. La superposición de propiedades radiométricas relativas a la longitud de onda (color) y sus potencias (reflectancia, irradiancia, etc) requiere una calibración radiométrica que se lleva a cabo en la sección 4.

Desde el punto de vista de la Reconstrucción 3D, la calibración pretende obtener modelos precisos tridimensionales de objetos B_α o de escenas de una manera lo más exacta posible, indicando el error y, a poder ser, las características de su distribución; esta cuestión afecta a la Reconstrucción Euclídea que se desarrolla en la sección 5, incorporando elementos de modelado geométrico más avanzado basado en la geometría del "espacio" de rectas. Por último, las necesidades planteadas por la producción de imágenes sintéticas de alta resolución (vinculadas a la producción de contenidos multimedia, p.e.) lleva a un control más preciso (análisis y síntesis, para su reproyección) de las condiciones radiométricas (color, iluminación) que se aborda en la última sección del capítulo.

[3] [Fis81] M. A. Fischler, R. C. Bolles: "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography". *Comm. of the ACM*, Vol 24, 381-395, 1981)

[4] El caso correspondiente a objetos de geometría variable es bastante más complicado y se aborda en el módulo 5

La referencia más completa para este módulo es [Har04] [5]. Las lecturas 2 a 8 del curso de Marc Pollefeys [6] proporcionan materiales gráficos más intuitivos, si bien los contenidos frecuentemente son más avanzados que los expuestos en este curso. Se recomienda usar ambos recursos a lo largo del módulo 2.

[5][Har04] R.Hartley and A.Zisserman: *Multiple View Geometry* (2nd ed), Cambridge Univ. Press, 2004

[6]Ver https://www.cs.unc.edu/ marc/mvg/slides.html

1.1. Modelos lineales para la calibración

El objetivo de esta sección es mostrar los modelos geométricos subyacentes a la calibración y mostrar las metodologías más relevantes para corregir los defectos o distorsiones asociados a proyecciones y modelos de perspectiva. Para poner en correspondencia datos que afectan a la imagen, la cámara y el mundo es necesario escribir las ecuaciones de proyección entre los sistemas de coordenadas del plano de la cámara y el mundo, corrigiendo las distorsiones asociadas a la cámara.

En primer lugar se presenta el modelo más simple que corresponde a una proyección lineal central a la que se etiqueta como *cámara pinhole*. A continuación se describen las ecuaciones de la proyección asociada y se analizan diferentes tipos de proyecciones. Obviamente, estos modelos son una simplificación que no recoge aspectos ópticos muy relevantes que se describen en la última subsección para motivar algunos desarrollos más realistas que se presentan en la sección siguiente.

Cámara tipo pinhole

El primer paso de la calibración consiste en identificar los puntos homólogos de una vista $\mathbf{p}_i$ en relación con los puntos d control $\mathbf{P}_i$ para construir la matriz $\mathbf{M}_\pi$ asociada a una proyección π con centro en $\mathbf{C}$ que denotaremos $\mathbf{M}_\mathbf{C}$. El modelo más simple de proyección es una proyección central para una cámara tipo pinhole.

Por consiguiente, el *objetivo* de esta subsección es resolver el siguiente *problema* Dada una colección de pares de puntos $(\mathbf{P}_i, \mathbf{p}_i$, calcular $\mathbf{M}_\mathbf{C}$ tal que $\mathbf{M}_\mathbf{C}(\mathbf{P}_i) = \mathbf{p}_i$ en términos de los *parámetros de la cámara*

Habitualmente desconocemos la posición del origen en coordenadas píxel que denotamos mediante $u = \alpha\frac{x}{z} + u_0$, $v = \beta\frac{y}{z} + v_0$

Denotemos mediante θ el ángulo que forma el eje vertical Ov con resp a Ou en el plano de imagen.

$$u = \alpha\frac{x}{z} - \alpha cot(\theta)\frac{y}{z} + u_0 \quad , \quad v = \frac{\beta}{sen\ \theta}\frac{y}{z} + v_0$$

En coordenadas homogéneas re-escribimos el resultado de la proyección como sigue

$$\mathbf{p} = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{z}\begin{pmatrix} \alpha & -\alpha cot(\theta) & u_0 & 0 \\ 0 & \frac{\beta}{sen\ \theta} & v_0 & 0 \\ 0 & 0 & 1/f & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \frac{1}{z}[\mathbf{K} \mid \mathbf{0}]\mathbf{P}$$

En el primer apartado de esta subsección se especifican los parámetros de la cámara (intrínsecos y extrínsecos) a estimar y se comenta el enfoque de la Fotogrametría Terrestre que motiva buena parte de los desarrollos de este capítulo. Tras esta introducción se muestra la utilidad de las imágenes rectificadas para la calibración. Por último, se comentan algunos problemas relacionados con el enfoque y desenfoque que modifican los parámetros, introduciendo distorsiones adicionales que es preciso delimitar y corregir.

Elementos significativos

Una cámara tipo pinhole se basa en un modelo de *proyección central* a través de un punto *material* $\mathbf{C}$ que se toma habitualmente como origen para el sistema de coordenadas centrado en la cámara. La adopción de este modelo tiene las siguientes consecuencias desde el punto de vista de la Óptica Geométrica:

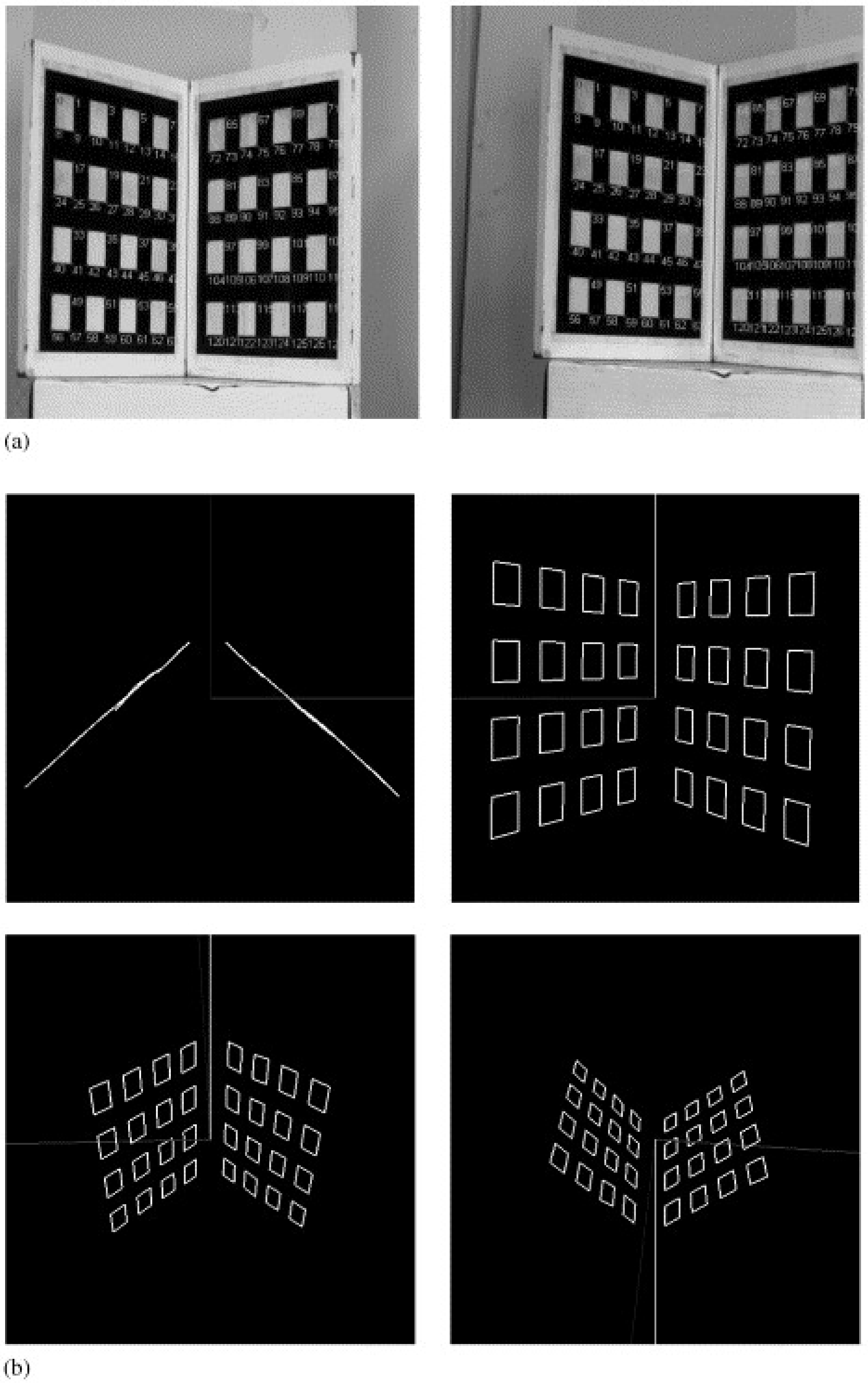

Figura 1.4: Ejemplo de parrilla de calibración con geometría conocida e imágenes en perspectiva de las parrillas reconstruidas

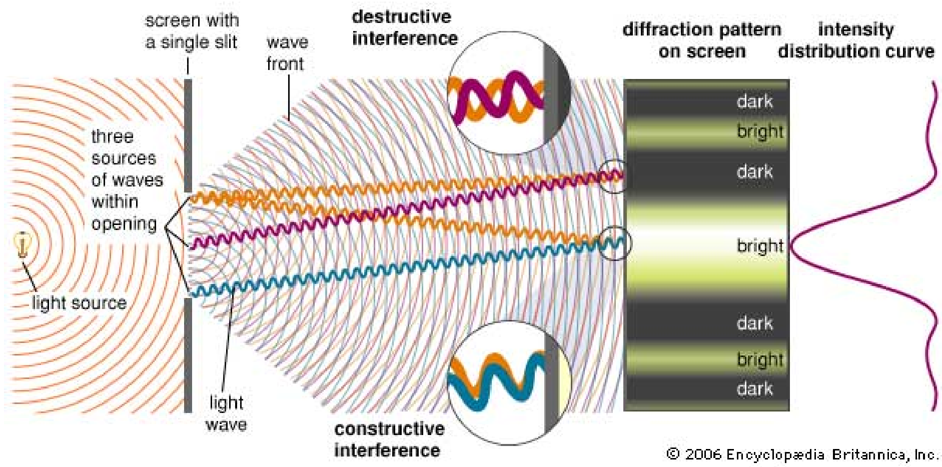

Figura 1.5: Ilustración del fenómeno de difracción

- El eje Oz mundo está alineado inicialmente con el *eje óptico*: la línea a través de $\mathbf{C} = \mathbf{O}$ es perpendicular al plano de imagen.
- Los ”rayos” (rectas orientadas) r_i que conectan $\mathbf{P}_i$ con el centro óptico $\mathbf{C}$ se proyectan sobre $\mathbf{r}_i \in \Pi_\mathbf{C}$

Un modelo más realista debe reemplazar el centro de proyección por un punto “más grueso” que se controla por la apertura. La modificación en la “apertura” de la cámara tiene consecuencias sobre la imagen:

- Si **C** es “más grueso” entonces los rayos próximos r_i se promedias, lo cual da lugar a efectos de *emborronamiento*
- Si **C** es ”diminuto” entonces aparecen efectos de *difracción*
- En una cámara oscura sólo un pequeño número de rayos alcanza $\Pi_\mathbf{C}$

En la calibración geométrica ”sólo” se corrigen defectos asociados a la captura de la geometría. Para obtener una representación con una mayor realismo es necesario utilizar la calibración radiométrica que permita incorporar el comportamiento de la luz con la materia bajo diferentes condiciones ambientales.

Campo de visión

Los *elementos ópticos más significativos* para el Campo de visión (FoV) son:

1. *Diámetro efectivo d* de la lente: Trozo de la lente alcanzable por los rayos de luz

$$tg\ w = \frac{d}{2f} \Rightarrow FoV = 2arc\ tg\frac{d}{2f}$$

   El último miembro muestra que el diámetro efectivo de una lente y la longitud focal determinan el campo de visión.

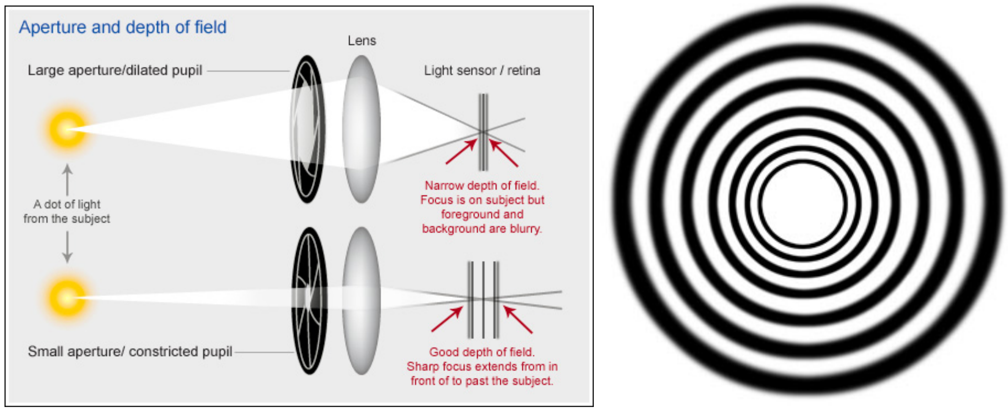

Figura 1.6: Ilustración del campo de profundidad y su efecto sobre círculos de confusión superpuestos sobre el punto focal (según vistaview360 y mononeil)

2. El *plano focal* es el plano al que enfoca la cámara. Los elementos situados por delante o por detrás en la dirección del eje visual están ligeramente desenfocados y presentan un "halo" que se modela en términos de círculos de confusión.

3. *Círculos de confusión:* Corresponde a la periferia de la zona (centrada en el punto focal) que está enfocada. Supongamos que $z = d(\mathbf{C}, \Pi_{\mathbf{C}})$ (distancia nominal) y z' es la distancia de enfoque para una lente "delgada" [7]. El radio r_c de las circunferencias de confusión es

$$r_c \ := \ \frac{d}{\mid z' - z} \mid$$

4. *Campo de profundidad:* rango en el que la cámara produce imágenes aceptables: depende de forma inversa del diámetro efectivo d y la longitud focal f. La discretización del campo de profundidad en los niveles "más significativos" es clave para linealizar el problema reproyectando la información sobre dichos niveles y resolviendo el problema de paraperspectiva asociado a cada nivel.

El *pegado simultáneo* de imágenes capturadas por diferentes cámaras requiere solapamiento suficiente de campos visuales, lo cual requiere resolver problemas de calibración geométrica y radiométrica que se expresan inicialmente en términos matriciales. El modelo más simple corresponde a estimar la matriz de proyección para cada vista y llevar el resultado de una matriz de proyección a otra mediante una trasformación del espacio; esta estrategia facilita una solución inicial en el entorno del eje óptico dentro del campo de visión (FoV: Field of View), que es necesario refinar en aproximaciones más precisas.

Los modelos que permiten evaluar las distorsiones siguen un orden creciente de complejidad. Los más sencillos están dados por transformaciones lineales salvo escala, es decir, colineaciones. Más adelante se consideran otros dados por aplicaciones no-lineales.

Las colineaciones permiten medir las coordenadas de imagen $(x, y)^\top$ de puntos $\mathbf{p}$ en imagen como proyección de puntos conocidos $\mathbf{P}$ con coordenadas $(X, Y, Z)^\top$. Permiten relacionar las coordenadas de pares de puntos $(\mathbf{p}, \mathbf{P})$ representados por $(x, y, ; X, Y, Z)$ con las incógnitas $(X_C, Y_C, Z_C, \omega, \phi, \kappa)^\top$ donde $(\omega, \phi, \kappa)^\top$ representan los ángulos de Euler para estimar la orientación.

[7]Ver http://www.circleofconfusion.ie/ para detalles y complementos

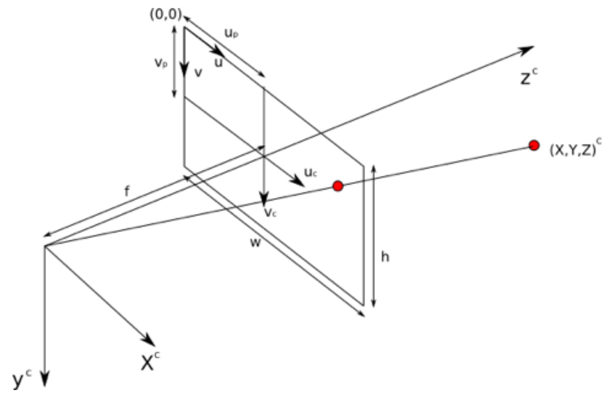

Figura 1.7: Transformación de coordenadas de imagen en coordenadas globales

En el caso general no-lineal es necesario estimar los parámetros de distorsión (radial y tangencial) geométrica mediante polinomios de grado bajo procedentes de una truncación del desarrollo de Taylor para modelos de distorsión más generales.

Transformaciones entre sistemas de coordenadas

En la aproximación lineal, hay que resolver el siguiente *problema*; Relacionar el sistemas de coordenadas mundo W con el de la cámara $\mathbf{C}_\alpha$. El espacio proyectivo y sus transformaciones proporcionan el marco geométrico más general para el estudio de proyecciones centrales.

Como se desea obtener una información métrica, se considera un sistema de *coordenadas afines* (no homogéneas) en el complementario del hiperplano $X_4 = 0$ caracterizado por la condición $X_4 \neq 0$. Dividiendo entre X_4, podemos suponer que la última coordenada es siempre no-nula. Por ello, la transformación asociada a un desplazamiento rígido (rotación y traslación) se describe como

$$\mathbf{P}^C = \mathbf{R}_W^C \mathbf{P}^W + \mathbf{t}_W^C$$

La expresión matricial afín de esta transformación está dada por:

$$\begin{pmatrix} u\mathbf{P}^C \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R}_W^C & \mathbf{t}_W^C \\ \mathbf{O}^\top & 1 \end{pmatrix} \begin{pmatrix} \mathbf{P}^W \\ 1 \end{pmatrix}$$

Expresión analítica de la matriz de proyección

En el marco proyectivo, la matriz asociada a la proyección $\pi : \mathbb{P}^3 \to \mathbb{P}^2$ está dada por una 3×4-matriz M_π en la que podemos suponer que la primera 3×3-caja tiene un determinante no-nulo. Deseamos describir $\mathbf{M}_\pi$ en términos de los 11 parámetros de calibración asociados a la cámara euclídea de una forma sintética. Para ello, denotemos mediante $\mathbf{r}_i$ a las filas de $\mathbf{M}$ para $1 \leq i \leq 3$. Sea θ el ángulo formado entre los ejes Ou y Ov en el plano de imagen. Entonces

$$\begin{pmatrix} \alpha\mathbf{r}_1^\top - \alpha cot(\theta)\mathbf{r}_1^\top + u_0\mathbf{r}_3^\top & \alpha\mathbf{t}_x - \alpha cot(\theta)\mathbf{t}_y + u_0\mathbf{t}_z \\ \frac{\beta}{sen\ \theta}\mathbf{r}_2^\top + u_0\mathbf{r}_3^\top & \frac{\beta}{sen\ \theta}\mathbf{t}_y + v_0\mathbf{t}_z \\ \mathbf{r}_3^\top & \mathbf{t}_z \end{pmatrix}$$

donde $\mathbf{t} = (\mathbf{t}_x, \mathbf{t}_y, \mathbf{t}_z)^\top$ es el vector traslación. La matriz de proyección es un objeto algebraico abstracto y por tanto afecta a todos los vectores del espacio; sin embargo la visualización está restringida al

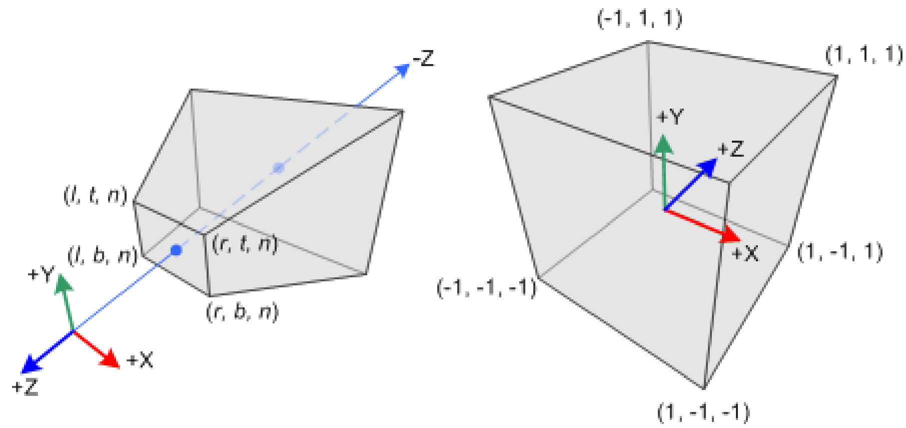

Figura 1.8: Diferencias entre la proyección en perspectiva y ortogonal

campo de visión descrito más arriba. La figura 1.8 ilustra esta idea simple pero que es necesario tener siempre en mente.

Revisitando modelos proyectivos de proyecciones

Los objetos fundamentales para el modelado geométrico son la *matriz de proyección* para diferentes tipos de proyección y la *homografía*; para estimarlas usamos cuadriláteros de referencia, incluyendo una estimación de la deformación de estos últimos. Para fijar ideas, se adopta una estrategia de complejidad creciente con respecto a estos elementos de referencia. Algunas propiedades generales a tener en cuenta a lo largo de esta subsección son

- Cualquier proyección central π_C está representada por una 3×4-matriz M_{π_C} que tiene rango 3 [8]
- El núcleo de la aplicación lineal es el centro **C** de la matriz de proyección [9]
- Existen diferentes estrategias para estimar los 11 parámetros (doce salvo factor de proporcionalidad) para la matriz de proyección.

La *hipótesis de partida* consiste en que los datos más significativos para la calibración están contenidos en un "plano dominante" (habitualmente el plano focal) bien delimitado en la escena que se proyecta sobre una región de la imagen. En el enfoque fotogramétrico, dicho plano es el de la parrilla de calibración; en los procedimientos de auto-calibración, puede ser cualquier plano del diedro ortogonal en el que se disponga de unidades de medida a lo largo de las rectas que resultan de intersecar el ángulo diedral con un plano perpendicular al eje del diedro. El plano dominante puede tener una orientación perpendicular (visión) fronto-paralela o bien oblicua con respecto a la línea de visión.

El *objetivo general* vinculado a la *calibración extrínseca* es la auto-localización (posición y orientación). La generalidad del marco proyectivo permite realizar síntesis de nuevas imágenes mediante

[8] Si el rango fuera más bajo, la imagen no sería un plano sino una recta

[9] En particular, para la proyección canónica dada por $(I \mid \mathbf{O})$ el núcleo es el punto con coordenadas afines $(0,0,0,1)^\top$

Figura 1.9: Las transformaciones euclídeas conservan distancias, pero la proyección sobre el plano de imagen genera una distorsión aparente

la (discretización de) transformaciones [10]. Los tópicos que abordamos en esta subsección son los siguientes:

1. Proyección ortogonal
2. Perspectiva débil y proyección escalada
3. Formulación de las ecuaciones de perspectiva
4. Paraperspectiva

Un *modelo de perspectiva* es una aproximación (no lineal) a las matrices de proyección. Una matriz de proyección permite visualizar la reconstrucción 3D de una escena. El carácter no-lineal de los modelos de perspectiva ralentiza la convergencia de los algoritmos utilizados para la gestión del modelo. Por ello, para aplicaciones que requieran aplicaciones on-line (incluyendo posible interacción con el entorno), es necesario introducir *modelos lineales de tipo afín*, entre los cuales cabe destacar el modelo de perspectiva débil y el de paraperspectiva como aproximaciones lineales a los modelos de perspectiva [11].

La descomposición general de una transformación proyectiva se ha mostrado en el capítulo anterior. En este capítulo se comentan casos relevantes que facilitan la actualización en tiempo real y la generación de proyecciones adaptadas a la localización del observador.

Proyección ortogonal

El *tipo más sencillo de proyección* corresponde a vistas fronto-paralelas asociadas a una proyección ortogonal del plano dominante sobre el plano de la cámara. En este caso, como el eje óptico es perpendicular a ambos planos y es suficiente corregir la distorsión radial.

Si tomamos $\mathbf{C}$ como origen de coordenadas y $Z = f$ como plano de imagen, cualquier. El punto $\mathbf{P} = (X.Y,Z)^\top$ se proyecta sobre $\mathbf{p} = (x,y,f)^\top$; en particular, en el plano (X,Z) formado por el eje óptico y la proyección de la recta $\overline{\mathbf{C},\mathbf{P}}$, por semejanza de triángulos se tiene $\frac{x}{f} = \frac{X}{Z}$ por lo que $x = f\frac{X}{Z}$ (dibujadlo); razonando análogamente para el plano (Y,Z) podemos escribir

[10] La simulación del movimiento de cámara mediante discretización a lo largo de un camino en el espacio de transformaciones se aborda en el capítulo 4 en relación con la Reconstrucción a partir de una sola vista

[11] El modelo de paraperspectiva es muy útil para escenas con objetos situados en diferentes planos de profundidad, pues permite desacoplar la información correspondiente a dichos planos y tratarla por separado. Esto resulta especialmente útil en Visión Estéreo Dinámica (ver módulo 5) o bien para aplicaciones de Video 3D (asignatura optativa del máster de Ingeniería Informática)

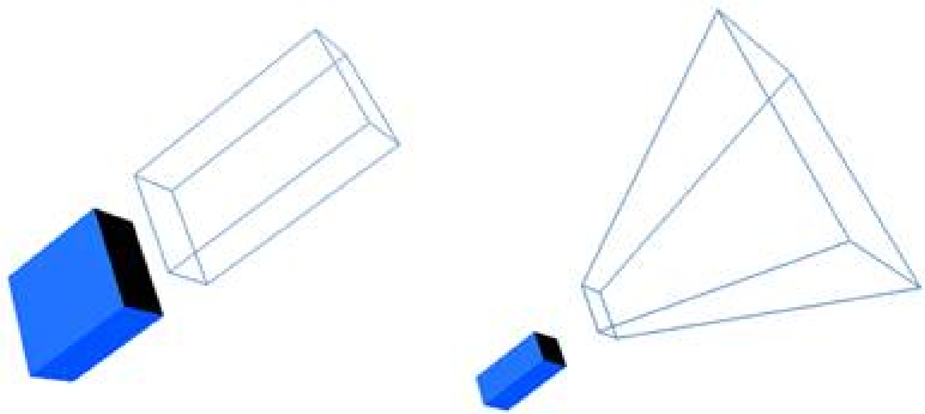

Figura 1.10: Proyección ortogonal frente a la de perspectiva

$$(x, y) = (f\frac{X}{Z}, f\frac{Y}{Z}) = f(\frac{X}{Z}, \frac{Y}{Z})$$

El carácter no-lineal procede de la división entre Z. Nótese que si los puntos observados están situados en un plano de profundidad constante $Z = Z_0$, entonces las ecuaciones de la proyección son *lineales* [12]

Usando coordenadas homogéneas $[x' : y' : z']$ (con $(x, y) = (x'/z', y'/z')$ para el plano de imagen, podemos escribir la ecuación de la proyección perspectiva sobre un plano ortogonal como

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

donde la matriz de la proyección sobre el plano ortogonal al eje óptico se puede reescribir como

$$\begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{pmatrix}$$

Las figuras siguientes ilustran la proyección ortogonal en relación con modelos de perspectiva angular presentados en el capítulo anterior y una vista en proyección ortogonal que resulta de gran utilidad para representa objetos en un rango con ”baja variabilidad” para la profundidad

Proyección ortográfica escalada

La matriz de la *proyección ortográfica* (3 puntos de fuga a distancia infinita) esta dada por

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

[12]Los modelos de paraperspectiva utilizan este hecho para realizar una estimación tosca de la profundidad, descomponer el espacio en diferentes planos de profundidad constante, abatir la información sobre dichos planos y resolver el problema lineal para cada plano por separado y pegar los datos. Es claro que este aproximación sólo es válida para cierto tipo de escenas con planos muy diferenciados y que las transiciones entre planos dan lugar a discontinuidades difíciles de controlar

(verificadlo como ejercicio): las líneas proyectantes son paralelas al eje óptico y no hay modificación de escala (es una representación 1 : 1). La matriz de proyección de perspectiva débil para los puntos situados en el plano $Z = Z_0$ tiene coordenadas en imagen

$$u = s\ ,\ v = sy\ ,\ s = \frac{f}{Z_0}$$

se puede representar como una "perturbación" lineal de la proyección ortográfica dada por

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & f/Z_0 \end{pmatrix}$$

La interpretación geométrica es muy simple: A diferencia de la proyección ortogonal, la proyección de perspectiva débil modifica la escala de forma isótropa (idéntica a lo largo de diferentes direcciones en el plano) de acuerdo con el factor de escala f/Z_0.

Esta descripción muestra que Si se dispone de un dispositivo de rango lineal (haz de infrarrojos, p.e.), entonces la estimación de la calibración para modelos de perspectiva débil se puede realizar en tiempo real. Este hecho tiene aplicaciones relevantes en cuestiones para producción de contenidos visuales, interacción basada en Wii, realidad aumentada o videojuegos, p.e.

La *proyección ortográfica escalada* es una proyección ortográfica sobre el plano $z = 0$ salvo escala $s = f/Z_0$. En este caso, idealmente se tiene que $f \to \infty$. Por ello, la matriz de proyección presenta la forma

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & Z_0/f \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

En el marco proyectivo general, este tipo de proyección lleva el plano afín $(su, sv, s)^\top$ con $s \neq 0$ en un plano cartesiano $\frac{U}{W}, \frac{V}{W})^\top = (u, v)^\top$

La *proyección ortogonal salvo escala* se describe en términos de la asignación

$$[X_0 : X_1 : X_2 : X_3] \mapsto (u, v)^\top = (sx, sy) \quad \text{con } s = f/Z_0$$

Este tipo de proyección es apropiada para objetos próximos al eje óptico. En este caso, las distorsiones son "pequeñas" cuando la dimensión del objeto es "pequeña" ($\delta z < \overline{Z}/20$) en relación con la distancia media $\overline{Z}$ con respecto a la cámara.

Una *ventaja de la proyección ortogonal escalada* consiste en que permite generar efectos globales (no locales) de perspectiva y, por consiguiente, agrupar puntos según profundidad relativa. Esta ventaja resulta de gran utilidad para generar una representación tosca on-line a bajo nivel de escenas que se capturan sobre un dispositivo móvil (una plataforma robótica o una kinect, p.e.), lo cual facilita una interacción con el entorno y el desarrollo de entornos inmersivos simplificados.

Un *inconveniente de la proyección ortogonal escalada* es que es un modelo equivocado desde el punto de vista métrico; en otras palabras no resulta apropiada para aplicaciones que requieran precisión (fotogrametría, cirugía asistida)

No obstante, este tipo de proyección puede resultar útil para *inicializar* procesos de reconstrucción a bajo nivel. Por ello y debido también a las múltiples aplicaciones de este tipo de proyección (para propósitos lúdicos o educativos), realizamos una *reinterpretación cartesiana* que resulta fácilmente implementable:

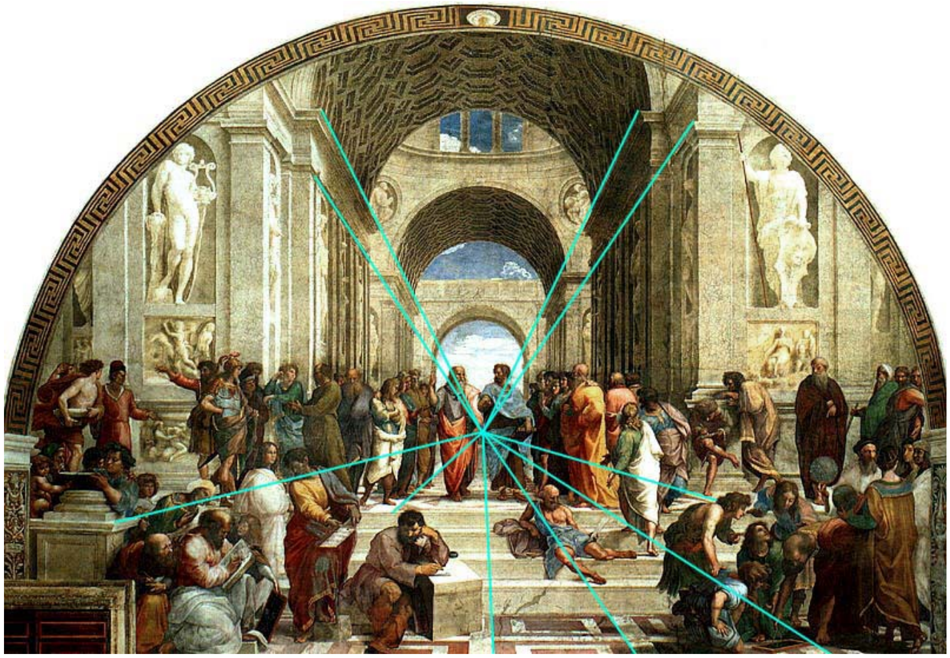

Figura 1.11: Aplicación de la perspectiva arquitectónica sobre la Escuela de Atenas de Rafael

Denotemos mediante (α_u, α_v) los factores de escala vertical y horizontal, y sean (u_c, v_c) las coordenadas del punto principal. Entonces

$$p_i = (u_i, v_i) \mapsto (x_i, y_i) := (\frac{u_i - u_c}{\alpha_u}, \frac{u_i - u_c}{\alpha_u})$$

Por ello, la *matriz de proyección escalada* está dada por

$$\begin{pmatrix} su_i \\ sv_i \\ s \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_c & 0 \\ 0 & \alpha_v & v_c & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{r}_1 & t_x \\ \mathbf{r}_2 & t_y \\ \mathbf{r}_3 & t_z \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}$$

Con esta notación se tienen las siguientes relaciones entre coordenadas cartesianas

$$(x_i, y_i) = (\frac{\mathbf{r}_1^\top \mathbf{P}_i + t_x}{\mathbf{r}_3^\top \mathbf{P}_i + t_z}, \frac{\mathbf{r}_2^\top \mathbf{P}_i + t_y}{\mathbf{r}_3^\top \mathbf{P}_i + t_z})$$

Proyección oblicua

Los haces de elementos paralelos contenidos en un plano no-ortogonal al eje óptico se cortan en puntos a distancia finita en la imagen. Si se adopta el punto de vista proyectivo, para escribir la ecuación de una proyección oblicua general, hay que reemplazar la 3×4-matriz de proyección correspondiente a la proyección ortogonal por una matriz que resulta de multiplicar a la izquierda por una 3×3 -matriz arbitraria (representando una homografía de $\mathbb{P}^2$) y a la derecha por una 4×4 -matriz arbitraria (representando una homografía de $\mathbb{P}^3$).

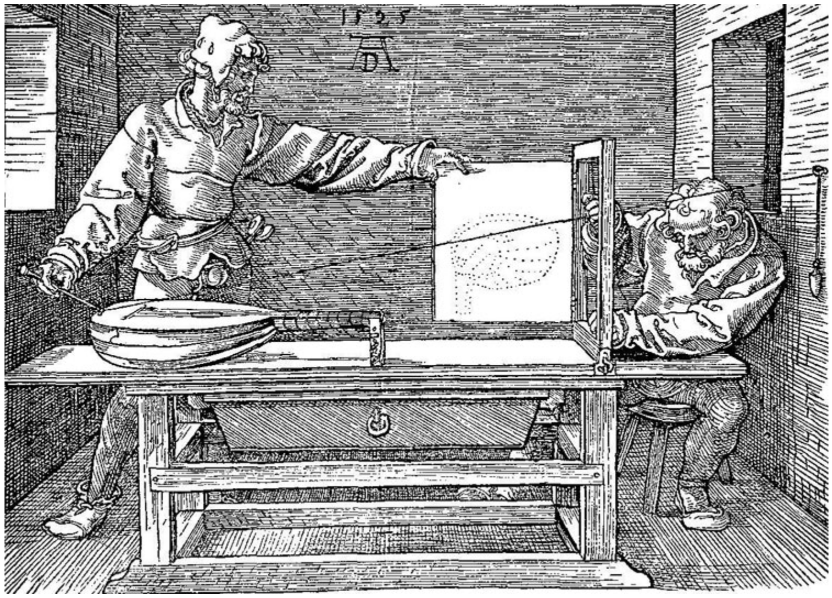

Figura 1.12: Grabado de Durero con la representación en perspectiva del laúd y su ilustración para la justificación matemática discreta del modelo

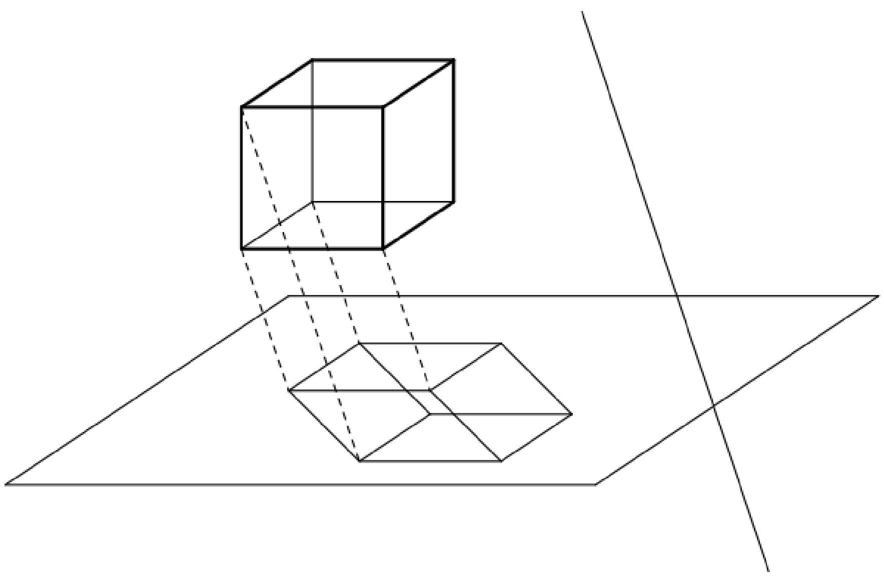

Figura 1.13: Ilustración de la proyección oblicua

Como la transformación en el espacio de llegada representa una modificación en los parámetros intrínsecos, la 3×3-matriz debe ser una matriz triangular superior; análogamente, como la transformación en el espacio de partida representa una transformación de semejanza salvo escala en los parámetros extrínsecos, la 4×4-matriz debe representar el producto semi-directo de una rotación y una traslación.

La estimación de parámetros más generales requiere una "rectificación proyectiva" que se aborda más abajo (ver Fig.14).

El plano focal cuya imagen se considera en cada vista se modela como (un trozo de) $\mathbb{P}^2$. Las transformaciones lineales más generales de dicho plano proyectivo están dadas por homografías, es decir, por elementos de $\mathbb{P}GL(3;\mathbb{R})$. Fijada una referencia de $\mathbb{P}^2$, una homografía $\mathbb{H}$ está dada salvo escala por

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}$$

que reescribimos como un sistema de tres ecuaciones

$$y_i = h_{i0}x_0 + h_{i1}x_1 + h_{i2}x_2 \qquad 0 \leq i \leq 2$$

Deshomogeneizamos con respecto a y_0:

$$\frac{y_1}{y_0} = \frac{h_{10}x_0 + h_{11}x_1 + h_{12}x_2}{h_{00}x_0 + h_{01}x_1 + h_{02}x_2}, \quad \frac{y_2}{y_0} = \frac{h_{20}x_0 + h_{21}x_1 + h_{22}x_2}{h_{00}x_0 + h_{01}x_1 + h_{02}x_2}$$

Dos puntos ($\mathbf{p} \in$ y $\mathbf{q}$) son *homólogos* si $\exists \mathbf{H} \in \mathbb{P}GL(3;\mathbb{R}) \mid \mathbf{q} = \mathbf{Hp}$. El problema a resolver consiste en *estimar* $\mathbf{H} \in \mathbb{P}GL(3;\mathbb{R})$. Este problema se resuelve por el método de coeficientes indeterminados usando puntos homólogos representados por 4-tuplas de vértices detectados en imagen.

Proyección afín

Este apartado es un caso particular del apartado anterior y por ello sólo se esboza la descripción dejando los detalles para que sean completados por el alumno interesado.

Supongamos fijado un punto arbitrario $\mathbf{p}_0 = (x_0), y_0, z_0)^\top$. Linealizamos las ecuaciones de proyección perspectiva $(u, v) = (f\frac{x}{z}, f\frac{y}{z})$ en torno a $\mathbf{p}$, se obtiene

$$u = \frac{f}{z_0}(x - x_0) + \frac{f x_0}{z_0^2}(z - z_0)$$

por lo que

$$u = [\frac{f}{z_0}, 0, \frac{f x_0}{z_0^1}, \frac{f x_0}{z_0}\frac{f x_0}{z_0}] = \frac{f}{z_0}[1 \ 0 \ \frac{x_0}{z_0} \ 0]$$

y análogamente para v. Usando esta expresión se obtiene la ecuación de la proyección afín que escribimos como

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} 1 & 0 & \frac{x_0}{z_0} & 0 \\ 0 & 0 & 0 & \frac{z_0}{f} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix}$$

Modelos de perspectiva y calibración

En esta subsección se revisan las ecuaciones de perspectiva de cara a su conversión en términos de coordenadas píxel (con objeto de facilitar su implementación computacional). A continuación se presenta el modelo de paraperspectiva que proporciona una aproximación lineal de primer orden

Figura 1.14: Ejemplo de rectificación proyectiva

asociada a la fijación de niveles de profundidad. Por último se esbozan algunas cuestiones de visualización que son de utilidad para aplicaciones en tiempo real a interacción basada en VR/AR o bien en videojuegos.

Reformulación de las ecuaciones de perspectiva

Una vez realizada la *inicialización* es necesario proceder al reescalado de elementos. Para ello, se lleva a cabo un *reescalado de la matriz de rotación* por la componente t_z del vector $\mathbf{t}$ de traslación:

$$R_1 := \mathbf{r}_1^\top / t_z \ , \ R_2 := \mathbf{r}_2^\top / t_z$$

El proceso se *inicializa* en el plano imagen con respecto a $\mathbf{t}$ haciendo

$$(x_0, y_0) \ = \ (\frac{t_x}{t_z}, \frac{t_y}{t_z}) \quad , \quad \varepsilon \ := \ \frac{\mathbf{r}_3^\top \mathbf{P}_i}{t_z}$$

Las *ecuaciones de perspectiva* se escriben

$$x_i = \frac{1}{1+\varepsilon_i}(R_1 \mathbf{P}_i + x_0) \ , \ y_i = \frac{1}{1+\varepsilon_i}(R_2 \mathbf{P}_i + x_0)$$

o bien como

$$x_i(1+\varepsilon_i) - x_0 - x_0 \frac{1}{t_z}(\mathbf{r}_3^\top \mathbf{P}_i) = \frac{1}{t_z}(\mathbf{r}_1^\top \mathbf{P}_i) - x_0 \frac{1}{t_z}(\mathbf{r}_3 \mathbf{P}_i)$$

a partir de las cuales se obtiene

$$\rightarrow \ (x_i - x_0)(1+\varepsilon_i) = \mathbf{I}_p \mathbf{P}_i \quad \text{con } \mathbf{I}_p = \frac{1}{t_x}(\mathbf{r}_1^\top - x_0 \mathbf{3}^\top)$$

Ejercicio.- La longitud focal f permite obtener datos relativos a coordenadas de las direcciones a partir de estimaciones coordenadas píxel (u_{pix}, v_{pix}) y recíprocamente mediante

$$(u_{mm}, v_{mm}) \ = \ (f\frac{x}{z}, f\frac{y}{z})$$

donde

$$(u_{mm} \ , \ v_{mm}) \ = \ (u_{pix} - o_x)s_u \ , \ v_{pix} - o_y)s_v$$

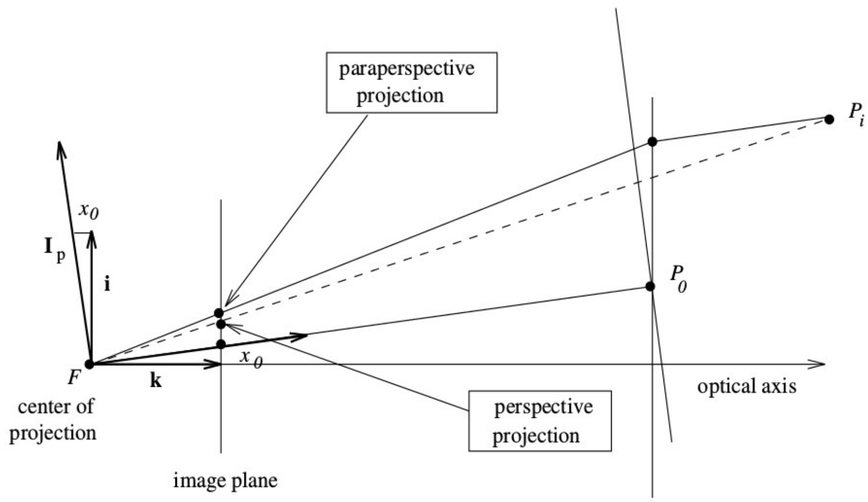

Figura 1.15: Principio de proyección con un modelo de cámara de paraperspectiva

o de forma alternativa

$$(u_{pix}\ ,\ v_{pix})\ =\ (\frac{u_{mm}}{s_u}+o_u\ ,\ \frac{v_{mm}}{s_v}+o_v)$$

Aplicad esta descripción a una imagen en perspectiva frontal capturada por el usuario para recuperar una aproximación a datos métricos sobre la escena.

Paraperspectiva

Existen diferentes aproximaciones para las ecuaciones de perspectiva que pueden ser reformuladas en términos de aproximaciones de orden 0 o de orden 1:

- orden 0: Tomamos $\frac{1}{1+\varepsilon_i} \simeq 1\ \rightarrow$ *perspectiva débil* $\rightarrow$ descartado.
- orden 1: Tomamos $\frac{1}{1+\varepsilon_i} \simeq 1-\varepsilon_i\ \rightarrow$ *paraperspectiva*

La *paraperspectiva* es un modelo lineal afín que proporciona una aproximación de orden 1 a modelos de perspectiva construidos inicialmente según modelos no-lineales (tipo LSM o Least Squares Methods). Para ello, se lleva a cabo una simplificación dada por

$$\frac{1}{1+\varepsilon}\ \simeq\ 1-\varepsilon$$

lo cual permite reescribir las ecuaciones de la paraperspectiva como

$$x_i = \frac{1}{1+\varepsilon_i}(R_1\mathbf{P}_i + x_0)\ \simeq\ (1-\varepsilon_i)(R_1\mathbf{P}_i + x_0)$$

En virtud de la ortogonalidad entre las líneas de la matriz de rotación se tiene

$$x_i = \mathbf{R}_1\mathbf{P}_i + x_0 - x_0\varepsilon_i = \frac{\mathbf{R}_1\mathbf{P}_i}{t_z} - x_0 - x_0\frac{\mathbf{R}_3\mathbf{P}_i}{t_z} = x_i^p$$

Razonando análogamente para y_i podemos reescribir la *ecuación de la paraperspectiva* como:

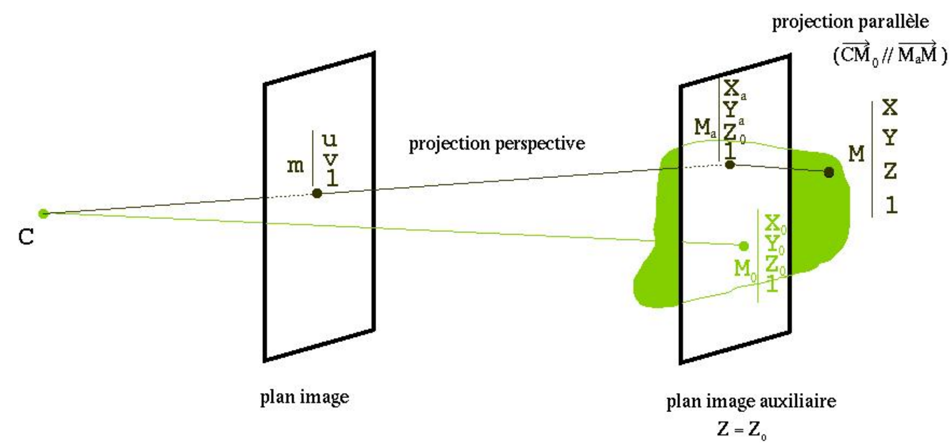

Figura 1.16: Modelo de proyección de paraperspectiva

$$x_i^p - x_i = \frac{\mathbf{r}_1^\top - x_0\mathbf{r}_3^\top}{t_z}\mathbf{P}_i \ , \ y_i^p - y_i = \frac{\mathbf{r}_2^\top - y_0\mathbf{r}_3^\top}{t_z}\mathbf{P}_i$$

La *Reconstrucción basada en Paraperspectiva* según (Horaud, 1996) sigue un *método iterativo* basada en reescribir las ecuaciones de perspectiva como

$$(x_i - x_0)(1+\varepsilon_i) = \mathbf{I}_p\mathbf{P}_i \ , \ y_i - y_0)(1+\varepsilon_i) = \mathbf{J}_p\mathbf{P}_i$$

con $\mathbf{I}_p = \frac{1}{t_x}(\mathbf{r}_1^\top - x_0\mathbf{3}^\top)$. Esta construcción se puede interpretar de dos formas diferentes:

- $(x_i, y_i$ es la proyección perspectiva de $\mathbf{P}_i$ (enfoque clásico)
- $(x_i(1+\varepsilon_i)) - x_0\varepsilon_i, y_i(1+\varepsilon_i)) - y_0\varepsilon_i)$ es la proyección para-perspectiva de $\mathbf{P}_i$

La *idea* para la Reconstrucción 3D basada en paraperspectiva consiste en estimar ε_i de forma incremental a partir de la inicialización dada por la proyección perspectiva.

La Reconstrucción pseudo-euclídea basada en Paraperspectiva sólo requiere un número muy bajo de iteraciones. Por ello es una solución apropiada para visualizar modelos en relación con aplicaciones a entornos inmersivos, videojuegos u otras aplicaciones que requieran respuesta en tiempo real.

Matriz de calibración intrínseca

Con la notación de la subsección anterior, utilizamos ahora las expresiones que relacionan las coordenadas píxel con las coordenadas métricas (medidas en mm) módulo el factor de escala

$$x_{mm} = (x_{pix} - o_x)s_x \quad y_{mm} = (y_{pix} - o_y)s_y$$

que reescribimos como

$$x_{pix} = -\frac{x_{mm}}{s_x} + o_x \quad , \quad y_{pix} = -\frac{y_{mm}}{s_y} + o_y$$

se obtiene la expresión matricial

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix}_{pix} = \begin{pmatrix} -\frac{1}{s_x} & 0 & o_x \\ 0 & -\frac{1}{s_y} & o_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ w \end{pmatrix}_{mm} = \mathbf{K}_{int}\mathbf{p}$$

siendo $\mathbf{K}_{int}$ la *matriz de calibración intrínseca* que no depende de la localización de la cámara sino tan sólo de los parámetros internos de la cámara.

El marco proyectivo proporciona una representación "compacta" de una proyección general como una "doble clase de conjugación" de la proyección canónica; en otras palabras cualquier punto **P** visible desde el centro de la cámara **C** se proyecta sobre un punto **p** de una vista de acuerdo con

$$\mathbf{p} = \lambda \mathbf{K}\mathbf{M}_{\pi}^{0}\mathbf{H}\mathbf{P}$$

donde λ es un factor de proporcionalidad, **K** es una matriz triangular superior de la forma especificada más arriba (que es una homografía particular de $\mathbb{P}^2$), $\mathbf{M}_{\pi}^{0}$ es la matriz de la proyección canónica (para una proyección ortogonal) y **H** es una transformación rígida espacial salvo escala (una homografía particular en el espacio $\mathbb{P}^3$)

Diferentes tipos de visualización

La visualización de escenarios o de objetos volumétricos debe ser compatible con diferentes requerimientos que pueden afectar a aspectos métricos (para preservar distancias y mejorar la gestión de las deformaciones aparentes), afines (para facilitar la navegación) o proyectivos (para enlazar diferentes modelos afines). Existe un gran número de modelos de perspectiva que facilitan la visualización. En este apartado sólo se presentan algunos de los más comunes, englobando el enfoque de los modelos de perspectiva en el marco afín.

En notación afín, se obtiene una proyección sobre el plano con coordenadas $(U,V,W)^{\top}$ siendo $W \neq 0$ que se aplica sobre el plano cartesiano $\frac{U}{W}, \frac{V}{W})^{\top} = (u,v)^{\top}$. Con esta notación, la *proyección ortogonal* esta dada analíticamente por la asignación:

$$[X_0 : X_1 : X_2 : X_3] \mapsto (U,V,W)^{\top} \mapsto (u,v)^{\top}$$

El *primer reto* relacionado con las aplicaciones de este modelo consiste en tratar de mejorar la estimación y la generación de proyecciones. La visualización asociada a dicha generación de proyecciones se realiza mediante modelos de perspectiva más realistas (usando proyecciones más acordes con la percepción visual humana) y facilitar su actualización en tiempo real

Para una *proyección ortogonal general* suponemos que los proyectores son perpendiculares al plano sobre el que se proyecta. Las proyecciones ortogonales presentan algunas *variantes significativas* tales como

- *Proyección axonométrica:* El plano de imagen no es perpendicular a ningún eje coordenado OX_i $\Rightarrow$ se muestran diferentes caras en cada vista.

- *Proyección isométrica:* Es una proyección axonométrica con vector director $(\pm 1, \pm 1, \pm 1)$.

Las proyecciones paralela, ortogonal y las variantes descritas son las más utilizadas en Arquitectura, Ingeniería o Videojuegos en 2,5D. Esta forma de representación presenta algunos *inconvenientes* tales como

- Modelo no-realista en relación con la percepción humana.

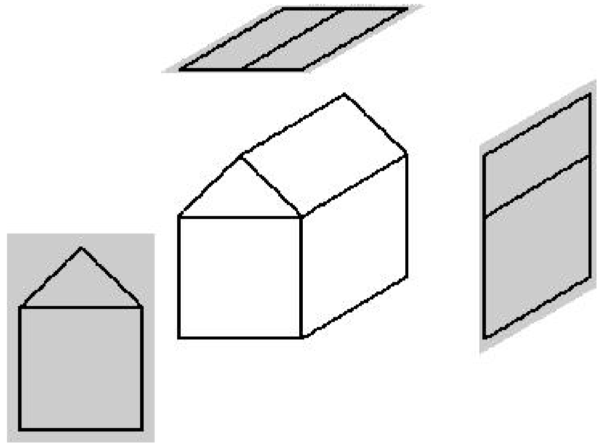

Figura 1.17: Ejemplo de proyección ortogonal


Figura 1.18: Ejemplo de proyección isométrica en arquitectura y videojuegos

- Elevado coste vinculado a la producción manual del modelo.

Rectificación y calibración interna de la cámara

En esta subsección se aborda el *problema siguiente:* Rectificar una imagen recuperando la métrica salvo escala , es decir, módulo el grupo de transformaciones de semejanza. Este problema se resuelve en los pasos siguientes:

1. Rectificación proyectiva de imagen
2. Puntos circulares del infinito
3. Cónica absoluta
4. Versión proyectiva de ángulos entre rectas

5. Cuádrica absoluta y ángulos entre planos

La rectificación de una imagen consiste en transformar haces de líneas de perspectiva en haces de rectas paralelas, enviando cualquier punto de fuga $\forall \mathbf{V}_i$ (situado a distancia finita) a la línea del infinito ℓ_∞ dada por $x_3 = 0$. En el primer apartado se muestra la relación entra la condición de conservar la ortogonalidad y la conservación de una forma cuadrática "en el infinito" a la que se llama "cuádrica absoluta" del espacio proyectivo. Los conceptos de cónica o cuádrica absoluta son más difíciles de visualizar. En ambos casos se trata de elementos imaginarios (que no es posible visualizar) pero que desempeñan un papel crucial para la estimación de propiedades euclídeas del espacio subyacente.

Una propiedad clave que muestra el interés de la cónica absoluta consiste en que permite estimar el ángulo formado entre dos rectas. Por ello, permite evaluar la distorsión asociada a efectos de perspectiva vinculados a una proyección central. Este argumento permite no sólo ortorectificar vistas, sino generar vistas sintéticas simulando rotaciones asociadas a cambios virtuales en la localización de la cámara. Con más generalidad, la conservación de la cónica absoluta ω_∞ por una homografía $\mathbf{H}$ es equivalente a la conservación de la métrica euclídea. Por ello, el grupo de las transformaciones de semejanza (transformaciones rígidas salvo escala) que es crucial para la Reconstrucción Euclídea resulta caracterizado por la condición de conservar la cónica absoluta ω_∞. Los dos últimos apartados de esta subsección están dedicados a presentar algunas propiedades básicas de la cónica y la cuádrica absoluta.

Rectificación de imagen

En este apartado se muestra cómo convertir en paralelas las líneas asociadas a haces proyectivos. La estrategia consiste en usar *Homografías* $\mathbf{H} \in \mathbb{P}GL(4;\mathbb{R})$ con restricciones adicionales:

- *Restricciones sobre la escena:* $\mathbf{V}_i$ para haces de LP en planos ortogonales.
- *Restricciones* sobre el movimiento donde se adopta un *pipeline* típico que consiste en estimar por separado y de forma consecutiva:
  1. Rotaciones puras
  2. Transformaciones lineales [Har94]]
  3. Extensiones de las transformaciones anteriores.

La homografía $\mathbf{H}$ deja fijo el plano del infinito Π_∞. Esta condición equivale a que $\mathbf{H}$ sea una transformación afín. Las *Condiciones de paralelismo* y Π_∞ dado por $X_3 = 0$ se expresan mediante

- $\ell_1 \| \ell_2 \;\Rightarrow\; \ell_1 \cap \ell_2 \in \Pi_\infty$ dado por $x_3 = 0$
- $\Pi_1 \| \Pi_2 \;\Rightarrow\; \Pi_1 \cap \Pi_2 = \ell_{12} \subset \Pi_\infty$ dado por $x_3 = 0$

De cara a la reconstrucción euclídea, estamos interesados en identificar las transformaciones proyectivas que conservan la condición de perpendicularidad (ó con más generalidad la de ángulo entre rectas). Los apartados siguientes proporcionan la respuesta a esta cuestión.

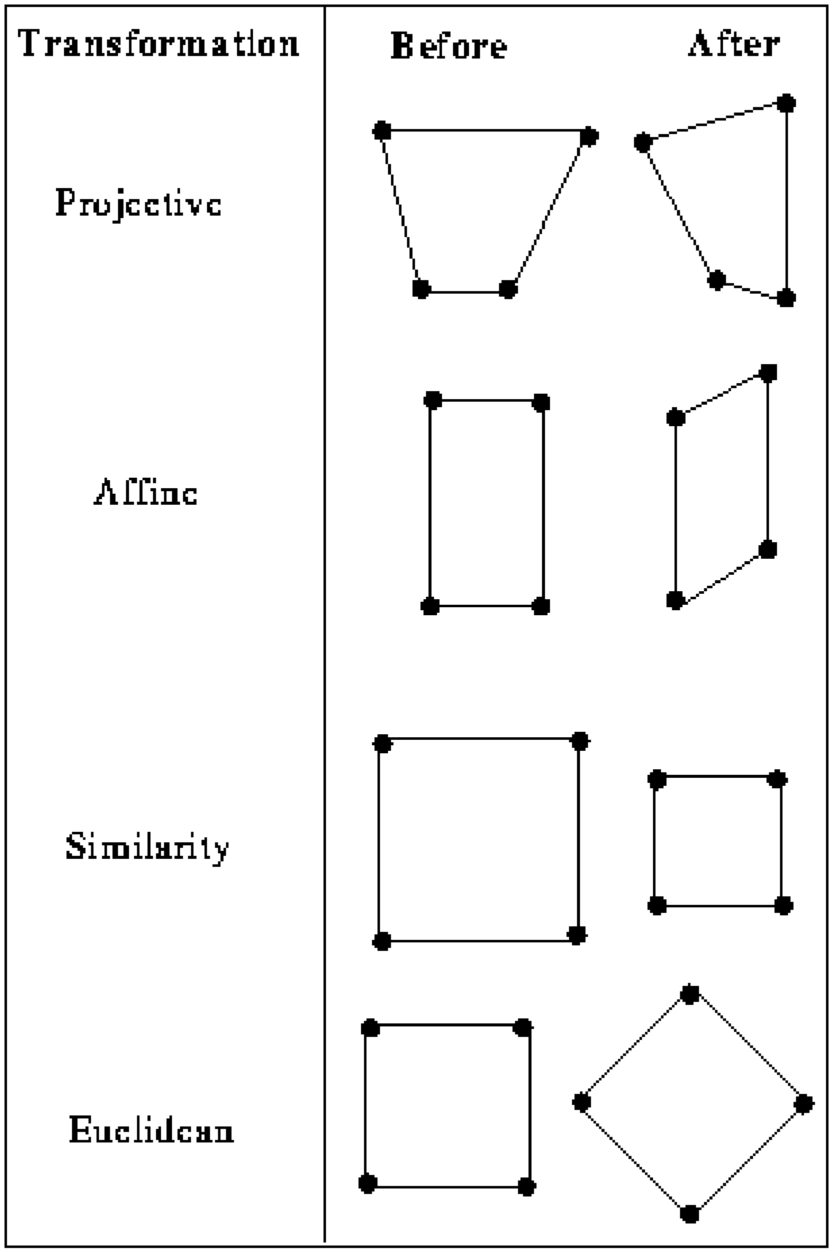

Figura 1.19: Comparación entre diferentes tipos de transformaciones

Puntos circulares del infinito

La primera descripción de la forma cuadrática absoluta Ω_∞ sobre una línea proyectiva se debe a I.Newton quien introdujo los puntos $\{\mathbf{I},\mathbf{J}\}$ (dados ahora en notación proyectiva mediante $[1 : \underline{i} : 0]^\top$) para representar los puntos de la forma cuadrática $x_1^2 + x_2^2 = 0$, $x_3 = 0$. En estos puntos se cortan todas las circunferencias del plano cartesiano que escribimos en notación proyectiva como $x_1^2+x_2^2+dx_1x_3+ex_2x_3+fx_3^2 = 0$; la intersección con $x_3 = 0$ da la cónica absoluta en la recta proyectiva $\mathbb{P}^1$ dada por dos puntos [13].

Ejercicio.- Verifica que los puntos circulares del infinito $[1 : \underline{i} : 0]^\top$ son puntos fijos por una homografía $\mathbf{H}$ (transformación proyectiva o colineación) del plano proyectivo $\mathbb{P}^2$ si y sólo si, dicha colineación es una semejanza. *Indicación:* Comprueba que para una transformación de semejanza $\mathbf{H}_\lambda$ se tiene

$$\mathbf{H}_S\mathbf{I} = \begin{pmatrix} \lambda cos\,\theta & -ssen\,\theta & t_x \\ \lambda sen\,\theta & scos\,\theta & t_y \\ 0 & 0 & 1 \end{pmatrix} = \lambda e^{i\theta} \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} = \mathbf{I}$$

extiende el argumento para $\mathbf{J}$ y verifica el recíproco.

Interpretación: La descomposición obvia $\mathbf{I} = [1 : 0 : 0]^\top + i[0 : 1 : 0]^\top$ (ejes ortogonales para la representación real de la recta compleja) y su análoga para $\mathbf{J}$ (dirección conjugada) muestra que la conservación de los puntos circulares del infinito equivale a la conservación de la ortogonalidad entre dos vectores (y su versión conjugada). Un truco algebraico elemental muestra que la (dual de la) cónica absoluta sobre una recta proyectiva $\mathbb{P}^1$ (correspondiente a un haz de líneas del plano ordinario) se expresa como

$$q_\infty^* = \mathbf{I}\mathbf{J}^\top + \mathbf{J}\mathbf{I}^\top = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Por ello, la conservación de esta cónica por transformaciones de semejanza se reescribe como

$$\mathbf{H}_S q_\infty^* \mathbf{H}_S^\top = q_\infty^*$$

que proporciona una forma sintética de expresar la conservación de la ortogonalidad por transformaciones de semejanza. Esta expresión se toma como definición para dimensión superior que se describe a continuación por separado para la cónica absoluta en $\mathbb{P}^2$ (en relación con la conservación de ángulos entre rectas) y para la cuádrica absoluta en $\mathbb{P}^3$ (en relación con la conservación de ángulos entre planos).

Cónica absoluta

La cónica y la cuádrica absoluta son variedades geométricas contenidas en el hiperplano del infinito cuya conservación caracteriza a las transformaciones euclídeas. Por ello, son cruciales para cuestiones de calibración que se abordan en la sección 3 de este capítulo.

Análogamente, todas las esferas del espacio ordinario se cortan a lo largo de la cónica definida por $X_1^2 + X_2^2 + X_3^2 = 0$ y $X_4 = 0$ [14] a la que se llama cónica absoluta y se representa mediante ω_∞. Este argumento se extiende a dimensión arbitraria de forma elemental.

[13] a mediados del s.XVIII no existía aún una descripción de la línea proyectiva como direcciones del plano

[14] Al homogeneizar $X^2 + Y^2 + Z^2 = r^2$ se obtiene $X_1^2 + X_2^2 + X_3^2 = X_4^2$, cortar por el hiperplano del infinito $X_4 = 0$ y extender al caso general por traslación

La cónica absoluta ω_∞ es invariante por $\mathbf{H} \in \mathbb{P}GL(4;\mathbb{R})$ por lo que $\mathbf{H}$ es una *transformación de semejanza*. La *ecuación de la cónica absoluta* es:

$$X_0^2 + X_1^2 + X_2^2 = 0 \quad , \quad X_3 = 0$$

Tiene las *Propiedades de incidencia* siguientes:

- $\#\{C \cap \omega_{infty}\} = 2$ $\forall$ circunferencia C
- $\forall$ esfera $\mathbb{S}^2$ se tiene $\mathbb{S}^2 \cap \Pi_\infty = \omega_\infty$

Nótese que bastan tres puntos para determinar una circunferencia; en este caso, la circunferencia en el plano del infinito $X_4 = 0$ está determinada por los dos puntos circulares $\mathbf{I}, \mathbf{J}$ y por un tercer punto a estimar. La "deformación aparente" de esta circunferencia imaginaria en una cónica imaginaria que proporciona idealmente una estimación de la localización relativa de la cámara en relación con el objeto.

Este argumento matemático requiere especificar un método para estimar dicha cónica imaginaria en términos de propiedades fáciles de evaluar como el ángulo entre rectas que se cortan. Para ello, se adopta una estrategia indirecta que consiste en estimar

1. Estimar la cuádrica absoluta dual Ω_∞ (DAC) [15].
2. Proyectar la cuádrica absoluta dual sobre el plano Π_i de imagen dada como $\omega_\infty^i \sim \mathbf{M}_i\Omega_\infty\mathbf{M}_i^T$ donde $\mathbf{M}_i$ es la matriz de proyección asociada a la proyección central $\pi_i : \mathbb{P}^3 \to \Pi_i$ sobre el plano de imagen $\Pi_i \simeq \mathbb{P}^2$
3. Calcular la relación entre la cónica absoluta ω_∞ (invariante del espacio proyectivo) y la imagen ω_∞^i de la cónica absoluta sobre el plano de imagen Π_i

Para obtener la *orientación relativa* es necesario estimar *ángulos entre rectas*. En el caso más sencillo, interesa conservar la ortogonalidad. La estimación depende del marco geométrico. El *ángulo euclídeo* entre líneas se define como

$$cos\theta = \frac{d_1^\top d_2}{\sqrt{(d_1^\top d_1)(d_2^\top d_2)}}$$

Con esta notación, la *ortogonalidad euclídea* se expresa como

$$d_1^\top d_2 \ = \ 0$$

La descripción del ángulo proyectivo entre líneas es más sofisticada y se aborda en el apartado siguiente. Esta presentación se justifica de cara a cuestiones de auto-calibración. En este caso, la "deformación aparente" de una circunferencia ideal (cónica absoluta= proporciona información sobre la localización relativa de la cámara; interesa disponer de una versión intrínseca de dicha localización en términos de ángulos de rectas conocidas en términos de un dato intrínseco (la cónica absoluta); una primera aproximación al problema se presenta en el apartado siguiente (saltar en primera lectura).

[15]La dual de una cuádrica regular se define como la envolvente de sus hiperplanos tangentes; en el caso 3D, los planos tangentes que no pasan por el centro de proyección se proyectan sobre líneas tangentes a la proyección sobre el plano de imagen que es la cónica absoluta ω_∞

Versión proyectiva de ángulos

La matriz de proyección $\mathbf{M}_i$ asociada a la proyección central $\pi_i : \mathbb{P}^3 \rightarrow \Pi_i$ sobre el plano de imagen $\Pi_i \simeq \mathbb{P}^2$ se representa en el marco euclídeo como

$$\mathbf{M}_i = \mathbf{K}\mathbf{R}_i^T [I_3 \mid -\mathbf{t}_i]$$

donde $\mathbf{K}$ es la matriz triangular superior correspondiente a los parámetros intrínsecos, $\mathbf{R}_i^T$ es la rotación (cambio en la orientación) y $\mathbf{t}_i$ es la traslación que conecta dos posiciones "consecutivas" de cámara. La cuádrica absoluta dual Ω_∞ tiene como matriz $diag(1,1,1,0)$. Por tanto, sustituyendo en $\omega_\infty^i \sim \mathbf{M}_i \Omega_\infty \mathbf{M}_i^T$ se obtiene

$$\omega_\infty^i \sim \mathbf{K}\mathbf{K}^T$$

Por ello, se obtiene el resultado siguiente [Har04]:

Proposición.- La cónica absoluta ω_∞ es equivalente a la inversa de $\mathbf{K}\mathbf{K}^T$ donde $\mathbf{K}^\top$ es la traspuesta de la matriz de calibración interna $\mathbf{K}$ de la cámara.

Actualización de la información métrica.- Para cada representación proyectiva de la escena, se tiene una forma particular de la dual de la cuádrica absoluta $\Omega_\infty^(T)$ que es conjugada de la forma canónica Ω_∞ [16]. En otras palabras:

$$\Omega_\infty^(T) = \mathbf{T}\Omega_\infty \mathbf{T}^T$$

donde $\mathbf{T}$ a la transformación que lleva una transformación métrica en una proyectiva. Como cada vista se ha obtenido en un marco euclídeo, la condición de tangencia es proyectivamente invariante y se conserva por proyección sobre el plano de imagen de cada vista, se tiene que la imagen $\omega_\infty^(i)$ de la DAC sobre cada plano de imagen sigue verificando la ecuación estructural $\omega_\infty^i \sim \mathbf{K}\mathbf{K}^T$, lo cual permite actualizar la información proyectiva a métrica.

En particular, la homografía $\mathbf{H}_{ij}^\infty$ que lleva un plano de imagen Π_i sobre otro plano de imagen Π_j transforma la imagen $\omega_i \subset \Pi_i$ de la DAC en la imagen $\omega_j \subset \Pi_j$ de modo que

$$\omega_j \sim (\mathbf{H}_{ij}^\infty)^{-T} \omega_i (\mathbf{H}_{ij}^\infty)^T$$

que en términos duales se reescribe como

$$\omega_j^\star \sim \mathbf{H}_{ij}^\infty \omega_i^\star (\mathbf{H}_{ij}^\infty)^T$$

Actualización de la información epipolar (Kruppa).- Como la condición de tangencia es proyectivamente invariante y rectas epipolares se transforman en rectas epipolares, se tiene que las rectas epipolares que son tangentes a ω_i en la visa i-ésima, deben seguir siendo tangentes a ω_j en la vista j-ésima. Esta condición se expresa algebraicamente en términos de las *ecuaciones de Kruppa*:

$$[e_{ij}]_\times^\top \mathbf{K}\mathbf{K}^\top [e_{ij}]_\times \sim \mathbf{F}_{ij}\mathbf{K}\mathbf{K}^\top \mathbf{F}_{ij}^\top$$

donde $\mathbf{F}_{ij}$ es la matriz fundamental para las vistas i y j, siendo e_{ij} el epipolo correspondiente. En este caso sólo se obtienen dos ecuaciones independientes en lugar de las 5 requeridas para determinar

[16] http://www.cs.unc.edu/ marc/tutorial/node87.html

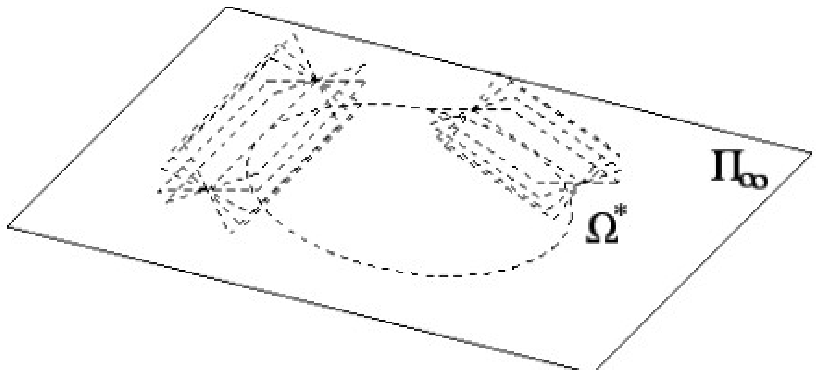

Figura 1.20: Representación de la cónica absoluta [Pollefeys]

una cónica [Zel96]. Esta reducción de las restricciones de auto-calibración a la geometría epipolar equivale a eliminar la estimación en la posición del infinito en las ecuaciones presentadas más arriba. Este enfoque da lugar a la aparición de degeneraciones artificiales que se pueden abordar an términos de la geometría de las cónicas y cuádricas duales [17]

Una interpretación alternativa consiste en interpretar las ecuaciones de Kruppa como la proyección sobre cada plano de imagen de una cuártica alabeada que es la intersección de las dos cuádricas (lugar base del haz) que aparecen en la descripción anterior. Esta cuártica es una curva elíptica que al proyectarse sobre un plano adquiere dos nodos. El problema a resolver consiste en estimar dicha curva de grado 4 sobre cada plano de imagen y "elevarla" a una curva alabeada en $\mathbb{P}^3$; este problema aún no ha sido resuelto.

Una nota sobre ángulos.- Denotemos mediante $\mathbf{d}_1$ y $\mathbf{d}_2$ dos direcciones del espacio ordinario que pasan por el centro de proyección y se proyectan sobre dos puntos $\mathbf{x}_1$ y $\mathbf{x}_2$ del plano de imagen. El ángulo entre las direcciones $\mathbf{d}_1$ y $\mathbf{d}_2$ se calcular mediante la conocida expresión

$$cos\theta \ := \ \frac{\mathbf{d}_1 \hat{A} \mathbf{d}_2}{|\,\mathbf{d}_1\,|\,|\,\mathbf{d}_1} \,| = \ \frac{\mathbf{d}_1^\top \mathbf{d}_2}{(\mathbf{d}_1^\top \mathbf{d}_1)^{1/2}.(\mathbf{d}_2^\top \mathbf{d}_2)^{1/2}}$$

Como $\mathbf{x}_i = \mathbf{K}\mathbf{d}_i$ para $i = 1, 2$, reemplazando el valor de $\mathbf{d}_i$ se obtiene la expresión intrínseca del ángulo entre rectas asociada a la matriz de calibración:

$$cos\theta \ = \ \frac{\mathbf{x}_1^\top (K^{-\top} K^{-1}) \mathbf{x}_2}{(\mathbf{x}_1^\top (K^{-\top} K^{-1}) \mathbf{x}_1)^{1/2}.(\mathbf{x}_2^\top (K^{-\top} K^{-1}) \mathbf{x}_2)^{1/2}} \ = \ \frac{\mathbf{x}_1^\top \omega_\infty \mathbf{x}_2}{(\mathbf{x}_1^\top \omega_\infty \mathbf{x}_1)^{1/2}.(\mathbf{x}_2^\top \omega_\infty \mathbf{x}_2)^{1/2}}$$

donde $\omega_\infty = (KK^\top)^{-1}$ representa la cónica absoluta lo cual permite reformular la *Conjugación proyectiva* como

$$d_1^\top \omega_\infty d_2 = 0$$

Una nota sobre la cuádrica absoluta y ángulos entre planos: La *Cuádrica absoluta* es una forma cuadrática no-degenerada imaginaria en $\mathbb{P}^3$ dada en coordenadas duales por

$$Y_0^2 + Y_1^2 + Y_2^2 + Y_3^2 = 0$$

[17] Ver la tesis de JF para los aspectos geométricos del problema

Esta cuádrica *degenera* en la cuádrica absoluta Ω_∞ que es dual de la cónica absoluta ω_∞

$$Y_0^2 + Y_1^2 + Y_2^2 = 0 \quad , \quad Y_3 = 0$$

Esta cónica está contenida en el plano del infinito Π_∞ dado por $Y_3 = 0$ que es el vector nulo de Ω_∞. La cuádrica absoluta *dual* permite calcular *ángulos entre planos*:

$$cos\theta = \frac{\Pi_1^\top \Omega_\infty \Pi_2}{\sqrt{(\Pi_1^\top \Omega_\infty \Pi_1)(\Pi_2^\top \Omega_\infty \Pi_2}}$$

Para más detalles ver [Har04]

Homografías y calibración

La 3×4-matriz de proyección $\mathbf{M}_\pi$ se estima por el método de coeficientes indeterminados a partir de una colección "suficiente" de imágenes $\mathbf{p}_i$ de puntos conocidos $\mathbf{P}_i$. resolviendo las condiciones $\mathbf{p}_i = \mathbf{M}_\pi \mathbf{P}_i$ salvo factor de proporcionalidad. A continuación es necesario incorporar restricciones adicionales para interpretar las soluciones en términos de la calibración de la cámara.

Para una localización fija, no hay transformación proyectiva en el espacio tridimensional ambiente y la matriz de proyección M_π se escribe de forma más simplificada como $M_\pi = K(\mathbf{R} \mid \mathbf{t}$ donde K es la matriz triangular superior asociada a la calibración intrínseca, $\mathbf{R}$ es la matriz de rotación y $\mathbf{t}$ es el vector traslación. Por ello, el *segundo paso* consiste en estimar K, ($\mathbf{R}$ y $\mathbf{t}$ a partir de la matriz de proyección $\mathbf{M}_\pi$

Estimación de homografías de un plano proyectivo

Cualquier proyección que no sea paralela al eje óptico da lugar a distorsiones; en particular, una parrilla plana de calibración se proyecta sobre un cuadrilátero idealmente plano. La aproximación lineal a estas distorsiones se puede evaluar en términos de la modificación asociada una malla regular dada por cuadrados que se visualiza como una malla cuadrangular. Por tanto, los elementos básicos a estimar son cuadrados (vista rectificada) y cuadriláteros (deformación lineal asociada a una vista en perspectiva arbitraria). La transformación más general que lleva un cuadrado en un cuadrilátero es una transformación proyectiva u homografía. Para facilitar la extensión del método propuesto se extiende el método a toda la imagen que se considera como un plano proyectivo.

La estimación de una homografía como transformación regular del plano proyectivo es sencilla: La homografía $\mathbf{H} \in \mathbb{P}GL(3;\mathbb{R})$ transforma un plano dominante en su homólogo en el espacio de llegada. Haciendo $x := \frac{x_1}{x_0}$, $y := \frac{x_2}{x_0}$, $x' := \frac{y_1}{y_0}$, $y' := \frac{y_2}{y_0}$ se tiene:

$$\begin{array}{rcl} x'h_{00} + xx'h_{01} + yx'h_{02} - h_{10} - h_{11}x - h_{12}y & = & 0 \\ y'h_{00} + xy'h_{01} + yy'h_{02} - h_{20} - h_{21}x - h_{22}y & = & 0 \end{array}$$

que se reescribe en forma matricial como

$$\begin{pmatrix} x' & xx' & yx' & -1 & -x & -y & 0 & 0 & 0 \\ y' & xy' & yy' & 0 & 0 & 0 & -1 & -x & -y \end{pmatrix} (h_{00} : \ldots : h_{22})^\top \; = \; \mathbf{O}$$

Haciendo $\mathbf{q}_i = \mathbf{H}\mathbf{p}_i$ se obtienen dos ecuaciones linealmente independientes. Un cómputo de parámetros muestra que se necesitan 4 pares de puntos homólogos $(\mathbf{p}_i, \mathbf{q}_i)$ en posición general para estimar $\mathbf{H}$

Los *problemas a resolver* en relación con una *única vista* son los siguientes:

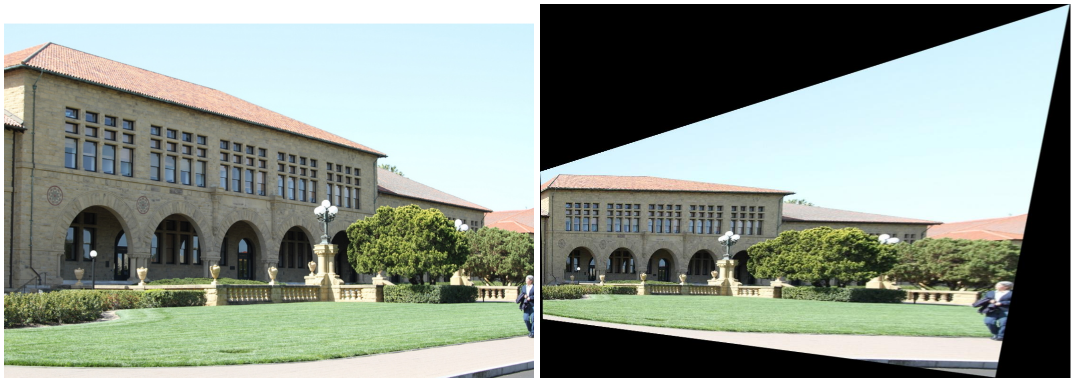

Figura 1.21: Ejemplo de imagen rectificada a partir de la extracción de cuadriláteros

- Automatizar la extracción de cuadriláteros (4-uplas de ”vértices” en posición general)
- Rectificar el cuadrilátero de referencia mediante una homografía.
- Fuera del cuadrado o del cuadrilátero de referencia (incluido en el plano dominante) puede haber distorsiones enormes.
- Las 4-uplas de referencia para un cuadrilátero cambian con la localización relativa de la cámara. En presencia de movimientos arbitrarios de cámara hay que controlar las degeneraciones de las 4-uplas de puntos.

Para ello, es necesario desarrollar una aproximación que resuelva de forma simultánea la puesta en correspondencia de elementos contenidos diferentes planos. La solución para un único plano dominante es muy simple (basta una homografía del plano de imagen basada en cuadriláteros, p.e.); esta solución justifica el tratamiento de fondos de la escena como si fueran texturas.

Sin embargo, la solución simultánea para diferentes planos de forma simultánea sin realizar un modelado 3D completo de la escena, requiere introducir transformaciones afines del espacio con un mayor número de puntos de control. La calibración es necesaria para identificar parámetros y corregir distorsiones que puedan facilitar el proceso de pegado o de simulación de movimientos de cámara asociados a una discretización del movimiento (ver capítulo siguiente).

Estimación de homografías para la calibración

En el último apartado de la subsección anterior se ha mostrado cómo la proyección está caracterizada por

$$\mathbf{p} = \lambda \mathbf{K}\mathbf{M}_{\pi}^{0}\mathbf{H}\mathbf{P}$$

donde **K** y **H** son homografías del plano y del espacio proyectivo, respectivamente (sometidas a restricciones adicionales). Por ello, la estimación de la matriz de proyección equivale a estimar dos homografías.

Para gestionar la información de una cantidad finita (habitualmente redundante) de puntos que aparecen en la imagen, adaptamos una notación matricial $\mathbf{U} = \mathbf{H}\mathbf{X}$ para la homografía **H** del plano proyectivo que transforma una n-tupla $\mathbf{X} = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$ de puntos del plano en una n-tupla $\mathbf{U} = (\mathbf{q}_1, \ldots, \mathbf{q}_n)$ de puntos del plano y que expresamos analíticamente como

$$[(x_1,y_1,1)^T,\ldots,(x_n,y_n,1)^T] \quad \mapsto \quad [(u_1,v_1,1)^T,\ldots,(u_n,v_n,1)^T]$$

Con esta notación, para calcular **H** tal que $\mathbf{U} = \mathbf{hX}$, es necesario "invertir" **X** (calculando una pseudo-inversa $\mathbf{X}'$ de la n-tupla original). En otras palabras, la homografía se obtendría como

$$\mathbf{H} = \mathbf{UX}'(\mathbf{XX}')^{-1}$$

Desafortunadamente, las cosas no son tan "fáciles" pues la ecuación que deseamos resolver no es $\mathbf{U} = \mathbf{hX}$, sino $\lambda_i \mathbf{q}_i = \mathbf{Hp}_i$ para $\lambda \neq 0$ y para $i = 1,\ldots,n$.

Por ello, es necesario realizar una adaptación del método de descomposición en valores singulares (SVD). Este método proporciona una *factorización*

$$\mathbf{A} = \mathbf{UDV}^T$$

de la matriz $\mathbf{A} \in M(m \times n; \mathbb{R})$ correspondiente a la homografía en producto de tres matrices donde

- $\mathbf{U} \in GL(m;\mathbb{C})$ y $\mathbf{V} \in GL(n;\mathbb{C})$ son matrices unitarias, es decir, las líneas (filas o columnas) son vectores unitarios dos a dos ortogonales (con respecto a la métrica compleja), pero no necesariamente rotaciones.
- la matriz **D** es diagonal
- Los valores propios son cuadrados de los elementos de **D** y los vectores propios son columnas de **V** [18]
- Los vectores propios con valores propios nulos son las únicas soluciones no-triviales.
- Las columnas de la matriz **V** de la SVD proporcionan una base ortogonal para el dominio de **A**. En particular, los valores singulares nulos tienen como vectores propios asociados los vectores de la base para el espacio nulo (núcleo de la aplicación que corresponde a las soluciones buscadas)

Por consiguiente, en este caso particular, $rk(\mathbf{A}) = n-1$ y $\mathbf{x}$ es un vector solución definido salvo escala.

Estimación directa de la matriz de proyección

La matriz de proyección $\mathbf{M}_\pi$ se puede estimar directamente a partir de las correspondencias entre $\mathbf{p}_i$ y $\mathbf{P}_i$ con $\mathbf{M}_\pi(\mathbf{P}_i) = \mathbf{p}_i$. De la misma forma que antes, esta estimación se realiza mediante el método de coeficientes indeterminados. Fijadas referencias en el abierto afín $D^+(x_3)$ de $\mathbb{P}^2$ y el abierto afín $D^+(X_3)$ de $\mathbb{P}^3$, escribimos la proyección en términos de coordenadas afines como

$$\begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

lo cual da dos ecuaciones

$$x_i = \frac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}}\ ,\ y_i = \frac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}}$$

[18] Recordemos que la ecuación $\mathbf{Ax} = 0$ admite soluciones no-nulas si y sólo si **A** es singular; de ahí el nombre de SVD

que imponen dos restricciones lineales sobre los coeficientes de la matriz:

$$m_{11}X+m_{12}Y+m_{13}Z+m_{14}-m_{31}x_iX-m_{32}x_iY-m_{33}x_iZ-m_{34}x_i = 0 m_{21}X+m_{22}Y+m_{23}Z+m_{24}-m_{31}y_iX-m_{32}y_iY-m_{33}y_iZ-m_{34}y_i$$

en el subespacio afín del producto de Segre $\mathbb{P}^2 \times \mathbb{P}^3 \hookrightarrow \mathbb{P}^{11}$ que parametriza los pares $(\mathbf{p}_i, \mathbf{P}_i)$ verificando que $\mathbf{p}_i = \mathbf{M}_\pi \mathbf{P}_i$. Expresamos matricialmente dichas restricciones mediante

$$\begin{pmatrix} X & Y & Z & 1 & 0 & 0 & 0 & 0 & -x_iX & -x_iY & -x_iZ & -x_i \\ 0 & 0 & 0 & 0 & X & Y & Z & 1 & -y_iX & -y_iY & -y_iZ & -y_i \end{pmatrix} \mathbf{m} = \mathbf{O}$$

donde $\mathbf{m} = (m_{11}, m_{12}, m_{13}, m_{14}, m_{21}, m_{22}, m_{23}, m_{24}, m_{31}, m_{32}, m_{33}, m_{34})^\top$ es el vector que representa la matriz de proyección salvo factor de proporcionalidad. Como cada par de puntos homólogos da lugar a 2 ecuaciones se necesitan (al menos) 6 pares de puntos homólogos $(\mathbf{p}_i, \mathbf{P}_i)$ verificando que $\mathbf{p}_i = \mathbf{M}_\pi \mathbf{P}_i$ para estimar la matriz de proyección $\mathbf{M} = (m_{ij})_{1\leq i\leq 3, 1\leq j\leq 4}$.

En general se tiene una colección redundante de n pares de puntos homólogos $(\mathbf{p}_i, \mathbf{P}_i)$ lo cual da lugar a una matriz $\mathbf{A}$ de tamaño $2n \times 12$ para $n \geq 6$. En la práctica, no es posible calcular en un tiempo razonable menores de tamaño 12×12 que permitan; por ello, es necesario dar los pasos siguientes:

1. *Linealizar el problema:* Minimizar $\mid \mathbf{Am} \mid$ bajo la restricción $\mid \mathbf{m} \mid = 1$ que se obtiene a partir del vector propio con el mínimo valor propio de la matriz simétrica $\mathbf{AA}^\top$ [19]

2. Utilizar la solución lineal para $\mathbf{A}$ identificada en el punto precedente para *inicializar el proceso de minimización no-lineal* asociado a la diferencia entre el punto medido y el punto proyectado:

$$min_{\mathbf{M}} \sum_{i=1}^{n} [(x_i, y_i) - \pi_{\mathbf{C}}(X_i, Y_i, Z_i, 1)]^2$$

En presencia de outliers el segundo apartado puede dar lugar a errores de bulto que ralenticen la convergencia en el proceso de estimación de la matriz. Por ello, es conveniente adoptar estrategias más finas de muestreo tipo *RanSaC* (Random Sampling Consensus) de amplio uso en casi todas las cuestiones de Reconstrucción. Este procedimiento parte de una colección minimal de puntos y trata de verificar la "bondad" del ajuste para otros pares de puntos candidatos a homólogos.. Consiste en los pasos siguientes:

1. *Identificar* una colección de N pares de puntos candidatos a homólogos $(\mathbf{p}_i, \mathbf{P}_i)$ (típicamente se lleva a cabo a partir de esquinas o máximos de intensidad)

2. *Seleccionar* una 6-upla de puntos "en posición general" dentro de la colección de N pares de puntos (habitualmente $N >> n \geq 6$)

3. *Estimar* la matriz de proyección $\mathbf{M}_\pi$ mediante la estrategia precedente (linealización y minimización no-lineal asociada a la linealización)

4. *Propagar* la solución obtenida aplicando la matriz de proyección estimada a otros candidatos validando el resultado:
   - Si el porcentaje de "éxito" es superior al 60% admitir la estimación como válida
   - En caso contrario seleccionar otra 6-upla y reiniciar el proceso

5. Terminar

[19] Esta formulación equivale a resolver el problema SVD (Singular Value Decomposition) correspondiente al vector con mínimo valor singular de $\mathbf{A}$

1.2. Calibración Fotogramétrica

Históricamente, la *calibración de una cámara* formaba parte de la Fotogrametría; por ello en este capítulo se califica como *Calibración Fotogramétrica* a la que se dedica la sección 2. El *objetivo de la Fotogrametría* es la extracción de información métrica mediante el uso de una parrilla de calibración; esta extracción se realizaba habitualmente de forma manual. La Visión Computacional proporciona modelos matemáticos y herramientas software para automatizar el proceso de extracción de datos a partir de varias imágenes usando la rigidez del objeto fotografiado sin un conocimiento previo de las coordenadas "absolutas" de los puntos de referencia $\mathbf{P}_i$.

La estimación de estas medidas (invariantes en el marco euclídeo) proporciona el soporte para llevar a cabo la calibración de la cámara que puede llevarse a cabo off-line (mayor precisión) o bien on-line (cuando se requiera rapidez para la fusión de información "al vuelo"). Los modelos físico-matemáticos permiten identificar procedimientos robustos para le estimación de los fenómenos de distorsión y su corrección semi-automática. Para realizar una estimación robusta de parámetros es necesario restringir el tipo de movimiento realizado por la cámara; la primera solución correspondiente a una rotación de una cámara fue desarrollada por Maybank y Faugeras (1992); un algoritmo más sencillo de implementar fue desarrollado por Hartley (1997). El seguimiento manual de personajes en escena (actores en una representación o jugadores de un encuentro, p.e.) requiere una extensión de estos métodos para evitar problemas con la visualización y facilitar la fusión de información en presencia de varias cámaras (esto último se aborda en el módulo 5 del CEViC).

La *estrategia general* para la calibración "absoluta" consiste en dos pasos;

1. Identificar una cantidad "suficiente" de puntos significativos $\mathbf{p}_i$ en imagen que son proyección de puntos de referencia $\mathbf{P}_i$ en la escena
2. Estimar la matriz de proyección M_π usando las coordenadas píxel (u_i, v_i) de $\mathbf{p}_i$.
3. Estimar los parámetros intrínsecos y extrínsecos de la cámara usando la matriz de proyección.

En Fotogrametría terrestre se conocen coordenadas absolutas 3D (mediante geoposicinamiento, p.e.) de puntos $\mathbf{P}_i$ en la escena. En caso contrario, se utiliza una cantidad redundante de puntos en imagen, eliminando los datos relativos a $\mathbf{P}_i$ en la primera fase.

Los *elementos significativos* pueden proceder de diferentes dispositivos o marcas (incluyendo dianas de calibración) o de un objeto arbitrario (pero fijo) en la escena. La primera aproximación se enmarca dentro de la *Calibración Fotogramétrica* y utiliza diferentes objetos 2D o 3D de calibración a los que se llama *parrillas de calibración*. La incorporación de métodos basados en Geometría Proyectiva ha dado lugar a calibración basada en líneas o bien a calibración basada en puntos de fuga para direcciones ortogonales que aparecen en la escena. La segunda aproximación utiliza sólo las condiciones de rigidez relativas al objeto y recibe el nombre de *auto-calibración*

La estimación de *parámetros intrínsecos* (características internas de la óptica a partir de los píxeles de la vista) y de los *parámetros extrínsecos* relativos a la localización (posición y orientación) de la cámara permite identificar *relaciones entre las coordenadas* de imagen C_I, de la cámara C_C y del mundo C_W. En el modelo lineal asociado a una cámara tipo pinhole, la relación entre datos del mundo y la imagen se lleva a cabo mediante una 3×4-matriz de proyección $\mathbf{M}_\pi$ que relaciona el sistema de coordenadas mundo C_W con el sistema de coordenadas de la cámara $C_\mathbf{C}$ utilizando propiedades del sistema de coordenadas de la imagen C_I.

En realidad, este modelo es una simplificación abusiva, pues no tiene en cuenta las características de las lentes ni tampoco, por consiguiente, las distorsiones (radiales o tangenciales) generadas por la captura de información a través de la lente. Por ello, para procesos en los que se requiera elevada precisión métrica es necesario estimar las distorsiones generadas por las cámaras utilizadas.

Existen *diferentes criterios* para la calibración

1. Geométrica (parámetros, proyección) vs Radiométrica (color, iluminación)
2. Corrección de perspectivas vs distorsiones/aberraciones
3. Fotogramétrica (dianas de calibración) vs Autocalibración
4. Sistema no-lineal (minimal) vs lineal (redundante)

La combinación de estos criterios da lugar a diferentes *taxonomías* para calibración, que se ha traducido en una gran diversidad de métodos y algoritmos

La calibración fotogramétrica se lleva a cabo off-linea usando parrillas de calibración. Loe elementos a determinar para el modelo lineal tipo pinhole afectan a elementos internos y externos de la cámara. Los parámetros intrínsecos no dependen de la localización de la cámara y afectan a elementos internos; son la longitud focal, punto principal y distorsión de los píxeles a lo largo de direcciones coordenadas afines (no necesariamente ortogonales). Los parámetros extrínsecos afectan a la localización; son 6 y están dados por la posición del centro de la cámara y la orientación del eje óptico.

Metodología para fotogrametría terrestre

La *calibración fotogramétrica* obtiene los parámetros de la cámara a partir de la estimación de datos de un objeto 3D (parrilla de calibración) con medidas geométricas previamente conocidas.

1. *Aspectos metodológicos*
2. *Estimación de parámetros*
3. *Actualización de la información*
4. *Problemas adicionales*

Aspectos metodológicos

En este apartado se adopta un *enfoque fotogramétrico* que incluye aspectos relativos a las localizaciones sucesivas de la cámara y el diseño de una red fotogramétrica para alcanzar una cobertura lo más completa posible del objeto

Los *Aspectos relativos a la cámara* son cruciales para simplificar la estimación de parámetros.

- *Orientación interior* cuyo objetivo es estimar el punto principal, longitud focal y distorsión de la lente (5 parámetros intrínsecos)
- *Orientación exterior* cuyo objetivo es estimar la localización (posición y orientación) de la cámara con respecto a sistema coordenado absoluto (6 parámetros extrínsecos)

Se recomienda alternar vistas frontales con vistas angulares (evitar vistas oblicuas muy sesgadas), y no utilizar ningún tipo de zoom o de angular que genere distorsiones adicionales.

El diseño de una *Red Fotogramétrica* afecta al geo-posicionamiento de las localizaciones de la cámara y a la inserción de puntos de control (llamados *dianas* de calibración) sobre el objeto o la escena para facilitar la puesta en correspondencia de las imágenes obtenidas. Ambos son cruciales para la captura geo-referenciada de una colección redundante de elementos para la calibración y posterior “pegado” de datos homólogos de cara a obtener un objeto global.

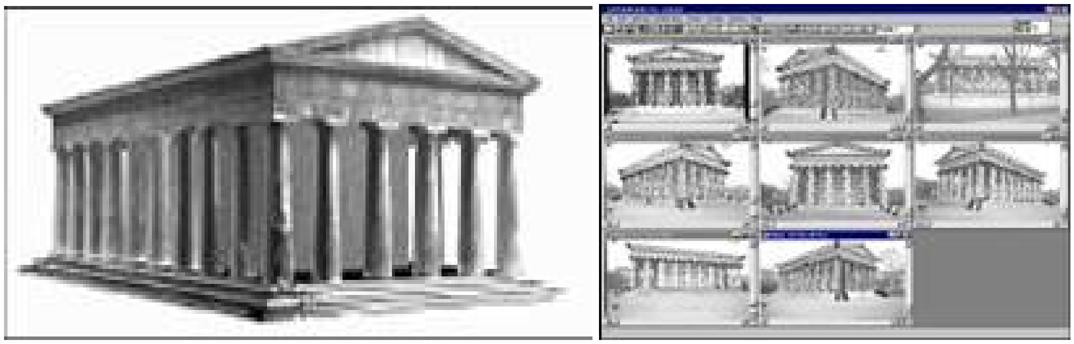

Figura 1.22: Ilustrando las 3x3 reglas clásicas de la fotogrametría

- *Orientación absoluta:* Identificar transformación entre sistemas de coordenadas a partir de dianas.
- *Orientación relativa:* Posición y orientación relativa entre cámaras a partir de proyecciones de dianas.

Una especificación más detallada de metodologías de Fotogrametría Terrestre relacionada con estas cuestiones aparece citada en la literatura como las 3×3 rules que se pueden consultar en 3×3: http://cipa.icomos.org

Estrategia para la estimación de parámetros

La estimación de parámetros es crucial para reproyectar la información procedente de diferentes dispositivos. Esta cuestión afecta no sólo a cámaras convencionales, sino también a cámaras de vídeo y dispositivos activos para evaluar la profundidad como infrarrojos o diferentes tipos de láser $3D$.

En el marco proyectivo, la estimación de parámetros afecta a las homografías de $\mathbb{P}^2$ y $\mathbb{P}^3$ que actúan a la izquierda y a la derecha sobre la matriz de proyección canónica; la estimación por "fuerza bruta" de las homografías requiere un elevado número de pares de puntos de control y puede proporcionar resultados que no se ajustan a lo esperado generando distorsiones adicionales. Por ello, es importante incorporar restricciones adicionales que afectan a una interpretación de las homografías desde el punto de vista de las cámaras.

En el marco de la calibración de cámaras, las matrices que representan las homografías están dadas por una matriz triangular (parámetros intrínsecos) y una transformación rígida salvo escala (parámetros extrínsecos). Las estrategias habituales proceden por desacoplamiento (separación entre transformaciones que actúan a la izquierda y a la derecha) y utilizan inicialmente el método de coeficientes indeterminados relativo a una k-tupla de pares de puntos homólogos (por la homografía o por la proyección) para estimar los coeficientes de la transformación afectada.

La estimación obtenida siguiendo esta metodología presenta problemas en relación con la re-proyección de los datos calculados, por lo que es necesario minimizar el error de re-proyección. Para evitar problemas relacionados con la falta de visibilidad o la falta de "posición general" para los pares de puntos homólogos, se suele tomar una cantidad redundante de pares de puntos homólogos; debido a la incertidumbre y errores en la estimación, el sistema es habitualmente incompatible, por lo que es necesario procedimientos de optimización para resolverlo, minimizando una función de coste. Las estrategias de optimización tienen asimismo dos fases:

1. Una primera fase *lineal* (basada en SVD) que proporciona una solución inicial tosca basada en diferentes tipos de distancia (euclídea frecuentemente, pero también L^1 u otras más rápidas y con mejor tratamiento para outliers)

2. una segunda fase *no-lineal* que se inicializa con la anterior (mínimos cuadrados, habitualmente) y que trata de controlar la distribución global del error.

Dependiendo del objeto algebraico a estimar (matrices de las homografías o de la proyección) existen diferentes tipos de error en le proceso de calibración o en el de la Reconstrucción $3D$. Los más frecuentes aparecen asociados a

- el error en una imagen (cuando se cuenta con una plantilla de calibración o se desea estimar algún tipo de distorsión);
- el error de transferencia simétrico (cuando se desea comparar los resultados correspondientes a dos imágenes asociado a la puesta en correspondencia automática);
- el error de reproyección (en la fase de Reconstrucción Euclídea propiamente dicha)

Actualización de la información

Es una cuestión crucial en aplicaciones que requieren interacción para estimar localización relativa y corrección de parámetros que puedan afectar a la re-localización de dispositivos. Las aplicaciones más avanzadas conciernen a la robótica asistida por visión con problemas de gran complejidad como la coordinación ojo-mano para facilitar la interacción en entornos complejos. El caso más difícil corresponde al diseño e implementación de estrategias compatibles con una realimentación en tiempo real como la que se requiere en aplicaciones industriales de *visual servoing* u otras más críticas relacionadas con la asistencia a cirugía basada en Visión.

Una *cuestión previa* a resolver que afecta a la calibración consiste en *minimizar el error de alineamiento*. Para simular el comportamiento de la visión humana en relación con este problema es conveniente utilizar 2 vistas para objetos volumétricos.

La actualización de la información requiere especificar las funciones de coste vinculadas al proceso de estimación y la minimización de diferentes tipos de errores. Este problema tiene una vertiente estadística que es necesario aplicar para reforzar la convergencia de la información. Para una única imagen los problemas iniciales a resolver son la corrección del ruido (que puede afectar al “alineamiento de primitivas”) y la corrección de las distorsiones.

La *corrección del ruido* implica contar con

- Un *modelo de ruido:* En el módulo 1 se han presentado diferentes modelos de ruido. Para simplificar y fijar ideas, a la vista del carácter estático que predomina en la captura de información de la mayor parte de este módulo, aquí nos restringimos a un modelo de ruido que suponemos Gaussiano isótropo y de media cero, es decir, con función de densidad

  $$Pr(\mathbf{x}) = \frac{1}{2\pi\sigma^2} e^{-(\mathbf{x}-\overline{\mathbf{x}})/(2\sigma^2)}$$

  donde $\overline{\mathbf{x}}$ denota el valor medio.

- Una *función de coste óptima* del modelo de ruido obtenida maximizando la verosimilitud asociada a la transformación realizada.

La primera hipótesis es abusiva y deberá ser reemplazada por modelos de ruido más realistas, sobre todo en presencia de movimiento. La segunda hipótesis depende del tipo de transformación a realizar (homografías vs proyecciones, p.e.). Para fijar ideas, en el caso de una homografía $\mathbf{H}$ un modelo más refinado para *error en una imagen* en torno a un número finito de puntos de control $\mathbf{x}_i$ para $1 \leq i \leq N$ está dado por

$$Pr(\{\mathbf{x}_i\} \mid \mathbf{H}) \;=\; \Pi_{i=1}^{N} \frac{1}{2\pi\sigma^2} e^{-d(\mathbf{x}_i, \mathbf{H}\overline{\mathbf{x}}_i)^2/(2\sigma^2)}$$

donde suponemos que la varianza σ^2 es la misma para un pequeño entorno de cada punto de control. Tomando logaritmos se obtiene una expresión

$$log(Pr(\{\mathbf{x}_i\} \mid \mathbf{H})) \;=\; -\frac{1}{2\sigma^2} \sum_{i=1}^{N} d(\mathbf{x}_i, \mathbf{H}\overline{\mathbf{x}}_i)^2 + \; cte$$

Bajo estas hipótesis simplificadoras, la *función de coste a optimizar* está dada por la *máxima verosimilitud* que escribimos como

$$\sum_{i=1}^{N} d(\mathbf{x}_i, \mathbf{H}\overline{\mathbf{x}}_i)^2$$

Obviamente, la función cuadrado de la distancia euclídea se puede reemplazar por otras funciones distancia más apropiadas para cada caso. Los modelos correspondientes a dos o más imágenes se abordan en la lección 3 de este módulo.

Nota.- En términos geométricos más abstractos, el paso de un grupo de Lie G a su álgebra de lie $\mathfrak{g} := T_eG$ (espacio tangente al grupo en el elemento neutro $e \in G$) se realiza mediante diferenciación a lo largo de caminos que pasan por $e \in G$. Este paso representa la linealización de la estructura de grupo. La aplicación exponencial (que es un difeomorfismo local) permite recuperar la estructural local del grupo G en un entorno de $e \in G$ y transforma la adición de matrices $X_1 + X_2 + \ldots$ (representando elementos de $\mathfrak{g}$) en producto de matrices $exp(X_1 + X_2 + \ldots) = exp(X_1)exp(X_2)\ldots$. Por ello, el procedimiento anterior de máxima verosimilitud reproduce en términos de modelos de optimización el mismo paso que se realiza del grupo al álgebra (y recíprocamente).

Como la función (cuadrado de la) distancia está asociada a los procesos de optimización tipo LSM (Least Squares Method), esta simple observación explica la eficiencia de la realimentación entre los procedimientos basados en mínimos cuadrados (para el espacio original) y los de máxima verosimilitud (para el espacio linealizado). Si ahora recordamos que el proceso de optimización presentado en relación con la estimación de las matrices de proyección tenía una fase lineal y otra no-lineal, esta observación muestra una realimentación no sólo entre procesos de optimización sino entre las simetrías locales asociadas a una versión infinitesimal asociada a las dos fases de la optimización.

Problemas adicionales de Calibración en Fotogrametría

Problema: Convertir en paralelas las líneas asociadas a haces proyectivos: LP $\rightarrow$ Líneas epipolares

1. *Diseño apropiado* para red topográfica $\rightarrow$ Reglas 3×3 de la Fotogrametría

2. *Corrección* de la distorsión relativa a la curvatura de la Tierra para imágenes aéreas/satelitares

3. *Corrección* de medidas tomadas bajo condiciones ambientales desfavorables $\rightarrow$ Procesamiento global de imagen

Figura 1.23: Detección de puntos de fuga mediante Ransac y el estimador Danés

4. *Reconstrucción* para ternas de imágenes tomadas en vuelo → Captura de la volumetría

Calibración fotogramétrica
Utiliza objetos con geometría conocida para la calibración

1. Parrilla 3D
2. Plantilla 2D
3. Objeto lineal

La *idea inicial* consiste en realizar una estimación pro mínimos cuadrados (LSM) que consiste en minimizar (la suma de) los cuadrados de las diferencias entre los datos predichos y los observados. Esta estrategia da lugar a distorsiones indeseadas debidas a outliers que ralentizan la convergencia de los métodos utilizados o bien dan lugar a modelos inestables.

Un análisis independiente realizado en colaboración con Diego G.Aguilera y M.Gonzalo muestra que el estimador que presenta mejor comportamiento para garantizar la convergencia es el estimador Danés basado en procedimientos Ransac.

Algunas recomendaciones adicionales para la calibración fotogramétrica

Para la calibración fotogramétrica relativa a los aspectos geométricos (de la cámara y la escena) es imprescindible utilizar una parrilla de calibración. La parrilla puede ser plana, formada por un ángulo diedral o incluso 3D. En Visión Dinámica Estéreo sería deseable contar con una parrilla transparente que se pudiera superponer a la escena. Para no interferir con la operabilidad en la escena, sería conveniente que dicha parrilla 3D se diseñara en un espectro no-visible. Este tipo de soluciones aún no están comercializadas y por ello, aquí sólo se expone una solución convencional al alcance de cualquier Laboratorio.

Parrilla 2D

Se basa en la estimación de datos de un patrón planar que se muestra con diferentes orientaciones. La clave inicial radica en una extracción robusta de datos (esquinas y segmentos) de los datos contenidos en la parrilla, la estimación de la deformación aparente relativa a dichos datos en relación con el modelo

Una ventaja es que no requiere conocer a priori el movimiento realizado por la cámara (localización relativa) en relación con el objeto capturado. Este método presenta el diseño experimental más sencillo, por lo que es el método off-line más popular

Un *pipeline típico* consiste en los pasos siguientes:

1. Estimar las deformaciones en plantilla
2. Extraer los parámetros intrínsecos de la cámara
3. Extraer los Parámetros extrínsecos de la cámara

El *problema más difícil* es la estimación de la distorsión en regiones centrales y periféricas que se aborda en términos de la distorsión radial y tangencial. [20]

Línea base de calibración

El cambio en la localización de la cámara con respecto a una posición inicial ideal se puede describir en términos geométricos mediante una traslación y una rotación. La línea que soporta el vector de traslación recibe el nombre de *línea base* y a la distancia entre los dos centros de la cámara se le denota mediante b.

La *línea base de calibración* es una técnica más reciente que las anteriores que se basa en utilizar un desplazamiento controlado de cámara. El esquema siguiente presenta los aspectos más relevantes de esta estrategia [21]

1. Utiliza un conjunto de $N \geq 4$ puntos con distancia conocida sobre una recta ℓ_{cal}
2. La calibración se realiza observando el desplazamiento de una línea en torno a un punto fijo (similar a zoom).
3. Puede combinar una rotación $\mathbf{R}_i$ en la cámara $\mathbf{C}_i \rightarrow$ Estimación lineal
4. Es un método apropiado para calibración simultánea de múltiples cámaras dispuestas en red $\rightarrow$ Retransmisiones deportivas para TV3D

Ya se han realizado transmisiones de eventos deportivos mediante vídeo 3D en relación con partidos de fútbol en la UE, baloncesto en la NBA ó bien pruebas de gimnasia en las Olimpiadas de Londres, p.e. En todos los casos, la existencia de marcas muy contrastadas en la escena permiten estimar tanto las transformaciones lineales (homografías) entre vistas capturadas por diferentes cámaras, como la deformación métrica (para esta última los círculos marcados en el terreno de juego son especialmente útiles).

Asimismo las cámaras están situadas en un plano elevado con respecto al terreno de juego; por ello, los modelos de perspectiva angular (y sus aproximaciones lineales) son los más apropiados para inicializar el proceso. Aunque el enfoque pueda ser variable, la distancia entre las cámaras (línea base) y el terreno de juego se mantienen en un rango casi-constante. Un *reto* de gran interés para el próximo futuro consiste en integrar la información proporcionada por una cámara en movimiento

[20] Una versión "intrínseca" más elaborada utiliza la estimación de curvas de grado 4 (Kruppa) que corresponden a las tangentes que pueden trazarse a dos cónicas. Esta versión se comenta en el Apéndice 2 de esta Lectura.

[21] *Nota:* No confundir con la línea base $b = d(\mathbf{C}_1, \mathbf{C}_2)$ para 2 cámaras en Visión Estéreo que se desarrolla en el capítulo siguiente

en los bordes del terreno de juego (modelo de perspectiva frontal con planos de profundidad cortos) con la información 3D proporcionada por cámaras alejadas del terreno de juego [22].

Relacionando las proyecciones euclídea y proyectiva

La utilización de ángulos presentada en la subsección anterior es muy sensible a error y en presencia de direcciones verticales puede ser altamente inestable. Por ello, para una aproximación automática es necesario desarrollar una versión vectorial que pueda resolver el problema de la autocalibración on-line. En esta subsección se revisan algunos de los argumentos presentados anteriormente pero en un contexto vectorial más sencillo de utilizar e implementar

Versión euclídea

Eligiendo de forma apropiada las coordenadas proyectivas suponemos que la oblicuidad es nula por lo que la proyección asociada a la primera cámara está dada inicialmente por

$$\mathbf{M}_{\mathbf{C}_1} = \begin{pmatrix} \alpha & 0 & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{pmatrix}$$

Si la ratio de aspecto es uno ("píxeles cuadrados") $\alpha = \beta$; "normalizando" coeficientes, podemos reescribir la matriz de proyección euclídea correspondiente a la primera cámara como

$$\mathbf{P}_{E1} \; = \; \mathbf{K}_1[\mathbf{I}_3 \,|\, \mathbf{O}]$$

donde $\mathbf{K}_1 = diag(f_1, f_1, 1)$. De forma análoga la matriz de proyección euclídea correspondiente a la segunda cámara $\mathbf{C}_2$ se escribe como

$$\mathbf{P}_{E2} \; = \; \mathbf{K}_2[\mathbf{R}_2 \,|\, \mathbf{t}_2]$$

donde $\mathbf{K}_2 = diag(f_2, f_2, 1)$ y el par $(\mathbf{R}_2, \mathbf{t}_2)$ proporcionan la localización (rotación, traslación) de la segunda cámara en relación con la primera.

Versión proyectiva

La matriz de proyección para el caso proyectivo es más general que la del caso euclídeo, pues está dada por una expresión de la forma $\mathbf{M}_\pi = [\mathbf{M} \,|\, \mathbf{m}]$ donde $\mathbf{M}$ es una 3×3-matriz regular y $\mathbf{m}$ es un 3×1-vector. La conversión de la matriz de proyección $[\mathbf{M} \,|\, \mathbf{m}]$ asociada $\mathbb{P}^3 \to \mathbb{P}^2$ en una matriz de proyección $\mathbf{K}_i[\mathbf{R}_i \,|\, \mathbf{t}_i]$ para la i-ésima cámara requiere identificar una homografía representada por una 4×4-matriz regular $\mathbf{H}$ tal que

$$\mathbf{P}_{E2} \; \simeq \; \mathbf{M}_\pi \mathbf{H} \quad \textbf{para } i = 1, 2$$

De la primera condición $\mathbf{P}_{E1} = \mathbf{K}_1[I \,|\, \mathbf{0}] \; \simeq \; \mathbf{M}_\pi \mathbf{H}$, se obtiene que la 3×3-caja superior izquierda de $\mathbf{H}$ es $\mathbf{K}_1$ y la columna a la derecha es idénticamente nula, por lo que la 4×4-matriz de la homografía $\mathbf{H}$ que induce la transformación entre ambas proyecciones (proyectiva y euclídea) es de la forma

[22] Estas cuestiones se abordan en el módulo 5 del CEViC y en la asignatura sobre Video 3D del máster de Ingeniería Informática

$$\mathbf{H} = \begin{pmatrix} \mathbf{K}_1 & \mathbf{O} \\ \mathbf{v}^\top & 1 \end{pmatrix}$$

Identificando la homografía

Eligiendo coordenadas proyectivas apropiadas podemos suponer que el plano π_∞ del infinito está dado por $X_4 = 0$, es decir, $\mathbf{H}^\top \pi_\infty = [0001]^\top$ por lo que

$$\pi_\infty = \mathbf{H}^\top \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{K}^{-T} & -\mathbf{K}^{-T}\mathbf{v} \\ \mathbf{0}^\top & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\mathbf{K}^{-T}\mathbf{v} \\ 1 \end{pmatrix}$$

En particular si $[\mathbf{n}_\infty : 1]^\top$ representa el "vector normal" del plano del infinito π_∞, la matriz $\mathbf{H}$ de la homografía presentada más arriba se puede rescribir como

$$\mathbf{H} = \begin{pmatrix} \mathbf{K}_1 & \mathbf{O} \\ \mathbf{n}_\infty^\top \mathbf{K}_1 & 1 \end{pmatrix}$$

Parámetros intrínsecos y extrínsecos

En esta subsección se describen los parámetros asociados a características internas y la localización de la cámara. Se supone inicialmente que la cámara es tipo pinhole con proyección central a través del foco o centro de la cámara.

Los *parámetros intrínsecos* son 5 parámetros que representamos mediante una *matriz intrínseca* $\mathbf{K}_\alpha$

- *Longitud focal* f: distancia entre el foco y el plano de cámara
- *Punto principal:* Coordenadas (u, v) de la proyección del foco de la cámara $\mathbf{C}$ sobre el plano imagen Π_α
- *Distorsión de la lente* para un píxel cuadrado con respecto a ejes principales.

Los *Parámetros extrínsecos* son 6 parámetros que representamos en una *matriz extrínseca* $[\mathbf{R}_\alpha \mid \mathbf{t}_\alpha]$ dada por una 3×3-matriz de *rotación* $\mathbf{R}_\alpha \in SO(3; \mathbb{R})$ y una 3×1-matriz de *traslación* $\mathbf{t}_\alpha \in \mathbb{R}^3$

Los parámetros extrínsecos relativos a la rotación corresponden a la *orientación externa de la cámara*, es decir, a los ángulos que describen la "posición" del eje óptico con respecto al sistema coordenados de referencia cuyo origen es el foco o centro $\mathbf{C}$ de la cámara [23]

Caso más simple: $(u_{mm}, v_{mm}) = (f\frac{x}{z}, f\frac{y}{z}) = ((u_{pix} - o_x)s_u, (v_{pix} - o_y)s_v) \to (u_{pix}, v_{pix})$

Calibración: Cámara proyectiva

Expresión de la proyección en términos de la *matriz de calibración*:

1. Para una *proyección centrada en el punto principal* el punto imagen está dado por

$$\mathbf{p} = \mathbf{K}[I \mid 0]\mathbf{P}$$

donde K es la *matriz de calibración* dada en el caso más simple por $\mathbf{K} := \begin{pmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{pmatrix}$

[23] Estos ángulos reciben el nombre de yaw, pitch y roll en la literatura anglosajona

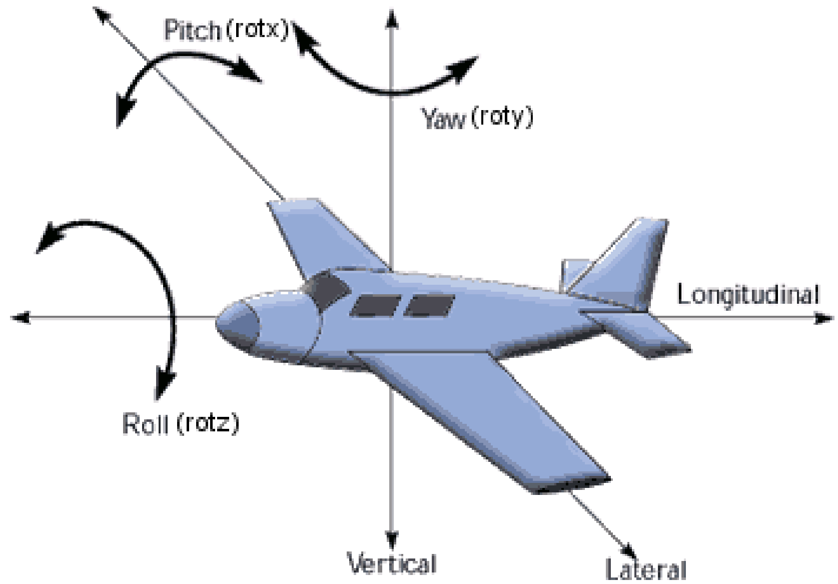

Figura 1.24: Parámetros de calibración extrínsecos de una cámara

2. Aplicando una traslación y una rotación R se tiene:

$$\mathbf{X}_{cam} = \begin{pmatrix} R & -R\mathbf{C} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Modelo para estimar la calibración

1. En el caso general, la *matriz de calibración* está dada por

$$K = \begin{pmatrix} \alpha_x & s & x_0 \\ o & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

   - s el *factor de oblicuidad* en la imagen
   - $(\alpha_x, \alpha_y) = (f m_x, f m_y)$ el número de píxels en cada dirección para corregir píxeles rectangulares

2. La *estimación de la matriz general de calibración* permite calcular la imagen de cada punto $3D$ en el plano de la cámara mediante

$$\mathbf{x} = K.R[I \mid -\mathbf{C}]\mathbf{X}$$

Recuperación de la estructura afín

Una *cámara afín* genera una "deformación aparente" para un rectángulo en un cuadrilátero que se corrige identificando los puntos del infinito correspondientes a lados paralelos

1. *Input:* Cuadrilátero $< \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4 > \rightarrow \ \ell_i = \mathbf{p}_i \times \mathbf{p}_{i+1}$

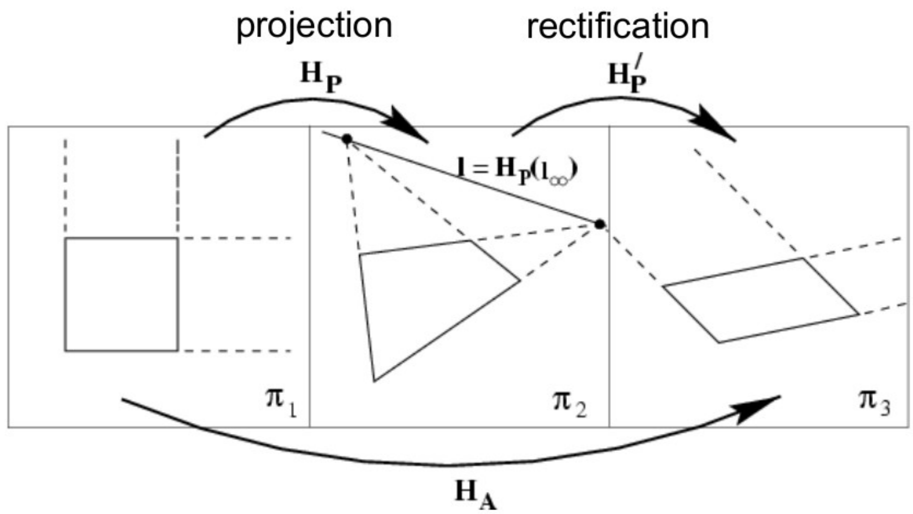

Figura 1.25: Ilustración del método para recuperar la estructura afín [Pollefeys]

2. *Puntos de fuga:* $\mathbf{v}_1 = \ell_1 \times \ell_3$ y $\mathbf{v}_2 = \ell_2 \times \ell_4 \rightarrow \; \ell_\infty = \mathbf{v}_1 \times \mathbf{v}_2$

3. *Rectificación afín:* dada por homografía $\mathbf{H}$ que transforma ℓ_∞ en línea a ”distancia infinita”

4. *Rappel sobre afinidades* (homografías que conservan ℓ_∞

$$\ell'_\infty = \mathbf{H}_A^{-T}\ell_\infty\ell_\infty = \begin{pmatrix} \mathbf{A}^{-T} & 0 \\ -\mathbf{At} & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \ell_\infty$$

5. *Ejemplo:* Si el cuadrado inicial tiene lados paralelos a ejes cartesianos rectangulares, entonces $\mathbf{H}_{cuad} = \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \ell_0 & \ell_1 & \ell_3 \end{matrix}$) con $\ell_\infty = (\ell_0, \ell_1, \ell_3)^\top$. En general, $\mathbf{H} = \mathbf{H}_{cuad}\mathbf{H}_{afin}$

Acoplamiento entre parámetros

El *acoplamiento* con parámetros intrínsecos se describe en los términos siguientes:

$$\mathbf{p} = \frac{1}{z}[||mathbf{K} \mid \mathbf{O}]\mathbf{P} = \frac{1}{z}\mathbf{K}(\mathbf{R}_W^C \mid \mathbf{t}_W^C)\mathbf{P} = \frac{1}{z}\mathbf{MP}$$

siendo $\mathbf{M} = \mathbf{K}(\mathbf{R}_W^C \mid \mathbf{t}_W^C)$ la 3×4-matriz de proyección.

Esta solución global permite “pegar” datos procedentes de cámaras con diferentes localizaciones, pero no es satisfactoria desde el punto de vista métrico. La re-interpretación de los parámetros extrínsecos desde el punto de vista proyectivo se realiza en términos de las transformaciones más elementales, es decir, de las colineaciones.

Recuperación de la estructura métrica

La recuperación de la métrica en la escena es similar a la presentada para la imagen, reemplazando la cónica absoluta por la cuádrica absoluta. Requiere una corrección de las distorsiones geométricas asociadas a la matriz de calibración intrínseca K que se realiza de acuerdo con la metodología presentada más arriba.

El elemento que se añade en este caso es la estimación de la transformación en el espacio ambiente en relación con forma general de la proyección. Esta cuestión afecta a las transformaciones que se realizan sobre el espacio y a la realimentación entre aspectos proyectivo y métricos a partir de la estimación de parámetros. En términos de invariantes, la inclusión de la geometría euclídea en la Proyectiva se realiza seleccionando las transformaciones proyectivas que dejan invariante la cuádrica absoluta.

Síntesis básica de imagen mediante interpolación

Aproximación ingenua a la fusión de imágenes para TV3D: Cámaras fijas idénticas en r localizaciones (no zoom):

1. *Inputs*:
   - *Mapas de color* $c(u_k, v_k)$ y profundidad $d(u_k, v_k)$ en cada píxel $p_k = (u_k, v_k)$
   - *Parámetros extrínsecos* para cada cámara: $\mathbf{K}^{(e)} = (\mathbf{R}_j, \mathbf{t}_j)$ para $2 \leq j \leq r$
   - *Matriz de proyección* $\mathbf{M}_j = \mathbf{K}_j^{(i)}[\mathbf{R}_j \mid \mathbf{t}_j]$
2. Interpolación simultánea entre valores ”inicial” 0 y ”final” 1 para cada par de cámaras
   - *Lineal* para el *Vector traslación:* $\mathbf{t}_\lambda = (1-_\lambda)\mathbf{t}_0 + \lambda\mathbf{t}_1$
   - *Esférica* (sobre $SO(2;\mathbb{R})$ para rotaciones: $\mathbf{R}_\lambda = (1-_\lambda)\mathbf{R}_0 + \lambda\mathbf{R}_1$
   - *Pseudo-lineal* para calibración externa: $\mathbf{K}_\lambda^{(e)} = (1-_\lambda)\mathbf{K}_0^{(e)} + \lambda\mathbf{K}_1^{(e)}$
3. *Unificación* mediante cuaterniones $\rightarrow$ robustez
4. *Reproyección* del color interpolado atendiendo a $c_\lambda(u_k, v_k)$ y profundidad $d_\lambda(u_k, v_k)$ para los píxeles $p_k = (u_k, v_k)$ cuyos homólogos estén bien identificados en $\neq$ imágenes (programación dinámica) y propagación muy rápida entre datos robustos verificados..

Movimiento controlado. Algoritmo de Tsai

El *algoritmo de Tsai* estima los parámetros de una cámara a partir de información contenida en objetos planares (como las parrillas de calibración) y no-planares (asociados a objetos rígidos). Para ello, desarrolla una estrategia en dos pasos que combina técnicas lineales y no-lineales. En el método propuesto por Tsai para evitar problemas numéricos sólo se reconocen las distorsiones radiales de tipo cuadrático (prescindiendo de las distorsiones tangenciales)

La *idea intuitiva* consiste en usar una traslación en Laboratorio, adaptando el procedimiento de calibración utilizando la línea base que ha sido descrito más arriba. La mayor *ventaja* de esta aproximación es la rapidez en la convergencia del procedimiento pues no depende de búsquedas no-lineales a diferencia de otros métodos utilizados en la Calibración Fotogramétrica “clásicas”

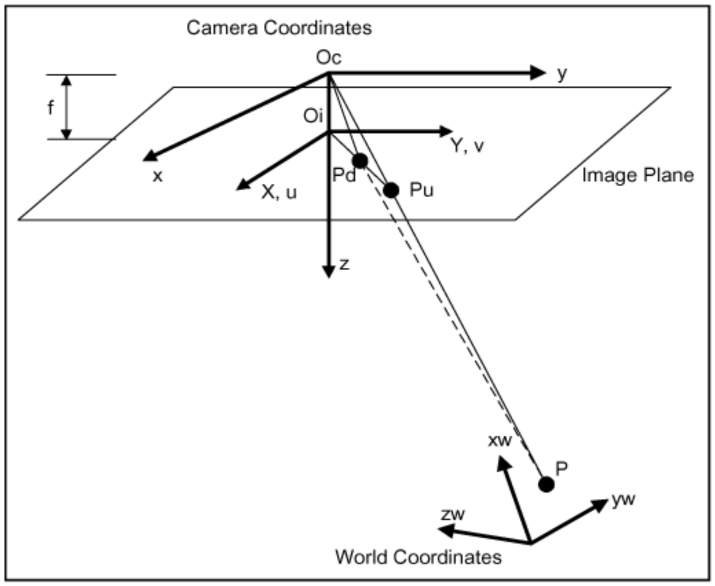

Figure 1. Pin-hole model and the Radial Alignment Constraint hypothesis.

Figura 1.26: Hipótesis sobre restricciones de alineamiento radial aplicadas sobre un modelo de cámara pinhole

1.3. Auto-calibración

La *autocalibración* consiste en la calibración a partir de vistas tomadas por cámaras no-calibradas sin uso de dianas/parrillas o red fotogramétrica, es decir utilizando solamente información sobre una escena o un objeto rígido procedente de las proyecciones. La *auto-calibración* es un área en la que se maximiza el solapamiento entre la Fotogrametría y la Visión Computacional; los métodos proyectivos de *ajuste de haces* utilizan objetos $1D$ (en lugar de vértices de una parrilla) y permiten incorporar información relativa a puntos de fuga para direcciones ortogonales.

El ajuste de haces es un tópico de la Geometría Proyectiva que proporciona soporte a la Fotogrametría y a la Reconstrucción 3D; de una manera intuitiva, el ajuste de haces minimiza el error de reproyección correspondiente a calcular las intersecciones de las rectas homólogas (llamadas epipolar a partir del capítulo 3) contenidas en diferentes vistas. Estas rectas deberían cortarse en un único punto; al no hacerlo, es necesario minimizar las distancias entre ellas [24]. Este enfoque extiende algunos principios básicos de la Geometría Epipolar que se presentan en el capítulo 3 (reconstrucción a partir de dos o más vistas). Dos *cuestiones importantes* afectan a

- *Modelado:* La *auto-calibración* consiste en la estimación de la métrica euclídea que es un invariante de un subgrupo H del grupo proyectivo. Dicho subgrupo está caracterizado por la condición de dejar invariante la cuádrica Q_∞ del hiperplano del infinito H_∞. Por ello, basta con mostrar un procedimiento para estimar dicha cuádrica; el resultado clave es la expresión de Q_∞ en términos de la matriz de calibración $\mathbf{K}$ y de su traspuesta.

[24]La selección de un punto medio sobre el segmento perpendicular común no es una buena estrategia, pues esta condición no es invariante desde el punto de vista proyectivo; es preferible utilizar triangulaciones, pero esta elección conduce a resolver polinomios de grado 6, tarea que tiene un elevado coste computacional; una discusión de este tópico se presenta en el apéndice al capítulo 3

- *Estimación:* En el caso más sencillo, la estimación de la cónica absoluta $x_0^2 + x_1^2 + x_2^2 = 0$ sólo requiere un punto, pues los puntos $\{I, J\} = [1 : \pm i : 0]$ pertenecen a esta cónica del plano $x_3 = 0$. Sin embargo, para objetos con una geometría complicada es necesario estimar distancias entre un gran número de puntos. lo cual afecta a distribuciones de probabilidad definidas sobre un objeto.

En todo este módulo se supone que las imágenes son digitales y que están sometidas a distorsiones mucho mayores que las procedentes de las cámaras métricas utilizadas en Fotogrametría profesional. La recuperación de la información métrica se realiza en Fotogrametría utilizando plantillas cuyas medidas son conocidas a priori; las distorsiones observadas en la cuadrícula proporcionan la clave para una calibración fotogramétrica previa; esta operación se puede llevar a cabo en el Laboratorio antes de realizar la toma de datos en campo o bien realizarse a posteriori para corregir los datos obtenidos. Las *distorsiones geométricas* más comunes producidas por la óptica de la cámara son de tipo radial o tangencial y se expresan en términos de series infinitas con elementos no-lineales con elementos significativos hasta grado 4; para simplificar y acotar los problemas numéricos, frecuentemente se consideran sólo los elementos de orden ≤ 2; el algoritmo de Tsai está basado en uno de los métodos más simples y prescinde de las distorsiones tangenciales y sólo considera las radiales de orden cuadrático. En presencia de movimiento o de cara a la calibración simultánea de múltiples cámaras resulta más apropiado utilizar el algoritmo de Zhang.

En el *caso no-calibrado* inicialmente (calibración en campo, fuera del Laboratorio) se dispone de imágenes tomadas por cámaras cuyas características se desconocen, los dispositivos de captura han podido sufrir modificaciones (traslado bajo condiciones inapropiadas, alunizaje) o bien proceden de diferentes dispositivos con calibración desconocida. La única información disponible está dada por una colección de puntos $3D$con coordenadas capturas mediante algún tipo de dispositivos (desde los teodolitos hasta dispositivos láser 3D). En todos estos casos, se requiere una calibración que permita fusionar la información procedente de una cámara móvil o diferentes dispositivos de captura recurriendo a la "verdad del objeto"; en términos geométricos, dicha condición se traduce en la conservación de distancias/ángulos entre puntos/líneas de las componentes de cada objeto rígido $B^{(\alpha)}$.

El desarrollo de este enfoque requiere plantear el problema en el *marco proyectivo* y suponer que, al menos, se cuenta con una reconstrucción proyectiva aproximada que conlleva información sobre los "elementos del infinito", es decir, un hiperplano del infinito H_∞ (complementario del espacio afín ambiente) y una forma cuadrática no-degenerada contenida en dicho plano que es invariante por transformaciones rígidas. En particular, el elemento geométrico clave para la Reconstrucción euclídea es la cónica absoluta q_∞ (en el plano proyectivo de imagen) o la cuádrica absoluta Q_∞ (en la comparación del espacio ambiente asociada a añadir los puntos del infinito). La imagen $\pi(q_\infty)$ de la cónica absoluta q_∞ "codifica" la calibración interna $\mathbf{K}$ de la cámara.

A lo largo de los años noventa se desarrollaron diferentes métodos para la autocalibración a partir de vistas tomadas por una única cámara; entre las diferentes aproximaciones cabe destacar:

- *Invariancia de la condición de tangencia:* Los planos que pasan por cada centro de la cámara y son tangentes a la cónica absoluta deben ser tangentes a las imágenes de la cónica absoluta. Las ecuaciones de Kruppa proporcionan la restricción adicional para la estimación de la autocalibración [Fau92a], [Fau92a]

- *Simplificación de la matriz fundamental* basada en en SVD y ecuaciones de Kruppa [Har92]

- *Utilización de la cuádrica absoluta Q_∞* como envolvente de planos tangentes a la cónica absoluta que se proyecta sobre la la dual (envolvente por planos tangentes) de la imagen de la cónica absoluta q_∞ [Tri97]

- *Restricción del módulo:* Si los parámetros intrínsecos son constantes, entonces la homografía infinita $\mathbf{H}_\infty$ es conjugada de una matriz de rotación $A \in SO(3;\mathbb{R})$, lo cual permite estimar la cónica absoluta [Pol99]

El *marco experimental inicial para la autocalibración* consiste en una cámara con calibración intrínseca desconocida pero fija a lo largo de las diferentes imágenes que se toman como inputs. La auto-calibración utiliza la "rigidez" de la escena sin ningún objeto auxiliar [25]

- *Inputs:* $N \geq 3$ imágenes tomadas por una misma cámara bajo condiciones de calibración intrínseca fijas.
- Sólo requiere correspondencias entre puntos.
- *Estrategia:* Realizar pequeños movimientos controlados de cámara.
- *Resolución* basada en métodos lineales frente a los no-lineales.

El *modelo geométrico inicial* para la Reconstrucción 3D esta dado por el embebimiento del marco euclídeo en el marco proyectivo que se realiza a partir de las trasformaciones proyectivas que dejan invariante la cuádrica absoluta $Q_\infty := \{\underline{X} \in \mathbb{P}^n \mid X_0 + \ldots + X_n = 0\}$ Valoración:

- *Inconvenientes:* Precisión menor que la calibración en Laboratorio. No permite incorporar imágenes procedentes de otras fuentes (disponibles en Internet, p.e.)
- *Ventaja:* Proporciona resultados aceptables para una cámara convencional. Es la única opción en algunos casos (exploración remota, reconstrucción a partir de películas antiguas)

El esquema de la sección es como sigue:

1. Cónica absoluta y parámetros intrínsecos
2. Estimación del plano del infinito
3. Estimación de la restricción epipolar
4. Algunas aplicaciones seleccionadas

El método no-lineal sólo requiere siete puntos para estimar la restricción estructural; sin embargo este método es inestable. El método lineal utiliza ocho puntos y, por consiguiente, es redundante pero converge más rápidamente y es más robusto. La mayor parte de los resultados presentados en esta sección fueron obtenidos a lo largo delos años noventa y están asociados a la reconstrucción a partir de una única cámara situada en diferentes localizaciones a poder ser próximas entre sí (pequeña línea base b).

La extensión de la metodología presentada al caso de línea base b más amplia o la reconstrucción a partir de múltiples vistas [26] tomadas con cámaras arbitrarias de calibración desconocida ha sido desarrollada desde finales de los noventa. Algunos de los hitos más conocidos de esta última corresponden a la reconstrucción 3D virtual de las estatuas de Buda en Bamiyahn o bien la reconstrucción del la Plaza Mayor de Praga a partir de imágenes extraídas de Internet.

[25] Para un desarrollo detallado de cuestiones relacionadas con calibración se recomienda el tutorial de Marc Pollefeys http://www.cs.unc.edu/ marc/tutorial/node3.html

[26] Ver apéndice al Cap.3 del módulo 2 para más detalles

Figura 1.27: Reconstrucciones 3D de la estatua de Buda y la Plaza Mayor de Praga

Cónica absoluta y Parámetros intrínsecos

Calibración a partir de rotación y zoom

En este apartado se adopta el enfoque desarrollado por R.Hartley (1997)

La imagen J_i obtenida tras rotación y zoom está relacionada con la original J_0 por una homografía $\mathbf{H}_j$

1. La imagen ω_j de la cónica absoluta (ICA) está dada por $\omega_j = \mathbf{H}_j^{-T} \omega_0 \mathbf{H}_j^{-1}$

2. Si $\mathbf{K}$ es la matriz de calibración con $K_{12} = 0 \Rightarrow \omega = \mathbf{K}^{-T}\mathbf{K}^{-1}$:

$$K = \begin{pmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \Rightarrow \omega = \begin{pmatrix} 1/\alpha_x^2 & 0 & -x_0/\alpha_x^2 \\ 0 & 1/\alpha_y^2 & -y_0/\alpha_y^2 \\ -x_0/\alpha_x^2 & -y_0/\alpha_y^2 & 1 + x_0/\alpha_x^2 + y_0/\alpha_y^2 \end{pmatrix}$$

3. *Problema:* Estimar ω_{ij}: 5 parámetros salvo escala. *Casos particulares*

   - Caso general: $s = K_{12} = 0 \Rightarrow \omega_{12} = 0$
   - Píxeles cuadrados: $s = 0$ y $\alpha_x = \alpha_y \Rightarrow \omega_{11} = \omega_{22}$
   - Punto principal conocido:
     - $s = 0$, $x_0 = 0 \Rightarrow \omega_{13} = 0$
     - $s = 0$, $y_0 = 0 \Rightarrow \omega_{23} = 0$
   - Cada caso da lugar a sistema lineal en ω_{ij}. Habitualmente se requieren 5 imágenes para J_0

El caso general

En presencia de traslaciones, no $\exists$ homografía entre vistas

- *Estrategia general* para autocalibración [Har99]

  1. Estimar una reconstrucción proyectiva tosca de la escena
  2. Identificar el "verdadero" plano del infinito $V = \Pi_\infty$ en la ref. de la Rec3D

3. Aplicar una transformación proyectiva G que lleve $V = \Pi_\infty$ a $X_3 = 0$ (recuperar paralelismo)
4. Actualizar la Rec3D proyectiva a una Rec3D afín: $\mathbf{P}_i \to G\mathbf{P}_i$ y $\Pi_j \to \Pi_j G^{-1}$
5. Identificar las "homografías infinitas"
6. Estimar la calibración de la cámara mediante ω_j
7. Realizar la Rec3D euclídea de la escena

Cálculo de la homografía del infinito

La *Homografía del infinito* para las cámaras i, j se define como la homografía entre J_i, J_j que relaciona las proyecciones respectivas de puntos que están sobre el plano del infinito $X_3 = 0$

1. Matriz transformada por $\mathbf{G}$ de la j-ésima cámara $[\mathbf{M}_j \mid \mathbf{t}_j] = \mathbf{P}_j = \mathbf{P}_j\mathbf{G}$
2. $\mathbf{X} = (\mathbf{x}^\top, 0) \in \{X_3 = 0\}$ se aplica en $\mathbf{M}'_i\mathbf{x} \in J_i$ y $\mathbf{M}'_j\mathbf{x} \in J_j$
3. Ambas imágenes están relacionadas por la *homografía del infinito*: $\mathbf{H}_{ij} = \mathbf{M}'_j\mathbf{M}'^{-1}_i$
4. Como $\omega \subset \{X_3 = 0\} \Rightarrow$ sus proyecciones están relacionadas por $\mathbf{H}_\infty$
5. Elegimos coordenadas para cámara de referencia t.q. $\mathbf{P}'_0 = [\mathbf{I} \mid 0]$ (no necesariamente una cámara física). Entonces:
   - $\mathbf{H}_\infty$ para el par $(0, j)$ es $\mathbf{H}_j = \mathbf{H}_{0j}$ que reescribimos como $\mathbf{M}'_j$
   - La ICA ω_0 en imagen 0 se transforma en $\omega_j = \mathbf{M}'^{-T}_j \omega_0 \mathbf{M}'^{-1}_j$
   - La matriz de calibración K_j se calcula por la factorización de Cholesky.

Estimando el plano del infinito

El esquema que se desarrolla en esta subsección incluye los apartados siguientes:

1. Estimación del plano del infinito
2. Métodos no-lineales: Algoritmo de 7 puntos
3. Métodos lineales: Algoritmo de 8 puntos

Estimación del plano del infinito

El *objetivo* de este apartado es construir una transformación casi-afín mediante programación lineal a partir identificar de datos visible extraídos mediante técnicas de Análisis de Imagen basadas en el agrupamiento de puntos y línea. Los *pasos a dar* son los siguientes:

1. *Orientación:* Multiplicar las matrices de proyección proyectivas $M^p_{\pi_j}$ y $\mathbf{X}_i$ por ± 1 de modo que $P_j\mathbf{X}_i = (u, v, w)^\top$ con $w > 0$ para inducir las dos orientaciones posibles en la imagen real.
2. Introducir *coordenadas de Grassmann* para representar las matrices de proyección :

$$M \leftrightarrow \mathbf{C}^M_G = (c_1, c_2, c_3, c_4)^\top \quad \text{con} \quad c_i = (-1)^k det(\hat{M}^k)$$

   es el determinantes que resulta de suprimir la columna k de M_π

3. Analizar los *semiespacios* asociados a las diferentes orientaciones: Para $\varepsilon = \pm 1$ construir $\mathbf{X}_i^\top \mathbf{V} > 0$ y $\varepsilon \mathbf{C}^{M_j T} \mathbf{V} > 0$

4. Introducir la técnica de *Programación Lineal* con $\varepsilon = \pm 1$ para resolver las dos desigualdades anteriores asociadas a los subespacios

5. Discutir las soluciones según la *Orientación en transformaciones* Elegir una matriz de transformación $\mathbf{G}$ con $r_4(\mathbf{G}) = \mathbf{V}$ y $sign(det(\mathbf{G})) = \varepsilon$

6. Llevar a cabo la *Reconstrucción casi-afín* reemplazando $\mathbf{X}_i \mapsto \mathbf{G}\mathbf{X}_i$ y $M_j \mapsto M_j \mathbf{G}^{-1}$

Una vez realizadas las transformaciones anteriores, hay que trasladar la reconstrucción casi-afín al origen de coordenadas y realizar un escalado anisótropo

Estimación de la matriz fundamental

La matriz fundamental proporciona una representación de la relación bilineal entre puntos homólogos asociados a dos vistas. Por ello, es la restricción estructural para la puesta en correspondencia entre datos homólogos correspondientes a dos vistas. Existen diferentes algoritmos para estimar la matriz fundamental. Si atendemos al mínimo número de puntos necesarios bastan 7 puntos (bajo ciertas condiciones, incluso 5 pares de puntos homólogos permiten estimar $\mathbf{F}$), lo cual proporciona una algoritmo no-lineal. Debido al carácter no-lineal de este algoritmo, la convergencia es más lenta y la inestabilidad mayor que con métodos lineales.

Por ello, el método más eficiente desde el punto de vista computacional está basado en 8 puntos y es *lineal*. Este método utiliza información redundante y por ello puede dar lugar a sistemas que resultan incompatibles. Para evitar este problema, es necesario combinarlo con procedimientos de optimización. Nuevamente, el método más robusto y eficiente viene dado por la metodología RanSaC (P.Torr, 1997).

Métodos no-lineales

Minimizar distancia entre puntos homólogos y líneas epipolares $\rightarrow$ Algoritmo de 7 puntos

1. Funcional a minimizar
$$\sum_{i=1}^{n} [d^2((\mathbf{p}'_i)^\top \mathbf{F}(\mathbf{p}_i) + d^2(\mathbf{p}_i^\top \mathbf{F}(\mathbf{p}'_i)]$$

2. Carácter *no-lineal*: procede de la parametrización elegida para $\mathbf{F}$

3. Se requieren 7 puntos como mínimo.

4. *Valoración:* Resultados mejores que Algoritmo de 8 puntos, pero convergencia más lenta.

Métodos lineales

El método lineal considerado más efectivo es el *algoritmo de* 8 *puntos* para dos imágenes. En este apartado, se muestra de una forma muy esquemática la *estrategia de Hartley* desarrollada para resolver este problema:

1. *Transformar coordenadas de imagen* usando traslación $\mathbf{T}(\mathbf{p}_i) = \hat{\mathbf{p}}_i$ y escalado $\mathbf{T}'(\mathbf{p}'_i) = \hat{\mathbf{p}}'_i$

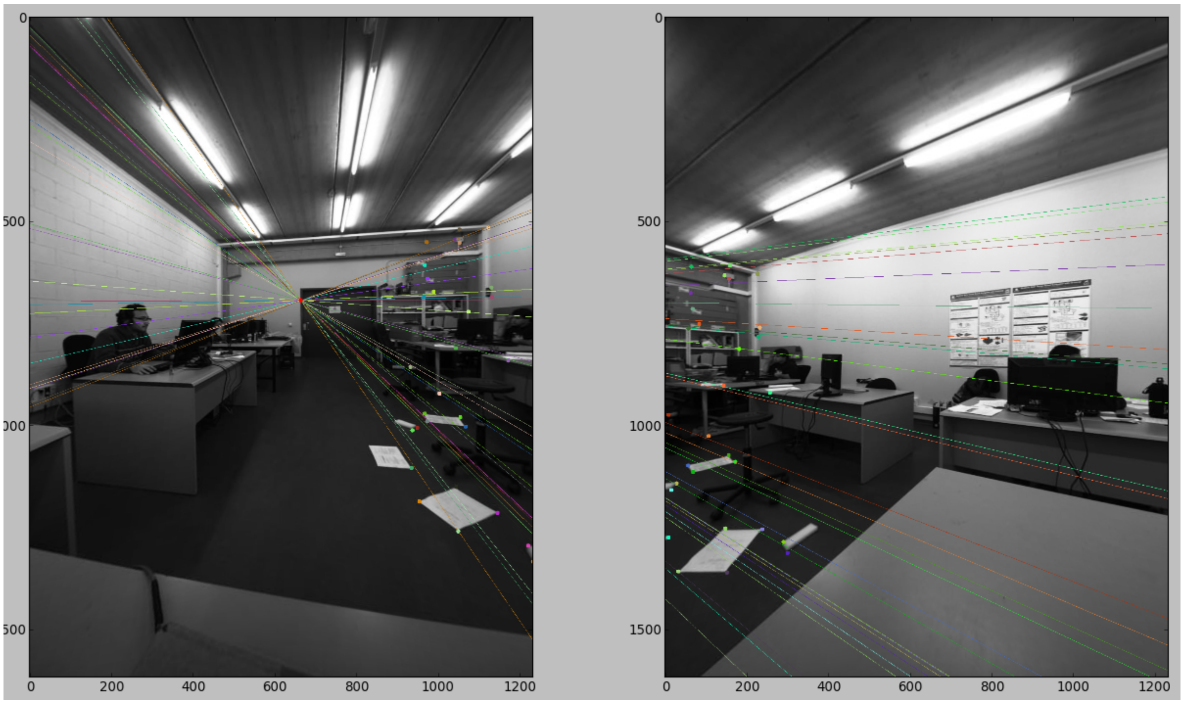

Figura 1.28: Estimación de la matriz fundamental mediante detección de puntos homólogos y líneas epipolares extraídas a partir de un par estéreo

2. *Utilizar LSM lineal para calcular* $\mathbf{F}$ minimizando $\sum_{i=1}^{n}[(\hat{\mathbf{p}}_i')^\top \mathbf{F}\hat{\mathbf{p}}_i]^2$
3. Reforzar $rk(\mathbf{F}) = 2$ usando SVD para $\mathbf{F} = USV^\top$ con $S = diag(r,s,t)$
   - La matriz fundamental $\hat{\mathbf{F}}$ que minimiza la norma de Frobenius es $\hat{\mathbf{F}}' = Udiag(r,s,0)V^\top$
4. *Estimación final de* $\mathbf{F} = (\mathbf{T}')^\top \hat{\mathbf{F}}'\mathbf{T}$