

Capítulo 1: El marco Proyectivo. Modelos de Perspectiva

Javier Finat y Fco Javier Delgado del Hoyo

De una forma muy simplificada, cada vista representa una proyección del mundo real tridimensional sobre el plano de imagen bidimensional ¹. Por ello, los diferentes tipos de proyección $\pi_{\mathbf{C}}$ con centro en un punto \mathbf{C} (al que llamamos foco ó centro de proyección) proporcionan una aproximación de primer orden a la representación de una proyección central. El modelo más simple de proyección corresponde al espacio cartesiano $\pi_{\mathbf{C}} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$.

El modelo cartesiano es insuficiente para abordar propiedades relacionadas con las diferentes entidades geométricas relativas a la distancia, la proporcionalidad asociada a localizaciones oblicuas o bien las relaciones entre diferentes factores de proporcionalidad correspondientes a diferentes localizaciones de la cámara. Por ello, es necesario incorporar estructuras adicionales como la euclídea, la afín ó la proyectiva, en relación con las entidades mencionadas.

Desde principios del siglo XV es común admitir que los modelos de perspectiva proporcionan la representación más apropiada de un objeto o una escena volumétrica. Esta afirmación sólo es válida para una zona próxima al eje óptico y para un ángulo de apertura “estrecho” y, frecuentemente, es incompatible con los modelos de proyección asociados a cámaras reales. Sin embargo, proporciona

- una representación intuitiva de la porción de “mundo” capturada por una cámara;
- un punto de partida para comparar diferentes representaciones o para inicializar procesos vinculados a operaciones de navegación en escenas o de manipulación de objetos;
- una motivación para el desarrollo de un marco geométrico capaz de integrar diferentes tipos de proyección sobre un plano de imagen.

El *espacio proyectivo* n -dimensional \mathbb{P}^n se describe en términos de direcciones del espacio ordinario $(n+1)$ -dimensional \mathbb{R}^{n+1} . Este espacio proporciona el marco geométrico lineal más general, pues incluye a los demás (euclídeo, afín) como casos particulares salvo escala. Asimismo, permite relacionar los puntos de vista asociados a diferentes localizaciones del observador. Por ello, se adopta como el soporte para representar el espacio ambiente y sus proyecciones sobre diferentes planos de imagen. Las transformaciones entre representaciones asociadas a

¹ Esta presentación contrasta con la predominante en el módulo 1, según la cual una imagen digital es un mapa de bits dado por una función a valores discretos en $[0, 255] \cap \mathbb{N}$

diferentes localizaciones se representan mediante transformaciones matriciales regulares, que son elementos de un grupo.

Un “marco geométrico” está dado por un espacio E y un grupo G_E de transformaciones regulares que actúa sobre E . Así, p.e. si E es un espacio vectorial n -dimensional real V , el grupo más “amplio” es el grupo lineal general $GL(n; \mathbb{R})$ que actúa mediante multiplicación a la izquierda:

$$\alpha : GL(n; \mathbb{R}) \times V \rightarrow V \mid \alpha(\mathbf{A}, \mathbf{v}) := \mathbf{A}\mathbf{v} \quad \forall \mathbf{A} \in GL(n; \mathbb{R}), \forall \mathbf{v} \in V$$

Las diferentes representaciones matriciales proporcionan una aproximación de primer orden (es decir, lineal) para representar el cambio en la localización de la cámara, ó el cambio en la apariencia del objeto, p.e.. La descripción de dichas transformaciones regulares mediante los elementos de un grupo permite componer transformaciones o bien invertir transformaciones para deshacer operaciones o bien para relacionar diferentes puntos representaciones.

En este capítulo se presentan elementos básicos de la Geometría Proyectiva orientados a proporcionar un soporte para la Reconstrucción 3D. Una de las claves para entender la elección del marco proyectivo radica en su capacidad de integración de las demás geometrías lineales (definidas localmente por transformaciones lineales). Según el programa de Erlangen (F.Klein, 1872), cada Geometría está caracterizada por un grupo G de transformaciones.²

De cara a las aplicaciones a la Reconstrucción 3D cada grupo G_E de un espacio E está caracterizado por la conservación de una forma cuadrática Q o una forma (bi)lineal B (ó con más generalidad multilineal) que es invariante por la acción del grupo. Una descripción intuitiva de los principales “ejemplos” que aparecen de forma recurrente a lo largo de todo este módulo es la siguiente:

- La *Geometría Euclídea* caracterizada por la conservación de la distancia euclídea (una forma cuadrática no-degenerada definida positiva); generada por rotaciones, traslaciones y simetrías
- La *Geometría Afín* caracterizada por la conservación del paralelismo, que se reinterpreta en términos de la conservación de una forma lineal (un “hiperplano” del infinito); está generada por traslaciones que conservan la “ratio” entre cantidades y, por consiguiente, el área salvo factor de escala.
- La *Geometría Proyectiva* caracterizada por la conservación de la razón (cociente) entre ratios (*cross ratio* desde finales el siglo XVII); generada por proyectividades que “relativizan” las condiciones de traslación en diferentes direcciones.

Esta descripción pone de manifiesto una “jerarquía natural” entre grupos que se traduce en una conservación de propiedades “métricas” a conservar tales como la distancia, la ratio de distancias (equivalente a la conservación del paralelismo)

² La afirmación recíproca también es cierta par cualquier grupo de dimensión finita, aunque no haremos uso de ello

o la razón doble (cociente de ratios). Dados 4 puntos alineados, la *razón doble* se define como

$$\text{Cross}(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4) := \frac{|\overline{\mathbf{p}_1\mathbf{p}_2}| |\overline{\mathbf{p}_3\mathbf{p}_4}|}{|\overline{\mathbf{p}_1\mathbf{p}_3}| |\overline{\mathbf{p}_2\mathbf{p}_4}|}$$

Esta definición se extiende de forma natural a 4 rectas a través de un punto o bien a 4 planos que pasan a través de una recta y así sucesivamente. Asimismo, permite introducir invariantes para 5-tuplas de puntos (rectas, planos) como la colección de 5 cross-ratios asociadas a las 5 formas de seleccionar una 4-tupla de puntos (rectas, planos), y así sucesivamente

La jerarquía entre las diferentes geometrías se describe algebraicamente en términos de inclusiones entre grupos: Todos los grupos considerados aquí son subgrupos del grupo lineal general $GL(n+1; \mathbb{R})$ salvo factor de escala, al que llamamos el proyectivizado del grupo lineal general y lo denotamos mediante $\mathbb{P}GL(n+1; \mathbb{R})$ ³. Por ello, se obtienen inclusiones naturales entre las geometrías lineales clásicas a la que llamaremos la *Jerarquía Geométrica*. En todos los casos, el grupo actúa sobre el vector columna que representa las coordenadas por multiplicación a la izquierda de las matrices que representan sus elementos.

Según el programa de Erlangen de F.Klein (1873) cada Geometría Lineal está determinada por un Grupo, al que se llama grupo estructural, y recíprocamente. De una forma abstracta, cualquier geometría está generada por simetrías continuas o discretas. La Reconstrucción 3D desarrolla el punto de vista de F.Klein proporcionando modelos y herramientas que permiten estimar, extraer y visualizar propiedades relativas a objetos o escenas tridimensionales a partir de la información contenida en varias imágenes 2D. En el caso más simple correspondiente a objetos o escenas fabricados por el hombre, las representaciones visuales aparecen asociadas a diferentes modelos de perspectiva⁴.

Desde un punto de vista de modelado matemático, la Reconstrucción 3D utiliza y extiende métodos de Geometría Proyectiva Clásica para representar objetos y simular movimientos de cámara en torno a los mismos⁵. Este marco geométrico permite reformular de una manera sintética

- la *calibración*, es decir, la estimación de parámetros asociados a las cámaras;
- las *restricciones epipolares* en términos de relaciones bilineales entre elementos homólogos (puntos ó líneas) para dos vistas;
- la extensión a *restricciones multipolares* en términos de relaciones multilineales (tensores) entre elementos lineales contenidos en múltiples vistas;

³ El grupo lineal general es el grupo de las transformaciones regulares, es decir, matrices de tamaño $(n+1) \times (n+1)$ con determinante no-nulo

⁴ Para una visualización de modelos de perspectiva con asistencia al dibujo ver <https://www.youtube.com/watch?v=KoOBsmYMGgM>

⁵ A partir del capítulo 3 se muestran modelos y herramientas que no necesitan modelos de perspectiva para la puesta en correspondencia, sino tan sólo propiedades de las proyectividades en términos de incidencia para haces de rectas, lo cual justifica a posteriori la necesidad del marco proyectivo

- la *visualización* de nuevas vistas mediante la simulación de cámaras virtuales;
- la *generación* semi-automática de modelos 3D.

El enfoque basado en la Geometría Proyectiva debe su popularidad a que permite conectar todas las aproximaciones basadas en elementos lineales. Su eficiencia procede del carácter sintético y la elegancia en la resolución global de las ecuaciones vinculadas a la Reconstrucción 3D [Lon81], la reducción a algoritmos de tipo lineal [Har97] y la posibilidad de integrar *todas* las vistas y cámaras posibles en un único objeto. Sin embargo, presenta la contrapartida asociada a deformaciones en la visualización de los objetos $B^{(\alpha)}$ presentes en la escena. La representación proyectiva de estos objetos se realiza mediante una inmersión en un marco geométrico que es fuertemente no-lineal: al alejarse de la dirección principal de visión, las distorsiones crecen de forma no lineal (difícil de controlar), incrementando el coste computacional de la visualización.

Asimismo, aunque existe una “métrica natural” en el espacio proyectivo (Fubini-Study), esta métrica distorsiona la métrica euclídea utilizada en fotogrametría terrestre o la esférica utilizada en imágenes satelitales. La corrección automática de las distorsiones asociadas al espacio proyectivo (curvatura negativa) es poco utilizada, requiere conocimientos que sólo se explican en cursos avanzados de Matemáticas y no ha sido implementada. Por ello, esta corrección resulta de poca utilidad para los levantamientos fotogramétricos convencionales que se necesitan en aplicaciones vinculadas a Arquitectura o Ingeniería, donde se sigue recurriendo a refinamientos de modelos de perspectiva que es preciso actualizar con el desplazamiento relativo del observador (real o virtual). Por ello y para evitar problemas con los cambios de sistemas coordenados asociados a la estructura global del espacio proyectivo, restringimos el uso de métricas al caso euclídeo, estrictamente

A pesar de estos “inconvenientes” que requieren una corrección posterior de las distorsiones asociadas al modelo, la Geometría Proyectiva proporciona el único marco general que es capaz de integrar las diferentes aproximaciones al problema de la Reconstrucción 3D y de hacer compatibles representaciones arbitrarias de una misma escena desde cualquier punto de vista, incorporando el punto de vista asociado a cámaras virtuales generadas por el usuario de forma interactiva para navegar el objeto o la escena.

Asimismo y en relación con cuestiones complementarias de Informática Gráfica (rayos de luz como semirrectas o elementos del espacio proyectivo) o más avanzadas de renderización (incorporación de la radiometría y propiedades de los materiales), el marco proyectivo permite gestionar la geometría asociada a diferentes agentes o varias fuentes de luz. Los “objetos” y transformaciones asociadas al marco proyectivo proporcionan un marco estructural para resolver rápidamente problemas de elevada complejidad que antes requerían un elevado número de iteraciones y una corrección (frecuentemente manual) de errores locales. De ahí que el enfoque basado en Geometría Proyectiva sea predominante desde principios de los noventa para la reconstrucción 3D de escenas o de objetos ([Fau93], [Har00], [For03], [Pol07], [Sze10]).

En todo el segundo módulo se supone que objetos y escenas son estáticos y rígidos, aunque la cámara puede estar en movimiento (real o simulado). El análisis de objetos en movimiento es el tópico principal del módulo 3, mientras que los objetos deformables se consideran en el módulo 4. La integración de todos estos puntos de vista se lleva a cabo en el módulo 5.

De acuerdo con ello, el capítulo está organizado en las siguientes secciones

1. Elementos de Geometría Proyectiva en baja dimensión
2. Estimación de las transformaciones proyectivas
3. Modelos geométricos de cámaras
4. Modelos de perspectiva
5. Estimación y Reconstrucción Euclídea

1. Elementos de Geometría Projectiva

Desde el punto de vista de la Visión Computacional, los *inputs* proceden del procesamiento y análisis de una o varias vistas. A partir de los datos (puntos y líneas) que se puedan extraer de la información se pretende alcanzar los siguientes *objetivos*:

1. *Comprensión de la estructura 3D*
2. *Generación de representaciones coherentes con los datos*
3. *Visualización interactiva de un modelo 3D*

Restricción de unicidad: El modelo geométrico debe ser único módulo la acción de un grupo de transformaciones G que depende del marco geométrico elegido (euclídeo, afín, proyectivo). Para evitar problemas de ambigüedad, es necesario determinar de forma única los parámetros asociados a las características geométricas de la escena y ópticas de la cámara, así como su evolución temporal para movimientos controlados de cámara.

Los *problemas a resolver* son:

1. *Problema de Correspondencia:* “Pegar” datos comunes contenidos en vistas en *hechos* \rightarrow Las vistas como mapas de bits.
2. *Problema de la Reconstrucción:* Recuperar la *estructura 3D* \rightarrow Las vistas como imagen de una proyección \rightarrow Correspondencia entre haces de rectas
3. *Problema de la Re-proyección* o generación de nuevas vistas mediante movimientos de ratón: Consistencia de una nueva vista *3D* con otras previas.

Nota.- Habitualmente, se requiere que los problemas anteriores tengan una solución única para evitar las ambigüedades que se derivarían de la existencia de diferentes soluciones; caso de no verificarse esta condición se dice que el problema está “mal puesto” (*ill-posed*) en el sentido de Hadamard. Cualquier problema no-lineal admite varias soluciones; incluso la mayor parte de los problemas lineales también admiten infinitas soluciones. Por ello, la etiqueta de “mal puesto” debe entenderse en el sentido de añadir condiciones adicionales para garantizar que la solución sea única. De este modo, se consigue acelerar la convergencia hacia “una” solución y evitar problemas de inestabilidad en el algoritmo a implementar

En cualquier caso, es preciso tener siempre presente que el objetivo de la Reconstrucción 3D es proporcionar nuevas vistas o bien una representación volumétrica de escenas u objetos a partir de varias cámaras o bien de una cámara en movimiento. Una representación esquemática típica (*pipeline* de interés tanto para el modelado 3D (que afecta a la reconstrucción) como para la renderización (que afecta a la visualización) se escalona en los pasos siguientes:

Hechos \rightarrow Modelos \rightarrow Movimiento \rightarrow Reproyección del color

Este esquema reaparece en diferentes lugares del módulo 2 y se aplica tanto a propiedades geométricas de la escena, como radiométricas de los objetos. La geometría de los rayos (semirrectas que pasan por el foco de la cámara, p.e.) proporcionan una conexión entre ambos tipos de propiedades.

1.1. Cámaras y proyecciones

La Reconstrucción 3D tiene como objetivo facilitar una *visualización interactiva* de un objeto 3D a partir de una o varias vistas capturadas por cámaras en diferentes localizaciones o bien por una cámara en movimiento.

- La visualización concierne a su representación como una pseudo-inversa asociada a la proyección de un espacio 3D sobre el plano 2D de cámara ⁶.
- El carácter interactivo consiste en generar nuevas vistas mediante movimientos de ratón.

Este tipo de representación formaliza el enfoque adoptado en el Renacimiento ilustrado por el grabado de Durero de la figura ??.

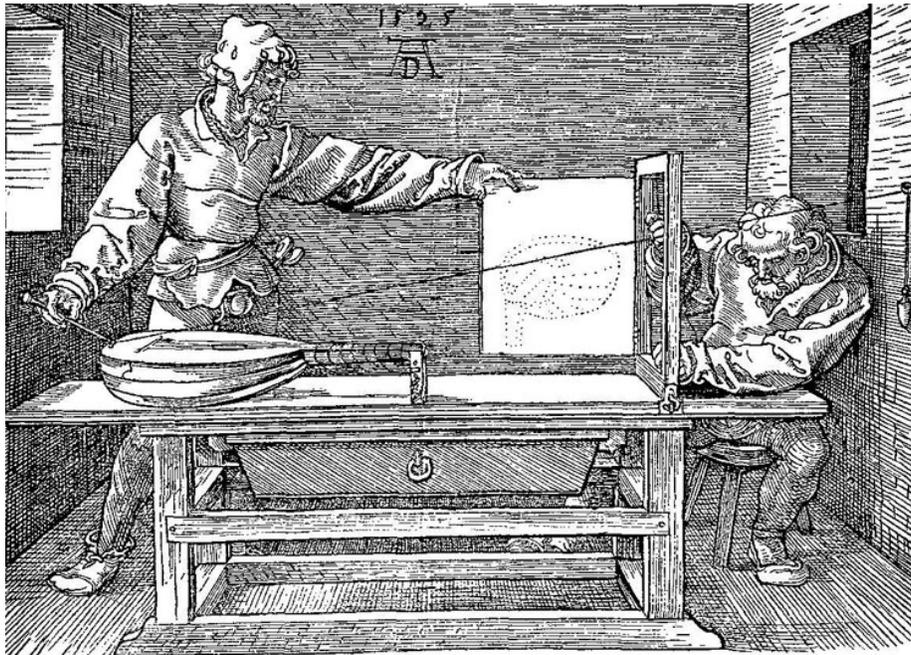


Fig. 1: Grabado de Alberto Durero que ilustra las representaciones proyectivas

⁶ Cualquier aplicación proyección está representada localmente por una matriz rectangular, por lo que no existe una inversa propiamente dicha

La *simplificación básica inicial* consiste en interpretar cada vista como una proyección central $\pi_{\mathbf{C}} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ sobre el plano de imagen (que extendemos a todo el plano \mathbb{R}^2) desde un punto ideal al que se llama foco o centro de la cámara \mathbf{C} eventualmente en movimiento. Es claro que esta interpretación no se corresponde con la realidad, pues el dispositivo óptico de una cámara no es un punto, sino un acoplamiento entre objetos materiales curvados (lentes); ello da lugar a diferentes tipos de distorsión (tipo tonel o tipo cojín, p.e.) cuya estimación se aborda más abajo. A pesar de no ajustarse a la realidad, el modelo de proyección central de una cámara proporciona una simplificación muy útil para inicializar el problema de la Reconstrucción.

En los tres primeros capítulos de este módulo predomina el enfoque geométrico, incorporándose un análisis radiométrico a partir del capítulo 4. Esta organización implica que inicialmente la cámara se considera como un dispositivo geométrico cuyo output debemos interpretar en términos de las primitivas extraídas en la fase de análisis de imagen. La interpretación geométrica de una vista arbitraria (como imagen de una proyección) requiere superponer diferentes estructuras geométricas (proyectiva, afín, euclídea) al espacio cartesiano. Esta superposición da lugar a diferentes “modelos geométricos” de cámaras.

En esta subsección sólo se presenta una descripción cualitativa, reservando una descripción más formal para las siguientes subsecciones; esta elección está motivada por el mayor grado de abstracción que presenta el espacio proyectivo en relación con otras representaciones más tradicionales.

1.1.1. Descripción cualitativa de una cámara

Las *cámaras fotográficas* están formadas por una colección de lentes (convergentes o divergentes) con ejes ópticos alineados. Este modelo es complicado y da lugar a una casuística poco relevante desde el punto de vista del análisis de datos contenidos en cámaras fotográficas digitales (otros tipos de cámaras asociados a diferentes componentes del espectro visible o no se comentan más adelante). Por ello, es necesario simplificarlo al máximo.

El modelo geométrico más sencillo para una cámara es el de una proyección central en el que cada rayo $r_i = \langle \mathbf{C}, \mathbf{P}_i \rangle$ pasa por el centro \mathbf{C} de la cámara y cortan al plano $\Pi_{\mathbf{C}}$ de la cámara en un punto \mathbf{p}_i . La proyección está representada mediante $\pi_{\mathbf{C}}(\mathbf{P}_i) = \mathbf{p}_i$. La versión más simple de este modelo (implícita en los tratados de perspectiva de L.B. Alberti o A. Durero, p.e.) se formula actualmente en términos de la ecuación de la proyección perspectiva dadas por:

$$x = f \frac{X}{Z} \quad y = f \frac{Y}{Z} \quad (1)$$

donde f es la *longitud focal* de la cámara, es decir, la distancia entre el plano de imagen de la cámara y el centro de proyección \mathbf{C} (suponemos que el eje es perpendicular al plano de imagen $\Pi_{\mathbf{C}}$). Esta descripción muestra que todos los puntos \mathbf{P} situados sobre un “rayo” r (línea a través del centro de proyección) se proyectan sobre el mismo punto \mathbf{p} perteneciente al plano de imagen. Este argumento proporciona la primera justificación para usar coordenadas homogéneas,

es decir, definidas salvo factor de proporcionalidad o de escala.

La igualdad precedente expresa la proporcionalidad entre dos triángulos asociados al eje principal de la cámara con extremo superior de la hipotenusa dado por $\mathbf{P} = (X, Y, Z)^T$ para el punto espacial y $\mathbf{p} = (x, y)^T$ para el punto situado sobre el mismo rayo a través del centro de cámara sobre el plano de cámara ⁷

Denotamos mediante $St(\mathbf{C})$ a la “estrella” de (semi)rectas que pasan por el punto \mathbf{C} . Si representamos cada dirección de proyección por un punto de una esfera \mathbb{S}^2 (vector director normalizado), la estrella está parametrizada por los puntos del cociente de la esfera \mathbb{S}^2 por la aplicación antipodal $a : \mathbb{S}^2 \rightarrow \mathbb{S}^2$ (en la que se identifican puntos diametralmente opuestos), es decir, es un plano proyectivo \mathbb{P}^2 . La representación proyectiva permite interpretar los elementos de la “estrella” como si fueran semirectas del espacio ordinario, lo cual es coherente con la interpretación que se realiza en Informática Gráfica en términos de “rayos” ó semirectas (de utilidad para Ray Casting y Ray Tracing, p.e.).

El *campo de visión* de la cámara restringe $St(\mathbf{C})$ a un *cono proyectante* de rayos $r_i = \langle \mathbf{C}, \mathbf{P}_i \rangle$ donde \mathbf{P}_i son puntos del espacio ordinario. La reconstrucción 3D de un objeto volumétrico $B^{(\alpha)}$ parte de la intersección $St(\mathbf{C}) \cap B^{(\alpha)}$. El *contorno aparente* es una curva alabeada (en el espacio tridimensional) definida como el borde $\partial(St(\mathbf{C}) \cap B^{(\alpha)})$.

La *silueta* $S_{\mathbf{C}}^{(\alpha)}$ capturada desde \mathbf{C} es la proyección sobre el plano de imagen $\Pi_{\mathbf{C}}$ del contorno aparente. Pequeños desplazamientos de una cámara digital generan siluetas próximas que nunca están relacionadas por una traslación, ni siquiera cuando el movimiento (real o aparente) de la cámara es muy sencillo como ocurre con pares de vistas tomadas de forma simultánea por un dispositivo estéreo. Es necesario desarrollar métodos que faciliten la puesta en correspondencia entre elementos comunes usando puntos significativos, siluetas o regiones con características comunes.

Ejercicio.- Si $\pi(\mathbf{P}) = \mathbf{p}$ para una proyección $\pi : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ con centro \mathbf{C} , verificar que todos los puntos situados sobre $(\mathbf{C})\mathbf{P}$ se proyectan sobre \mathbf{p} (*Indicación:* Elige \mathbf{C} como origen de coordenadas y representa cualquier punto sobre el rayo que pasa por $\mathbf{P} = (X, Y, Z)^T$ como $\lambda\mathbf{P}$).

1.1.2. Representaciones coordenadas para datos geométricos

Inicialmente los *elementos significativos* 0D para la reconstrucción 3D de escenas se expresan en términos de coordenadas píxel (en imágenes digitales) o vóxel (unidades básicas volumétricas asociadas a escaneos, p.e.) que admiten una representación en términos de diferentes representaciones coordenadas con la siguiente notación:

- En el plano de imagen se tienen *Coordenadas euclídeas* $(x, y)^T$ (cuya discretización se asimila a coordenadas píxel), *Coordenadas afines* $(x, y, 1)^T$

⁷ En lo sucesivo, utilizaremos letras mayúsculas para denotar puntos del espacio y minúsculas para puntos del plano, insertando un subíndice para representar diferentes elementos (punto, línea, etc) y un superíndice α, β, \dots cuando se desee especificar el plano al que pertenecen



Fig. 2: Ejemplos de proyecciones en perspectiva

y *Coordenadas proyectivas* $[x_0 : x_1 : x_2]^T$ para un punto \mathbf{p} o \mathbf{x} en cada vista $2D$.

- En el espacio ambiente se tienen *Coordenadas euclídeas* $(X, Y, Z)^T$ (cuya discretización se asimila a coordenadas vóxel), *Coordenadas afines* $(X, Y, Z, 1)^T$ y *Coordenadas proyectivas* $[X_0 : X_1 : X_2 : X_3]^T$ para un punto \mathbf{P} o \mathbf{X} en la escena $3D$.

La captura automática de información en una imagen digital aparece asociada a una *discretización* de una vista de escena que representamos mediante $\mathbf{P}_i = \pi(\mathbf{p})$. En la práctica, sólo se pretende almacenar la información que sea más significativa ("hechos salientes") para facilitar su gestión. En ocasiones la extracción de hechos puede presentar una elevada complejidad (ver figura más abajo), por lo que para simplificar en este capítulo nos restringimos a objetos o escenas creados por el hombre, centrando la atención en representaciones arquitectónicas o escenarios urbanos.

El *problema a resolver* consiste en describir la proyección π a partir de una "cantidad suficiente" de puntos \mathbf{p}_i en imagen. Por su propia naturaleza, la proyección requiere disponer de una representación de la proyección en términos de los "rayos" que pasan por el foco de la cámara \mathbf{C} .

Si se denota mediante $\ell_i = \langle \mathbf{C}, \mathbf{P}_i \rangle$ la línea que conecta cada punto de la escena \mathbf{P}_i con el foco \mathbf{C} de la cámara, se tiene que $\pi(\ell_i) = \mathbf{p}_i$. Análogamente, cualquier línea del espacio ambiente $L^{(\alpha)}$ que no pase por \mathbf{C} se proyecta sobre una línea $\ell^{(\alpha)} \subset \Pi_{\mathbf{C}}$ del plano de la cámara. Estas observaciones sugieren un tratamiento más sistemático en términos de rectas del espacio ordinario como elementos básicos para la representación (este enfoque es común a la Informática Gráfica).

1.1.3. Noción de espacio proyectivo

Los puntos situados sobre las rectas del espacio cartesiano \mathbb{R}^n que pasan por el origen se representan mediante $\mathbb{R}^n - \{0\}$. Las direcciones de dichas rectas son las clases por la relación de equivalencia $\mathbf{a} \sim \mathbf{b}$ si y sólo si $\exists \lambda \neq 0$ tal que

$\mathbf{a} = \lambda \mathbf{b}$. Cada dirección representa un punto del espacio proyectivo \mathbb{P}^{n-1} . Por tanto,

$$\mathbb{P}\mathbb{R}^n := (\mathbb{R}^{n+1} - \{\mathbf{0}\}) / \sim \simeq \mathbb{S}^n / \sim$$

donde el segundo isomorfismo proporciona una parametrización del espacio proyectivo nD en términos de la esfera: cada dirección del espacio está representada por dos puntos antipodales que se identifican por la relación \sim . En particular,

1. La *recta proyectiva* $\mathbb{P}^1 := (\mathbb{R}^2 - \{\mathbf{0}\}) / \sim \simeq \mathbb{S}^1 / \sim$ y proporciona un modelo geométrico (pencil) de las direcciones del plano cartesiano $2D$ "completado" (rectas que pasan por el origen del plano) con los elementos del infinito (dados en este caso por un único punto proyectivo).
2. El *plano proyectivo* $\mathbb{P}^2 := (\mathbb{R}^3 - \{\mathbf{0}\}) / \sim \simeq \mathbb{S}^2 / \sim$ y proporciona un modelo geométrico (star) de las direcciones del espacio cartesiano $3D$ "completado" con los elementos del infinito (una recta proyectiva)
3. El *espacio proyectivo* $\mathbb{P}^3 := (\mathbb{R}^4 - \{\mathbf{0}\}) / \sim \simeq \mathbb{S}^3 / \sim$; intuitivamente, proporciona un modelo geométrico de las direcciones del espacio cartesiano $4D$ "completado" con los elementos del infinito (un plano proyectivo).

En Matemáticas se suelen denotar mediante $[x_0 : x_1 : \dots : x_n]$ a las coordenadas proyectivas del espacio proyectivo n -dimensional \mathbb{P}^n ; en Reconstrucción $3D$ está más extendida la notación $[x_1 : x_2 : \dots : x_{n+1}]$ (utilizada sobre todo en [Har00]); por ello, habitualmente se adopta esta última en estos apuntes.

En las subsecciones siguientes se abordan diferentes tipos de espacios lineales (proyectivo, afín, euclídeo) y sus transformaciones estructurales que proporcionan una jerarquía natural entre los diferentes marcos estructurales para la Reconstrucción $3D$. Antes de abordar esta descripción explícita, se presentan dos apartados relativos a una justificación perceptual del espacio proyectivo y algunas observaciones muy generales sobre la estimación de datos en el marco proyectivo.

Ejercicio.- Verifica que un polinomio real $f(\underline{x})$ de grado d está bien definido sobre el espacio proyectivo \mathbb{P}^n si y sólo si $f(\lambda \underline{x}) = \lambda^d f(\underline{x})$ para cualquier constante no-nula $\lambda \in \mathbb{R}^*$.

En particular, la anulación de ecuaciones lineales $\sum_{i=1}^{n+1} a_i x_i$ definen hiperplanos en \mathbb{P}^n , la anulación de ecuaciones cuadráticas $\sum_{1 \leq i, j \leq n+1} a_{ij} x_i x_j$ definen hipercuádricas en \mathbb{P}^n , y así sucesivamente. En este Curso estamos interesados sobre todo en los casos $n = 2$ y $n = 3$; la mayor parte de los resultados se extienden sin mayores problemas a dimensión arbitraria.

1.1.4. Motivación perceptual del espacio proyectivo

La idea básica inicial consiste en interpretar las líneas del espacio cartesiano ordinario que pasan por un origen (interpretable como la posición del observador) como si fueran los puntos del espacio proyectivo

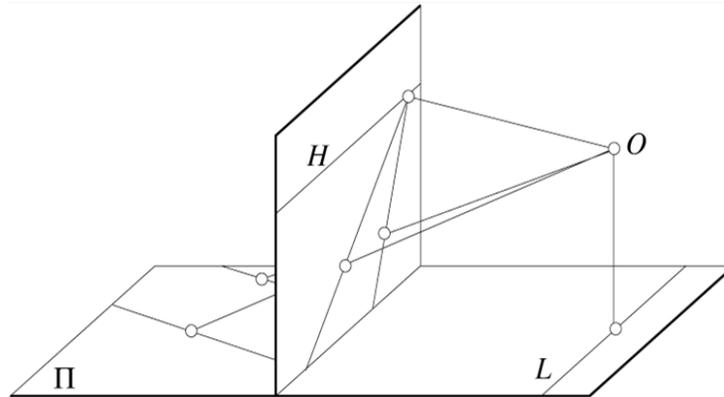


Fig. 3: Invariancia de las condiciones de incidencia para puntos y líneas en relación con proyecciones y secciones

Este enfoque es compatible con la *percepción humana* según la cual las líneas paralelas se “juntan” en un “punto del infinito”. Esta idea se utiliza en los modelos de perspectiva utilizados desde comienzos del Renacimiento y se trasladan en el s.XIX para representar la proyección central para una cámara tipo agujero de alfiler (*pinhole*, en lo sucesivo). Una formalización de esta idea se presenta en el apartado siguiente.

Según esta *representación* Los rayos r a través del foco o centro \mathbf{C} de la cámara son puntos del espacio proyectivo \mathbb{P}^3 . Por consiguiente, para una proyección central $\pi_{\mathbf{C}}$ cualquier recta que pase por el foco de la cámara \mathbf{C} se proyecta sobre el propio foco, pero cualquier otra recta $\mathbf{r} = \langle \mathbf{P}_i, \mathbf{P}_j \rangle$ se proyecta sobre una única recta $\pi(\mathbf{r}) = \langle \mathbf{p}_i, \mathbf{p}_j \rangle$ contenida en el plano de imagen $\Pi_{\mathbf{C}}$.

Como $\mathbf{C} \notin \mathbf{r}$, cualquier recta contenida en el plano $\langle \mathbf{P}_i, \mathbf{P}_j \rangle$ se proyecta sobre la misma recta $\pi(\mathbf{r})$ del plano imagen, es decir, necesitamos añadir condiciones adicionales relacionadas con la incidencia entre elementos lineales para representar de forma unívoca los elementos del espacio. La figura 3 ilustra el significado de la invariancia de las condiciones de incidencia para puntos y líneas (operaciones básicas de la Geometría Projectiva).

Esta observación plantea el problema de seleccionar qué datos son *visibles* y qué elementos son *invariantes* con respecto a las representaciones basadas en rectas. La visibilidad es un problema no trivial que se resuelve mediante haces de rectas y se representa usando mapas de profundidad en los que sólo se visualiza los objetos que están más cercanos con respecto a los más lejanos parcialmente ocluidos⁸. La figura 4 ilustra la idea básica.

Es fácil ver que las distancias, los ángulos, las áreas, la ratio de longitudes, las condiciones de paralelismo ó de perpendicularidad no se conservan en diferentes vistas de un mismo objeto. Más abajo veremos que la “razón doble” (cross-ratio)

⁸ Los algoritmos de visibilidad se presentan en el capítulo 4 en relación con conos proyectantes y envolventes visuales

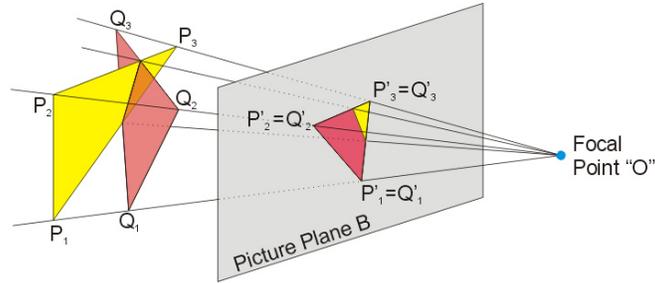


Fig. 4: Principio de visibilidad en perspectiva

entre longitudes de segmentos cortados por 4 rectas de un haz es el invariante fundamental. En la práctica, es conveniente contar con una estimación de la tasa de variación de las cantidades geométricas mencionadas.

Por el momento, nótese que la existencia de diferentes representaciones en perspectiva de una misma escena plantea el problema de *compatibilizar vistas* y *generar de forma semi-automática nuevas vistas* a partir de las ya conocidas. Para ello es necesario identificar elementos que puedan permanecer invariantes para vistas próximas, realizar una implementación computacional de las transformaciones proyectivas, corregir distorsiones y visualizar los resultados.

Ejercicio.- Expresa vectorialmente (1) el ángulo entre dos líneas del plano; (2) el ángulo dos planos del espacio; (c) el área correspondiente a un triángulo en el plano; (d) volumen de un tetraedro en el espacio.

1.1.5. Estimación de hechos geométricos básicos

Para construir objetos geométricos lineales en la escena, es necesario estimar hechos lineales básicos, es decir, puntos \mathbf{p}_i y líneas ℓ_{ij} contenidos en cada vista y sus relaciones de incidencia $\mathbf{p}_i \in \ell_{ij}$. Las condiciones de incidencia más sencillas y relevantes corresponden a

- Recta $\ell_{ij} = \mathbf{p}_i \times \mathbf{p}_j$ determinada por dos puntos \mathbf{p}_i y \mathbf{p}_j .
- Punto $\ell_k \times \ell_m = \mathbf{p}_{km}$ como intersección de dos rectas ℓ_k y ℓ_m .

donde el producto cruzado $\mathbf{v}_i = (a_i, b_i, c_i)^T$ y $\mathbf{v}_j = (a_j, b_j, c_j)^T$ se calcula como

$$\mathbf{v}_i \times \mathbf{v}_j = \begin{pmatrix} 0 & -c_1 & b_1 \\ c_1 & 0 & -a_1 \\ -b_1 & a_1 & 0 \end{pmatrix} \begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix} = \begin{pmatrix} b_1 c_2 - b_2 c_1 \\ a_2 c_1 - a_1 c_2 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}$$

que permite calcular una recta $\ell_{ij} = \mathbf{p}_i \times \mathbf{p}_j$ que pasa por dos puntos y un punto $\mathbf{p}_{km} = \ell_k \times \ell_m$ como intersección de dos rectas.

La estimación de objetos o de escenas debe tener en cuenta diferentes tipos de información.

- *Información geométrica:* Debemos identificar “hechos salientes aislados” (vértices ó máximos de intensidad, p.e.) y elaborar estrategias para la puesta en correspondencia entre datos homólogos. Los *elementos a estimar* más importantes para la Reconstrucción 3D son la disparidad (diferente localización de datos homólogos en imágenes próximas) y la profundidad (distancia con respecto a la cámara).
- *Información radiométrica:* Identificar datos densos sobre regiones que contienen información sobre el color, las texturas, la iluminación, caracterizando de forma aproximada la distribución suavizada. Los *elementos a estimar* más importantes para la Reconstrucción 3D están relacionados con el comportamiento de la respuesta de los materiales con respecto a las condiciones de iluminación. Esta información se “codifica” en términos del mapa de reflectancia bidireccional. Una buena estimación facilita la renderización, es decir, la representación planar de la escena que incluye efectos ópticos para dar la sensación de profundidad.

Algunos prerrequisitos de las soluciones a desarrollar son la precisión (acotación del error) o la *invariancia* con respecto a transformaciones en el espacio y sus proyecciones planas. La información radiométrica es muy sensible con respecto a las condiciones de iluminación; por ello, su análisis para la Reconstrucción presenta una mayor dificultad.

Un *esbozo de soluciones* debe tener en cuenta la información que se puede extraer de forma automática que incluya

- *Aspectos geométricos* que en una primera aproximación es posible imaginar como mapas de perspectiva. Existen diferentes formas de representar modelos de perspectiva (frontal, angular, oblicua); todos ellos facilitan la emergencia de una estructura que puede ser visualizada desde diferentes puntos de vista. La clave para la visualización actual o la generación de nuevas vistas radica en una implementación eficiente de la puesta en correspondencia entre diferentes vistas. El objetivo es alcanzar una simulación tosca de la visión binocular a partir de varias vistas cuyo objetivo es la obtención de la estructura o de la forma (structure/shape) del objeto a partir del movimiento (SFM/sfm) que es uno de los tópicos centrales de este módulo.
- *Aspectos radiométricos:* Bajo condiciones de iluminación controladas, mediante diferentes técnicas de filtrado de color (ver módulo 1) es posible establecer una correspondencia entre regiones con propiedades radiométricas “alineadas”, eventualmente bajo diferentes condiciones de iluminación. Formalmente, el “alineamiento” se realiza sobre una versión simplificada de “campos de luz” que se desarrolla más adelante ⁹.

⁹ La función plenóptica proporciona el marco habitual estática para los campos de luz,

En la práctica, debido a la naturaleza discreta y contaminada en fases iniciales de captura (ruido, p.e.) ó procesamiento (filtrado, efectos de escalonamiento, etc), no se puede esperar que las ecuaciones algebraicas correspondientes a relaciones de incidencia se verifiquen de forma exacta. Por ello, la estimación de datos geométricos ó radiométricos debe minimizar la expresión algebraica esperada; esta expresión corresponde en este caso a ecuaciones de primer grado para las que existen diferentes procedimientos de optimización.

Ejercicio.- Comprueba que

- La intersección $\ell \times \ell'$ de dos rectas $\ell = (a, b, c)^T$ y $\ell' = (a, b, c')^T$ es el punto $\ell \times \ell' = (c - c')(b, -a, 0)^T$.
- La recta $\ell_{12} = \mathbf{p}_1 \times \mathbf{p}_2$ que pasa por los puntos $\mathbf{p}_1 = (x_1, y_1, 0)^T$ y $\mathbf{p}_2 = (x_2, y_2, 0)^T$ es $\ell = (0, 0, x_1 y_2 - x_2 y_1)^T$

Nota.- Si se interpretan las componentes como coordenadas homogéneas de rectas o de puntos, la primera afirmación muestra la intersección de dos rectas paralelas en un “punto del infinito”; la segunda afirmación muestra la recta que pasa por dos puntos del infinito como una “recta del horizonte”.

1.2. El espacio proyectivo y sus transformaciones

El espacio proyectivo \mathbb{P}^n se obtiene “añadiendo” los “elementos del infinito” al espacio ordinario; de una manera intuitiva, los elementos del infinito corresponden a los puntos en los que se cortan las líneas paralelas del espacio cartesiano ordinario. En esta subsección se introducen algunas nociones básicas y se ilustran con resultados asociados al análisis de imagen. Para fijar ideas, se agrupan los contenidos en los siguientes apartados:

1. Descripción del espacio proyectivo
2. Dualidad entre puntos \mathbf{p} y líneas ℓ del plano proyectivo \mathbb{P}^n
3. Elementos del infinito: líneas del horizonte y puntos de fuga
4. Cónicas del plano proyectivo

El primer paso de la estrategia que acabamos de describir consiste en identificar los elementos del infinito. En principio, cualquier punto o recta contenido en una vista puede ser un elemento del infinito, por lo que las transformaciones a definir e implementar deben ser aplicables a dichos elementos. Por ello, empezamos completando una vista con las direcciones del infinito, lo cual conduce a una representación de las direcciones del espacio ordinario mediante un plano proyectivo.

aunque esta formulación presenta limitaciones que se abordan en el capítulo 6 del módulo 2 (Reconstrucción 3D) y en el módulo 5 (Visión Estéreo Dinámica)

1.2.1. Descripción del espacio proyectivo

El *modelo proyectivo real para una vista* está dado por el conjunto de direcciones del espacio cartesiano ordinario. Formalmente:

$$\mathbb{RP}^2 := \mathbb{R}^3 - \{0\} / \mathbb{R}^* = \mathbb{S}^2 / \text{sim}$$

donde el paso al cociente por el grupo multiplicativo de las unidades \mathbb{R}^* (elementos no nulos) representa la identificación de puntos situados a lo largo de la misma dirección (relacionados entre sí mediante homotecias) y la relación \sim representa el paso al cociente por la aplicación antipodal. De forma análoga, el *modelo proyectivo real para el espacio ambiente* (mundo) está dado por:

$$\mathbb{RP}^3 := \mathbb{R}^4 - \{0\} / \mathbb{R}^* = \mathbb{S}^3 / \text{sim}$$

con una interpretación similar al del caso planar, pero en una dimensión más para el espacio proyectivo tridimensional. Habitualmente, omitiremos el símbolo \mathbb{R} dentro de la descripción de los espacios proyectivos, sobrentendiendo que es el real, salvo que se indique lo contrario ¹⁰. Con esta notación, la *cámara proyectiva* se representa mediante una proyección

$$\pi_C : \mathbb{P}^3 \rightarrow \mathbb{P}^2$$

Los puntos del espacio proyectivo \mathbb{P}^n representan rectas de \mathbb{R}^{n+1} . Esta idea básica se expresa en términos de *coordenadas homogéneas*, es decir, coordenadas definidas salvo factor de proporcionalidad; para $n = 3$ se representa mediante

$$[X_1 : X_2 : X_3 : X_4]^\top = \lambda [X_1 : X_2 : X_3 : X_4]^\top$$

Aunque toda la Geometría Projectiva se puede construir de forma sintética sin utilizar coordenadas (conocido al menos desde Von Staudt), la introducción de coordenadas homogéneas facilita la representación y la gestión de información desde el punto de vista computacional.

Fijadas referencias en los espacios de partida y llegada, la proyección central $\pi_C : \mathbb{P}^3 \rightarrow \mathbb{P}^2$ sobre un plano (en particular, de imagen) desde el punto central \mathbf{C} de la cámara (al que llamamos foco se representa mediante una 3×4 -matriz P_C definida asimismo salvo factor de proporcionalidad).

El *grupo de transformaciones proyectivas* (colineaciones, homografías) está dado por $\mathbb{PGL}(n+1, \mathbb{R}) := GL(n+1, \mathbb{R}) / \mathbb{R}^*$ que actúa por multiplicación a la izquierda con respecto a las $(n+1)$ -coordenadas homogéneas. La acción es libre y transitiva y dota al espacio proyectivo de estructura de espacio homogéneo por la acción del grupo. Por ello, es el grupo estructural del espacio proyectivo n -dimensional \mathbb{P}^n en el sentido del programa de Erlangen (F.Klein, 1873).

¹⁰ El modelo complejo tiene interés para cuestiones relacionadas con Teoría de la Señal y en relación con cuestiones más avanzadas de Reconocimiento que se revisan en el módulo 4 del CEViC

1.2.2. Puntos del infinito

El plano proyectivo permite incorporar las intersecciones de líneas paralelas en el plano ordinario como puntos del infinito y darles el mismo tratamiento que a cualquier otro punto del espacio ambiente. Este hecho tiene profundas implicaciones no sólo desde el punto de vista matemático (como espacio homogéneo), sino computacional, pues permite evitar la casuística asociada al modelo euclídeo o afín y además facilita un tratamiento más compacto del formalismo. Para ilustrar esta idea veamos cómo se calcula la *intersección de dos rectas*:

La intersección de dos líneas $\mathbf{l}_1 = (a_1, b_1, c_1)^\top$ y $\mathbf{l}_2 = (a_2, b_2, c_2)^\top$ está dada por un producto vectorial (menores de tamaño 2×2 de la matriz de coeficientes que reescribimos mediante una matriz antisimétrica de la forma

$$\begin{aligned} \mathbf{x} &= \mathbf{l}_1 \times \mathbf{l}_2 & (2) \\ &= \begin{bmatrix} 0 & -c_1 & b_1 \\ c_1 & 0 & -a_1 \\ -b_1 & a_1 & 0 \end{bmatrix} \begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix} = \widehat{\mathbf{l}}_1 \mathbf{l}_2 & (3) \end{aligned}$$

Las matrices antisimétricas forman un espacio vectorial que juega un papel crucial para un gran número de cuestiones relacionadas con movimientos en el espacio ordinario ¹¹

Ejercicio.- Verifica utilizando el Teorema de Rouché que dos líneas paralelas en el plano proyectivo siempre se cortan. *Indicación:* La intersección de dos líneas paralelas $\mathbf{l}_1 = (a, b, c_1)^\top$ y $\mathbf{l}_2 = (a, b, c_2)^\top$ en el plano proyectivo \mathbb{P}^2 está dada por

$$\mathbf{l}_1 \times \mathbf{l}_2 = (bc_2 - c_1b, ac_1 - ac_2, 0)^\top = (c_1 - c_2)(b, -a, 0)^\top \sim (b, -a, 0)^\top$$

donde \sim representa la equivalencia proyectiva y $(b, -a, 0)^\top$ es el “punto del infinito” en el que se cortan las líneas paralelas.

1.2.3. Dualidad entre puntos y líneas en el plano

Las condiciones de incidencia entre un punto \mathbf{p} y una línea ℓ en el plano cartesiano se expresan en términos algebraicos como

$$0 = ax + by + c = (x, y, 1)(a, b, c)^\top$$

donde $(x, y, 1)^\top$ representa el punto afín asociado al punto \mathbf{p} de coordenadas $(x, y)^\top$ (píxel o cartesianas) y $(a, b, c)^\top$ la línea ℓ . Para $x_3 \neq 0$, haciendo $x = x_1/x_3$ e $y = x_2/x_3$, la ecuación anterior se reescribe como

¹¹ Más adelante veremos que es el álgebra de Lie del grupo especial ortogonal $SO(3; \mathbb{R})$ que parametriza la rotaciones del espacio ordinario

$$\mathbf{a}^T \cdot \mathbf{x} = a_1x_1 + a_2x_2 + a_3x_3 = (x_1, x_2, x_3)(a_1, a_2, a_3)^T = 0$$

donde para facilitar una expresión más compacta hemos hecho $(a, b, c) = (a_1, a_2, a_3)$.

Esta representación da lugar a una *dualidad* entre punto \mathbf{p} y línea ℓ en el plano proyectivo representados por los vectores homogéneos \mathbf{x} y \mathbf{a} , respectivamente. Esta dualidad facilita una formulación sin excepciones para las condiciones de incidencia entre rectas (incluyendo la intersección de rectas paralelas). Así p.e. se tiene

- *Incidencia:* $\mathbf{p} \in \ell \Leftrightarrow \mathbf{p}^T \ell = 0 \Leftrightarrow \ell^T \mathbf{p} = 0$.
- *Intersección:* $\mathbf{p} \in \ell \cap \ell' \Leftrightarrow \mathbf{p} = \ell \times \ell'$ (usar la condición de incidencia $\ell(\ell \times \ell') = \ell'(\ell \times \ell') = 0$).
- *Línea ℓ por dos puntos \mathbf{p} y \mathbf{p}' :* $\ell = \mathbf{p} \times \mathbf{p}'$

La primera representación en términos de $(x, y, 1)$ es asimétrica, pues la tercera coordenada de la representación afín del punto es 1, pero la segunda no tiene excepciones. La consideración de $(n + 1)$ -uplas de coordenadas no todas nulas (no existe ninguna dirección cuyos elementos sean todos nulos) se extiende a dimensión arbitraria.

La representación proyectiva facilita la caracterización de los elementos del infinito como los puntos de intersección de las líneas paralelas que se aborda en el apartado siguiente.

1.2.4. Elementos del infinito

Los elementos del infinito permiten expresar las condiciones de incidencia sin excepciones en el plano proyectivo \mathbf{P}^2 . La idea intuitiva consiste en que *dos rectas paralelas se cortan en un punto del infinito*.

Ejercicio.- Verifica que la representación proyectiva de las rectas del plano cartesiano $\{x = 1\}$ y $\{x = 2\}$ se cortan en el punto $(-1, 0, 1) \times (-1, 0, 2) = (0 : 1 : 0) = \mathbf{p}_{02, \infty}$. Calcular los puntos de intersección de las rectas proyectivas asociadas a dos rectas paralelas al eje Ox o a la diagonal del primer cuadrante en el plano cartesiano.

En realidad, cualquier haz de rectas paralelo a una dirección del espacio cartesiano se corta en un punto del infinito, por lo que cualquier dirección puede representar un punto del infinito en el plano proyectivo.

Asimismo, la recta que pasa por dos puntos del infinito es nuevamente una *recta del infinito* también llamada una *línea del horizonte*. En el caso particular del plano proyectivo \mathbb{P}^2 , denotamos mediante $\ell_{i, \infty}$ a la recta dada por $x_i = 0$ en coordenadas homogéneas a las que se llama “líneas coordenadas”.

De una forma intuitiva, las tres líneas coordenadas $x_i = 0$ para $1 \leq i \leq 3$ en el modelo proyectivo de una imagen se pueden visualizar como líneas del horizonte. Una extensión de este argumento justifica que, en ocasiones, se elija una referencia de puntos de fuga \mathbf{V}_i (vanishing points) a la dada por los puntos

$\mathbf{V}_1 = [1 : 0 : 0]^T$, $\mathbf{V}_2 = [0 : 1 : 0]^T$, $\mathbf{V}_3 = [0 : 0 : 1]^T$, a la que se añade el punto $\mathbf{V}_4 = [1 : 1 : 1]^T$ por razones que se verán más adelante.

Mediante la acción del grupo lineal proyectivo $\mathbb{P}GL(3; \mathbb{R})$ es fácil ver que *cualquier recta* puede ser transformada en una recta del infinito. Por ello, la representación dada por $x_i = 0$ para $1 \leq i \leq 3$ es sólo para fijar ideas; en ningún caso se pretende decir que sólo haya 3 líneas del infinito.

La selección de una línea del plano proyectivo como línea del infinito, permite realizar una *descomposición del plano proyectivo*. Para ello, construimos el *abierto coordenado*

$$D_+(x_i) := \{[x_1 : x_2 : x_3] \in \mathbb{P}^2 \mid x_i \neq 0\} = \left\{ \left[\frac{x_1}{x_i} : \frac{x_2}{x_i} : \frac{x_3}{x_i} \right] \in \mathbb{P}^2 \mid x_i \neq 0 \right\}$$

donde forzosamente alguna de las entradas es igual a 1. Con esta notación es claro que

$$\mathbf{P}^2 = D_+(x_i) \cup \ell_{i,\infty} \simeq \mathbf{A}^2 \cup \mathbf{P}^1$$

De forma análoga, cualquier *punto del infinito* se puede expresar como “punto de fuga” (correspondiente a la intersección de rectas paralelas del espacio cartesiano) o bien como intersección $\ell_\infty \cap \ell'_\infty$ de dos líneas del infinito ℓ_∞ y ℓ'_∞ ; en particular, si se eligen dichas líneas como líneas coordenadas se tiene una “referencia estándar” formada por tres puntos del infinito

$$\mathbf{P}_{ij,\infty} = \ell_{i,\infty} \cap \ell_{j,\infty} = \ell_{i,\infty} \times \ell_{j,\infty}$$

con coordenadas $[1 : 0 : 0]$, $[0 : 1 : 0]$ y $[0 : 0 : 1]$, respectivamente a los que en ocasiones se les denota como puntos de fuga “canónicos” y se les denota mediante \mathbf{V}_i . Nuevamente, éste es un convenio, pues cualquier punto puede ser un punto de fuga y cualquier terna de puntos de fuga l.i. se transforma en la canónica mediante una transformación proyectiva del plano (homografía).

Al complementario \mathbf{A}^2 de una recta proyectiva \mathbb{P}^1 se le llama un *abierto afín* y tiene coordenadas no-homogéneas que dependen de la elección de la recta proyectiva como línea del infinito ℓ_∞ . En Visión Computacional habitualmente se elige $x_3 = 0$ como ℓ_∞ , lo cual da lugar a una descomposición del plano proyectivo como unión disjunta:

$$\mathbf{P}^2 = D^+(x_3) \cup \{x_3 = 0\} \simeq \mathbf{A}^2 \cup \mathbf{P}^1$$

donde $D^+(x_3) := \{\mathbf{p} = [x_1 : x_2 : x_3] \mid x_3 \neq 0\}$ es un abierto “afín”. El apartado siguiente está dedicado a analizar este tipo de conjuntos.

Ejercicio.- Verifica que

$$\mathbf{P}^3 = D^+(x_i) \cup \{x_i = 0\} \simeq \mathbf{A}^3 \cup \mathbf{P}^2 \quad \forall 1 \leq i \leq 4$$

La aplicación que da el cambio de coordenadas sobre $D^+(x_i) \cap D^+(x_j)$ está dada por $x_i x_j^{-1}$ ó su inversa $x_j x_i^{-1}$ que es claramente no-lineal con distorsiones

tanto mayores cuanto más cerca está el punto \mathbf{P} a alguno de los hiperplanos coordenados $x_i = 0$.

Ejercicio.- Verifica que las condiciones de incidencia $\mathbf{p} \in \ell$ y $\ell_1 \cap \ell_2 = \mathbf{p}_{12}$ en el plano proyectivo son invariantes por la acción del grupo proyectivo. Extiende y verifica este argumento para dimensión 3.

1.3. El espacio afín y sus transformaciones

La recta afín ya ha sido introducida anteriormente como una extensión del segmento s_{AB} ¹² que pasa por dos puntos A y B , es decir, como

$$r_{AB} := \{C = (1 - \lambda)A + \lambda B \mid \lambda \in \mathbb{R}\}$$

En estos términos, la recta afín se describe como una combinación lineal de puntos que también se puede describir en términos vectoriales como

$$r_{AB} := \{C = A + \lambda \mathbf{v} \mid \lambda \in \mathbb{R}, \mathbf{v} = B - A\}$$

La segunda descripción contiene el germen de los desarrollos que se presentan en este apartado, pues muestra a la recta afín como el resultado de trasladar un punto A a lo largo de una dirección determinada por el vector $\mathbf{v} = B - A$. Esta idea se generaliza a cualquier dimensión

De una forma intuitiva, el espacio afín n -dimensional \mathbb{A}^n se obtiene como “traslación” del conjunto de puntos del espacio por la acción de cualquier vector libre. Por ello, localmente tiene la misma estructura que el espacio cartesiano, pero sus elementos no son vectores libres (con origen en el punto de coordenadas $\mathbf{0}$).

El espacio afín proporciona una representación en términos de *puntos* (en lugar de vectores) del espacio ordinario y de referencias cambiantes que se adapta fácilmente a representaciones paramétricas y permite resolver muy rápidamente operaciones booleanas sencillas o problemas de discretización. Por ello, facilita el transvase de información entre las imágenes digitales (modelos cartesianos para coordenadas píxel) y los modelos proyectivos utilizados para transformar imágenes para generar nuevos puntos de vista o “pegar” datos procedentes de diferentes localizaciones para cámaras \mathbf{C}_i .

Formalmente, un espacio afín n -dimensional es un conjunto \mathbb{A}^n dado por un espacio vectorial n -dimensional V sobre un cuerpo \mathbb{K} y una acción de grupo libre y transitiva de V sobre \mathbb{A}^n dada por traslaciones¹³

De una forma menos abstracta, un espacio afín es un conjunto \mathbb{A}^n y una ley de composición externa

$$\ell : V \times \mathbb{A}^n \rightarrow \mathbb{A}^n \quad | \quad (\mathbf{v}, a) \mapsto \mathbf{v} + a$$

verificando las propiedades siguientes:

¹² Recordemos que $s_{AB} := \{C = (1 - \lambda)A + \lambda B \mid \lambda \in [0, 1]\}$

¹³ En términos abstractos (extensión del programa de Erlangen de Klein) un espacio afín es un espacio principal homogéneo para la acción de V dada por traslaciones.

- *Elemento neutro a la izquierda:* $\forall a \in \mathbb{A}^n \mathbf{0} + a = a$
- *Asociativa:* $\mathbf{v} + (\mathbf{w} + a) = (\mathbf{v} + \mathbf{w}) + a$ para $\forall \mathbf{v}, \mathbf{w} \in V, \forall a \in \mathbb{A}^n$.
- *Unicidad:* La aplicación $V \rightarrow \mathbb{A}^n : \mathbf{v} \mapsto \mathbf{v} + a$ es biyectiva $\forall a \in \mathbb{A}^n$.

Esta subsección comienza con una descripción coordenada del espacio afín e introduce los efectos asociados a diferentes localizaciones con respecto a distintos sistemas de referencia; a continuación se introduce un marco basado en homogeneización y deshomogeneización para comparar datos correspondientes a diferentes vistas. Por último se abordan algunas cuestiones elementales relacionadas con las transformaciones afines que son clave para comparar imágenes tomadas desde localizaciones próximas.

A diferencia del caso euclídeo superpuesto al modelo cartesiano, en el espacio afín no existe un origen de coordenadas. Toda la información se expresa en términos de operaciones formales con puntos. En cierto sentido, se puede decir que el modelado afín es el más robusto y acorde con la percepción humana. No obstante, hay que tener presente que la propia representación geométrica puede inducir a interpretaciones erróneas, lo cual requiere una revisión a posteriori de los modelos representados. Este hecho aparece ya en representaciones de mediados del s.XIV (ver figura 5), pero también pueden responder a un propósito expreso de ilustrar geometrías hiperbólicas como las de los grabados de Escher.

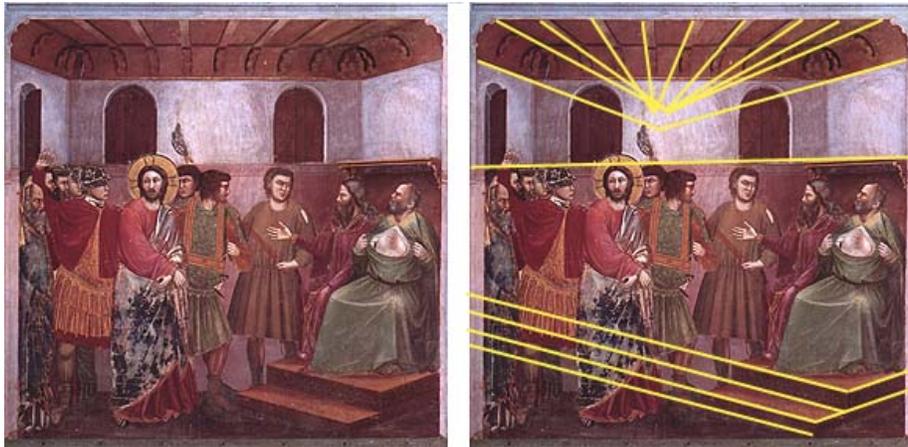


Fig. 5: Ejemplo de inconsistencia en los representaciones globales creados por Lorenzetti (1338)

1.3.1. Descripción coordenada del espacio afín

Una referencia afín \mathcal{R}_a de \mathbb{A}^n está dada por una colección de $n + 1$ "puntos independientes"; la noción de recta afín se ha presentado en la sección 4 del

capítulo 1 del módulo 1. Por ello, aquí nos limitamos a recordar la de plano afín y la de espacio afín tridimensional.

1. Un *plano afín* \mathbb{A}^2 tiene como referencia afín \mathcal{R}_a cualquier terna de puntos no-alineados $\mathbf{A}_0, \mathbf{A}_1, \mathbf{A}_2$. Cualquier punto $\mathbf{A} \in \mathbb{A}^2$ se expresa como una única combinación lineal $\mathbf{A} = \lambda_0\mathbf{A}_0 + \lambda_1\mathbf{A}_1 + \lambda_2\mathbf{A}_2$ con $\lambda_i \in \mathbb{R}$ y $\sum_{i=0}^2 \lambda_i = 1$. Se dice entonces que $(\lambda_0, \lambda_1, \lambda_2)$ son las coordenadas afines o baricéntricas de \mathbf{A} con respecto a la referencia afín $\mathbf{A}_0, \mathbf{A}_1, \mathbf{A}_2$. En particular, los puntos de la referencia tienen como coordenadas afines $(1, 0, 0)$ para \mathbf{A}_0 , $(0, 1, 0)$ para \mathbf{A}_1 y $(0, 0, 1)$ para \mathbf{A}_2 .
2. Un *espacio afín* \mathbb{A}^3 tiene como referencia afín \mathcal{R}_a cualquier 4-upla de puntos $\mathbf{A}_0, \mathbf{A}_1, \mathbf{A}_2$ y \mathbf{A}_3 no-coplanarios, siendo no-alineados cualquier terna. Si conectamos tres de ellos con el cuarto se obtienen direcciones linealmente independientes que permiten generar 3 planos de referencia (combinándolas dos a dos); proyectando de forma paralela a cada una de estas 3 direcciones se obtienen coordenadas. Alternativamente, cualquier punto $\mathbf{A} \in \mathbb{A}^3$ se expresa como una única combinación lineal $\mathbf{A} = \lambda_0\mathbf{A}_0 + \lambda_1\mathbf{A}_1 + \lambda_2\mathbf{A}_2 + \lambda_3\mathbf{A}_3$ con $\lambda_i \in \mathbb{R}$ y $\sum_{i=0}^3 \lambda_i = 1$. Se dice entonces que $(\lambda_0, \lambda_1, \lambda_2, \lambda_3)$ son las coordenadas afines o baricéntricas de \mathbf{A} con respecto a la referencia afín $\mathbf{A}_0, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$. En particular, los puntos de la referencia tienen como coordenadas afines $(1, 0, 0, 0)$ para \mathbf{A}_0 , $(0, 1, 0, 0)$ para \mathbf{A}_1 , $(0, 0, 1, 0)$ para \mathbf{A}_2 y $(0, 0, 0, 1)$ para \mathbf{A}_3 .

El caso correspondiente al espacio afín \mathbb{A}^n se deja como ejercicio.

La relación con el enfoque vectorial (utilizado también en el espacio cartesiano habitual \mathbb{R}^n) es clara: Si $R_{\mathbb{A}^n} = \{\mathbf{A}_0, \mathbf{A}_1, \dots, \mathbf{A}_n\}$ es una referencia afín de \mathbb{A}^n , haciendo $\mathbf{v}_i = \mathbf{A}_i - \mathbf{A}_0$, se obtiene una colección de n vectores linealmente independientes, es decir, una base del espacio vectorial subyacente a \mathbb{R}^n . Recíprocamente, dada una base arbitraria, añadiendo el origen a la colección de los afijos de los n vectores de la base se obtiene una referencia afín $\mathcal{R}_{\mathbb{A}^n}$ a la que, en ocasiones se representa más abreviadamente como \mathcal{R}_a^n (el subíndice a corresponde a afín y el superíndice a la dimensión del espacio).

Una *ventaja* del enfoque afín con respecto al cartesiano consiste en que no hay ningún punto “privilegiado” (como origen de coordenadas); por ello el mantenimiento de la coherencia de la información tiene un menor coste computacional para el caso afín que para el cartesiano. No obstante, es necesario controlar la posible propagación de errores en cada paso ó cada vez que se actualiza una referencia, garantizando que la actualización de las $n + 1$ -uplas dan lugar a referencias afines en cada paso ¹⁴. Como *conclusión* cualquier referencia minimal de \mathbb{A}^n está dada por $n + 1$ puntos asociados a n vectores l.i.; asimismo, es conveniente maximizar los ángulos entre las direcciones correspondientes a los vectores.

¹⁴ Obviamente, la matriz que da el cambio de referencia debe ser regular, pero además el valor del determinante debe estar por encima de un umbral para minimizar problemas de dependencia entre puntos que puedan dar lugar a degeneraciones del sistema

La relación entre un espacio afín y el espacio vectorial subyacente permite trasladar de forma automática todo el cálculo vectorial al espacio afín. Sintetizamos los aspectos a utilizar en las afirmaciones siguientes:

- Cada punto \mathbf{P} de un espacio afín real $\mathbb{A}_{\mathbb{R}}^n$ tiene un entorno *globalmente equivalente* a \mathbb{R}^n de forma canónica ¹⁵
- La colección de puntos $\{\mathbf{A}_0, \dots, \mathbf{A}_n\}$ es una referencia afín $\mathcal{R}_{\mathbb{A}^n}$ si $\{\mathbf{v}_i = \mathbf{A}_i - \mathbf{A}_0 \mid 1 \leq i \leq n\}$ son vectores l.i.
- Dada una referencia minimal $\mathbf{A}_0, \dots, \mathbf{A}_n$ cada punto se expresa de forma única como $\mathbf{P} = \sum_{i=0}^n \lambda_i \mathbf{A}_i$
- Los cocientes $\lambda_i / (\sum_{i=0}^n \lambda_i)$ se llaman *coordenadas baricéntricas*. ¹⁶
- La referencia dual \mathcal{R}_a^ν de $\mathcal{R}_a = \{\mathbf{A}_0, \dots, \mathbf{A}_q\}$ está dada por $\{< \mathbf{A}_0, \dots, \hat{\mathbf{A}}_i, \dots, \mathbf{A}_q \mid 0 \leq i \leq q\}$ donde cada uno de los elementos representa el hiperplano que resulta de suprimir el punto \mathbf{A}_i de las combinaciones lineales; es decir, se trata de puntos caracterizados por la condición $\lambda_i = 0$ que juegan un papel similar a los “hiperplanos coordenados” para el caso afín.

La descripción afín proporciona un soporte fácilmente implementable para la *posición relativa* de un punto con respecto a referencias afines, que resultan de utilidad para la Robótica asistida por Visión; este modelo es de hecho las utilizadas por los sistemas de percepción humana en las tareas cotidianas que no requieran precisión milimétrica.

Ejercicio 1.- Dada una referencia afín en el plano \mathbb{A}^2 , identifica la posición relativa de un punto \mathbf{A} en relación con las rectas $< \mathbf{A}_i, \mathbf{A}_{i+1} >$ para $i = 0, 1, 2(mod, 3)$ en términos de los signos de las coordenadas baricéntricas. (*Indicación:* resolver previamente el caso correspondiente a la recta afín utilizando la referencia dual). Extiende la construcción efectuada al caso del espacio \mathbb{A}^3 , reemplazando las rectas $< \mathbf{A}_i, \mathbf{A}_{i+1} >$ por los planos de referencia $< \mathbf{A}_i, \mathbf{A}_{i+1}, \mathbf{A}_{i+2} >$ para $i = 0, 1, 2, 3(mod, 4)$.

Ejercicio 2.- Describe un procedimiento para controlar el *movimiento relativo de cámara* mediante el paso por líneas de referencia $\lambda_i = 0$ asociadas a referencias triangulares en una representación planar de la escena que se actualiza con el movimiento de la cámara. ¿Es aplicable este método a mapas de elevación digital (DEM) para el terreno? Diseña un algoritmo para la actualización de la información a partir de un mapa de alturas y otro de profundidades generados de forma sintética.

Una cuestión importante para delimitar / seguir objetos o acotas escenas es el cálculo de envolventes lineales. Hay varios tipos de envolventes por primitivas

¹⁵ El resultado no depende del cuerpo real \mathbb{R} y sigue siendo válido para los complejos o cualquier otro cuerpo \mathbb{K} . La adaptación a coeficientes racionales interesa para el enfoque discreto adoptado en esta materia.

¹⁶ Esta descripción es equivalente a la dada más arriba para espacios afines de baja dimensión presentada en la última sección del capítulo 1 del módulo 1. Verificalo como ejercicio

geométricas lineales a trozos (PL-primitives donde PL corresponde a Piecewise Linear):

- *Envolvente convexa* $\mathcal{H}(\mathbf{A}_0, \dots, \mathbf{A}_r) := \{\mathbf{P} = \sum_{i=0}^r \lambda_i \mathbf{A}_i \mid \sum_{i=0}^r \lambda_i = 1, 0 \leq \lambda_i \leq 1 \forall i\}$
- *Envolvente afin* $\mathcal{A}(\mathbf{A}_0, \dots, \mathbf{A}_r) := \{\mathbf{P} = \sum_{i=0}^r \lambda_i \mathbf{A}_i \mid \sum_{i=0}^r \lambda_i = 1\}$
- *Envolvente visual*: es la formada por el cono proyectante $\mathcal{C}_{\mathbf{C}} S^{(\alpha)}$ que tiene por vértice el foco de la cámara \mathbf{C} y por directriz la silueta aproximada $S^{(\alpha)}$ o PL-borde de un objeto $B^{(\alpha)}$, eventualmente en movimiento.

La estimación de una silueta eventualmente cambiante puede tener un elevado coste computacional; cuando se desea obtener resultados en tiempo real, es conveniente simplificar la forma reemplazándola por algún tipo de envolventes; además de las lineales, frecuentemente se recurre a primitivas cuadráticas a trozos (PQ-primitives donde PQ corresponde a Piecewise Quadratic). Los momentos de orden 2 proporcionan herramientas para estimar la forma de las elipses o elipsoides que mejor se ajustan a los datos obtenidos.

Ejercicio 3.- Realiza una implementación para cada una de las envolventes lineales mencionadas usando datos generados de forma sintética.

Nota.- En la práctica no es elemental generar siluetas de forma exacta a partir de datos contenidos en imágenes. Requiere técnicas de segmentación avanzadas que se han esbozado en el módulo 1. La silueta aproximada $S^{(\alpha)}$ está dada inicialmente por una poligonal simple situada en un plano ortogonal a la línea de visión (a una profundidad que se obtiene independientemente) y se obtiene por “recorte” usando propiedades radiométricas o bien información de rango obtenido a partir de sensores adicionales (infrarrojos, p.e.) o de las discontinuidades asociadas a mapas de profundidad.

El enfoque basado en información de rango se desarrolla sobre todo en el módulo 5 del CEViC en el que se presentan elementos de Visión Estéreo Dinámica y sus aplicaciones al Vídeo 3D; este tipo de cuestiones está relacionado con la aproximación volumétrica a objetos en movimiento obtenida mediante intersecciones de conos proyectantes y técnicas de “excavación” que utilizan marching cubes para generar volúmenes mejor ajustados a la intersección de conos proyectivos. Es importante avanzar en una implementación computacional eficiente de estos tópicos en relación con la Reconstrucción 3D de objetos en movimiento.

Homogeneización y deshomogeneización

La deshomogeneización facilita el paso del proyectivo al afin a partir de la identificación de un hiperplano del infinito:

1. Supongamos que $\mathbb{P}^n \simeq \mathbb{A}^n \cup \mathbb{P}^{n-1}$ con H_∞ dado por $x_0 = 0$
2. *Deshomogeneización*: $\mathbf{P}^n \rightarrow \mathbf{A}^n \mid [x_0 : x_1 : \dots : x_n] \mapsto (1, \frac{x_1}{x_0}, \dots, \frac{x_n}{x_0})$

3. *Homogeneización:* $\mathbf{A}^n \rightarrow \mathbf{P}^n \mid (1, \frac{x_1}{x_0}, \dots, \frac{x_n}{x_0}) \mapsto [x_0 : x_1 : \dots : x_n]$
4. *Pegado* de datos locales sobre $D^+(x_i) \cap D^+(x_j)$ con $D^+(x_i) = \{\underline{x} \in \mathbb{P}^n \mid x_i \neq 0\}$
5. *Distorsiones* tanto mayores cuanto más cerca de cada hiperplano del infinito $x_i = 0$

La composición de ambos procesos es la clave para el pegado de copias afines y, por tanto, para la construcción de modelos proyectivos globales (que puedan ser visualizados desde cualquier localización de cámaras virtuales) a partir de datos afines locales.

Ejercicio.- Muestra el cambio de carta entre dos abiertos coordenados $U_i = D^+(x_i)$ e implementa la transformación sobre la intersección $U_{ij} := U_i \cap U_j$. Evalúa la distorsión en las proximidades de $x_i = 0$ o de $x_j = 0$ (*Indicación:* Discretiza en las proximidades de ambas rectas)

1.3.2. Transformaciones afines

Definición.- A es una transformación afín si deja invariante un plano del infinito H_∞ , p.e. $x_{n+1} = 0$

Las *transformaciones afines* de \mathbb{A}^n generan el *grupo afín*

Definición.- El grupo grupo afín es el producto semidirecto de $GL(n; \mathbb{R}) := \{A \in M(n \times n; \mathbb{R}) \mid \det(A) \neq 0\}$ y el grupo de traslaciones \mathbb{R}^n . Por ello, las transformaciones afines están representadas por $(n+1) \times (n+1)$ -matrices de la forma

$$M = \begin{pmatrix} A & \mathbf{v} \\ 0 & 1 \end{pmatrix}, \quad A \in GL(n; \mathbb{K}), \quad \mathbf{v} \in \mathbb{R}^n$$

que actúan a la izquierda sobre los puntos $\mathbf{p} = (x_1, \dots, x_n, 1)^\top \in \mathbb{A}^n$ mediante

$$M\mathbf{p} := A\underline{x} + \mathbf{v}$$

Ejercicio.- Verificar que cualquier transformación afín conserva las relaciones de paralelismo para rectas que no pasan por el centro de proyección.

1.3.3. El grupo especial lineal

Un subgrupo especialmente importante es el producto del *grupo especial lineal* $SL(n; \mathbb{K})$ y del grupo de traslaciones \mathbb{K}^n donde

$$SL(n; \mathbb{K}) := \{A \in GL(n; \mathbb{K}) \mid \det(A) = 1\}$$

es el grupo de las matrices regulares que conservan el volumen orientado. El grupo $SL(n; \mathbb{K})$ tiene estructura de variedad y, por ello, se puede calcular el espacio tangente que es un álgebra de Lie (espacio vectorial con producto interno) dado por

$$sl(n; \mathbb{K}) := T_I SL(n; \mathbb{K}) = \{X \in M(n \times n; \mathbb{K}) \mid tr(X) = 0\}$$

Para $n = 2$ se tiene que $sl(2; \mathbb{K})$ es un espacio vectorial tridimensional generado por

$$\mathbf{e}_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{e}_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{r} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Los generadores $\{\mathbf{e}_+, \mathbf{e}_-, \mathbf{r}\}$ forman una $sl(2)$ -terna y tienen un gran interés para multitud de aplicaciones; proporcionan los “ladrillos básicos” para construir “representaciones” de objetos matriciales que no se pueden diagonalizar (que son la mayoría). De ahí su carácter ubicuo en la linealización de problemas complejos; aparecen en problemas tan diferentes como el análisis de deformaciones, la estimación del movimiento o la compresión de imagen. Por ello, reaparecen en distintos módulos del CEViC.

Desde el punto de vista matemático, cada una de las transformaciones que generan el álgebra de Lie de $SL(2)$ pueden ser interpretadas como un “vector velocidad” que indican una dirección de “desplazamiento infinitesimal”. La transformación original (en el espacio ambiente) se obtiene como la exponencial de dicha transformación infinitesimal X , es decir, como

$$\exp(tX) = I + tX + \frac{t^2}{2!}X^2 + \frac{t^3}{3!}X^3 + \dots + \frac{t^n}{n!}X^n + o(n+1)$$

En particular, como $\mathbf{e}_+^2 = \mathbf{e}_-^2 = 0$, se tiene que su exponencial es un “segmento” en el grupo de matrices que expresa una distorsión direccional. Así por ejemplo

$$\exp(t\mathbf{e}_+)(x, y)^\top = [I + t\mathbf{e}_+(x, y)^\top] = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} (x, y)^\top = (x + ty, y)^\top$$

Análogamente,

$$\exp(t\mathbf{e}_-)(x, y)^\top = [I + t\mathbf{e}_-(x, y)^\top] = \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix} (x, y)^\top = (x, tx + y)^\top$$

que se pueden aplicar para representar una distorsión de un rectángulo relacionada con efectos de perspectiva o de cambios de orientación ¹⁷.

Ejercicio.- Implementar las transformaciones anteriores y mostrar los efectos debidos a las distorsiones de dichas transformaciones sobre los vértices de un rectángulo. Esta descripción es de gran importancia para:

- actualizar en tiempo real de los mapas de perspectiva dependiendo de la localización relativa de la cámara (ver más abajo);

¹⁷ En el capítulo 2 se desarrolla un análisis más sistemático de los diferentes tipos de distorsión. De una forma simplificada, la distorsión puede ser centrada (radial o simétrica) ó por el contrario, descentrada

- estimar las distorsiones para objetos móviles pues permite desacoplar el movimiento de objetos rígidos en componentes muy sencillas de estimar (ver el módulo 3 de CEViC).

1.4. El espacio euclídeo y sus transformaciones

En esta subsección se introduce elementos básicos de la geometría euclídea -basada en la conservación de la distancia euclídea- que resultan de utilidad para la Reconstrucción 3D en el marco proyectivo. En el marco euclídeo, la conservación de la distancia implica la conservación de ángulos (Geometría Conforme) y la conservación de áreas (Geometría Especial); el recíproco no es cierto *En el módulo 5 se desarrollan una aproximación al vídeo 3D basada en la casi-conservación de ángulos; el control de una variación acotada corresponde a la Geometría Conforme.*

En este marco la conservación de la cónica absoluta o la estimación de su distorsión en diferentes planos de proyección (diferentes del correspondiente a la proyección ortogonal) juega un papel fundamental para la recuperación de aspectos métricos de objetos industriales o escenarios arquitectónicos, p.e.. El grupo estructural de transformaciones es el grupo especial ortogonal caracterizado por la conservación de la matriz identidad que representa la métrica euclídea.

La extensión del enfoque presentado inicialmente objetos volumétricos curvados requiere una aproximación lineal a trozos (PL: Piecewise Linear) que se aborda en el capítulo 6 de este módulo. Para objetos curvados se reemplaza la conservación de la métrica euclídea por la conservación de la imagen de dicha métrica por una aplicación suave a trozos. Por ello, el enfoque presentado en este apartado se extiende de forma natural a objetos curvados eventualmente deformables, extendiendo el ámbito de la aplicación a situaciones en las que se requiera una precisión elevada.

1.4.1. Descripción del espacio euclídeo

Desde el punto de vista conjuntista, el espacio euclídeo \mathbb{E}^n está dado por el par (\mathbb{R}^n, ds^2) donde ds^2 es la distancia euclídea: $d(\mathbf{P}, \mathbf{Q}) = [\sum_{i=1}^n (q_i - p_i)^2]^{1/2}$

Si se utiliza la estructura subyacente de espacio vectorial asociada al espacio cartesiano \mathbb{R}^n , conectando el afijo de cada vector libre con el origen se obtiene una representación vectorial $v_{\mathbf{P}} = \overline{\mathbf{OP}}$ vector libre asociado a \mathbf{P} a la que se aplica el Álgebra Lineal de la forma usual.

La existencia de una métrica euclídea permite construir entornos e introducir criterios de proximidad. Los procedimientos de búsqueda y localización se representan mediante mapas de distancias o de curvas de nivel; un ejemplo típico corresponde a los puntos que se encuentran situados a la misma distancia o que presentan el mismo valor para una función definida sobre las coordenadas locales (la altura o la profundidad, p.e.).

Ejercicio.- Realiza una implementación computacional de los procedimientos de búsqueda mediante procedimientos de barrido (paralelo o rotacional, típicamente). *Indicación:* Utiliza procedimientos de búsqueda recursiva incrementando el valor de la función distancia discreta sobre el plano de imagen ¹⁸. La implementación de L^1 es más rápida y eficiente para procedimientos de búsqueda y se adapta mejor a la naturaleza discreta de la imagen y, por consiguiente, de los entornos a inspeccionar.

1.4.2. Transformaciones euclídeas

Las transformaciones euclídeas están caracterizadas por conservar la distancia euclídea. El grupo de las *transformaciones euclídeas* es un subgrupo del grupo proyectivo que se define como el producto semidirecto del grupo de las transformaciones ortonormales $SO(n; \mathbb{R})$ por las traslaciones \mathbb{R}^n . El grupo especial ortogonal $SO(n; \mathbb{R})$ es la intersección del grupo especial lineal $SL(n; \mathbb{R})$ y del grupo ortogonal $O(n; \mathbb{R})$ donde

- Grupo Especial Lineal real $SL(n; \mathbb{R}) := \{A \in GL(n; \mathbb{R}) \mid \det(A) = 1\}$: conservan el volumen
- Grupo Ortogonal $O(n; \mathbb{R}) := \{A \in GL(n; \mathbb{R}) \mid A^\top \cdot A = I\}$: conservan las relaciones de ortogonalidad

En otras palabras el grupo $SO(n; \mathbb{R})$ está formado por las $(n+1) \times (n+1)$ -matrices de la forma

$$M = \begin{pmatrix} A & \mathbf{v} \\ 0 & 1 \end{pmatrix}, \quad A \in SO(n), \quad \mathbf{v} \in \mathbb{R}^n$$

Estas matrices actúan a la izquierda sobre los puntos $\mathbf{p} = (x_1, \dots, x_n, 1)^\top \in \mathbb{A}^n$ mediante

$$M\mathbf{p} := A\mathbf{x} + \mathbf{v}$$

1.4.3. Conservación de la métrica y cónica absoluta

Cualquier métrica se representa como una forma cuadrática no-degenerada \mathbf{Q} ; en el caso euclídeo la forma cuadrática es definida positiva.

Nota histórica.- Isaac Newton demostró que cualquier par de circunferencias C_1 y C_2 se cortan en a lo sumo dos puntos reales; sobre el cuerpo de los números complejos \mathbb{C} el número de soluciones complejas debe ser 4. Por tanto, además de las soluciones reales (si las hay), debe haber al menos otras dos soluciones complejas. Dos de estas soluciones son los *puntos circulares del infinito* $(1 : \pm i)$

¹⁸ En lugar de la distancia euclídea, etiquetada como L^2 , en otras ocasiones se utiliza la métrica L^1 ó, con más generalidad, L^k para algún valor natural $1 \leq k \leq \infty$; aunque la apariencia del entorno de cada punto cambia, le topología es la misma

donde $i = \sqrt{-1}$ que fueron denotados mediante I, J por I.Newton. Estos dos puntos representan la “cónica del infinito” para la recta proyectiva $\mathbb{P}^1_{\mathbb{C}}$. En efecto, si resolvemos $x_0^2 + x_1^2 = 0$ se obtiene $(x_0/x_1)^2 = -1$, es decir, $x_0/x_1 = \pm i$. En particular, cualquier transformación euclídea de la recta deja invariante los puntos circulares del infinito. Nótese que

- ${}^{\top}\mathbf{I}\mathbf{J} = {}^{\top}\mathbf{J}\mathbf{I} = 0$, es decir, los puntos \mathbf{I} y \mathbf{J} de $\mathbb{P}^1 := \mathbb{S}^1 / \sim_a$ corresponden a direcciones ortogonales del plano ordinario.
- $\mathbf{I} \cdot {}^{\top}\mathbf{J} = I_2$ que representa la matriz de los puntos circulares (de hecho de la cónica absoluta) sobre \mathbb{P}^1
- Los puntos circulares representan el soporte de una forma cuadrática no-degenerada sobre \mathbb{P}^1

Más de medio siglo más tarde, Laguerre demostró que cualquier transformación euclídea del plano proyectivo \mathbb{P}^2 deja invariante la *cónica absoluta*

$$q_{\infty} := \{\underline{x} \in \mathbb{P}^2 \mid x_0^2 + x_1^2 + x_2^2 = 0\}$$

y recíprocamente. Nótese que cualquier punto de q_{∞} es nuevamente imaginario (la cónica no es observable, aunque sí se pueda calcular). Dicha cónica se puede interpretar asimismo como una sección plana de la *cuádrica absoluta*

$$Q_{\infty} := \{\underline{x} \in \mathbb{P}^3 \mid x_0^2 + x_1^2 + x_2^2 + x_3^2 = 0\}$$

por el hiperplano del infinito $H_{\infty} := \{\underline{x} \in \mathbb{P}^3 \mid x_3 = 0\}$, es decir,

$$q_{\infty} = Q_{\infty} \cap H_{\infty}$$

Ejercicio (Laguerre).- Demostrar que la cónica absoluta q_{∞} se conserva por el grupo de semejanza generado por las transformaciones euclídeas salvo escala (*Indicación:* Verificar que la matriz diagonal se transforma en otra diagonal y re-escalar).

Nota.- A efectos de la Visión Computacional este resultado de la Geometría Projectiva clásica (principios del siglo XIX) implica que si realizamos desplazamientos rígidos (traslaciones y rotaciones) salvo escala con la cámara, la cónica absoluta q_{∞} se conserva con tal que no modifiquemos los parámetros internos de la cámara (salvo efectos de zoom). La estimación de los parámetros internos se aborda en el capítulo 2 de este módulo.

El problema de la estimación.- Si estimamos la cónica absoluta tendremos una restricción estructural para las transformaciones euclídeas. La forma canónica de la ecuación $x_0^2 + x_1^2 + x_2^2 = 0$ indica que, para una elección apropiada de coordenadas, la cónica absoluta es una circunferencia; como ya conocemos dos puntos $(1 : \pm i : 0)$, bastaría identificar un tercero.

Lamentablemente, en la práctica las cosas no son tan fáciles: Para una estimación precisa deberíamos saber que el plano (de hecho H_{∞}) sobre el que se

está realizando la estimación es “perfectamente ortogonal” a la línea de visión (paralelo al plano de la cámara) torsión y que no presenta distorsión alguna. En la realidad, lo que se puede llegar a estimar es una distorsión de la cónica absoluta (proyección de la cónica absoluta sobre el plano del infinito asociado a la captura actual). En consecuencia es necesario estimar las cónicas que pueden aparecer y sus posibles degeneraciones en relación con los problemas de optimización asociados (que requieren un paso al espacio tangente). Por ello, lo que se está estimando es una cónica como lugar de puntos y (en el proceso de optimización) una cónica como lugar envolvente de las tangentes; al par formado por una cónica lugar y su cónica envolvente se le llama una *cónica completa* (incluye las degeneraciones posibles).

De una forma intuitiva, si se desea estimar la distorsión de una circunferencia $q_\infty \subset H_\infty$ en una elipse $q \subset \Pi$ podemos recurrir al comportamiento de las tangentes en relación con el plano del infinito. Ello lleva a evaluar los ángulos que forman dichas tangentes y, en particular, al problema de la estimación de los ángulos en el proyectivo.

1.4.4. Conservación de ángulos

Aplicando argumentos bien conocidos de la trigonometría elemental (especificar cuáles), se obtiene que la conservación de distancias implica la conservación de ángulos. Sin embargo, el recíproco no es cierto, lo cual da lugar a una distinción importante entre la Geometría Euclídea (conservación de distancias) y la Geometría Conforme (conservación de ángulos) que se aborda con técnicas más avanzadas relacionadas con el Análisis Complejo y la Geometría Diferencial. Las herramientas basadas en Geometría Conforme son de gran utilidad para un modelado adaptativo de formas volumétricas complicadas, dentro del marco de la Visión Estéreo Dinámica para objetos deformables (actores reales y su conversión en actores virtuales, p.e.)

En el marco fotogramétrico tradicional, los antiguos teodolitos proporcionan una información muy precisa sobre ángulos; una estación total debidamente georeferenciada permite automatizar el proceso de captura para objetos en un rango medio o lejano. Si se realiza una triangulación conectando las ternas de puntos más próximos, la información sobre ángulos de cada triángulo es equivalente salvo escala en el plano a la información sobre distancias gracias al teorema de los senos:

$$\frac{\text{sen } \alpha}{a} = \frac{\text{sen } \beta}{b} = \frac{\text{sen } \gamma}{c}$$

La disponibilidad de 4-uplas de puntos próximos no-coplanarios, permite recuperar propiedades volumétricas de la escena.

Sin embargo, en general *conservación de ángulos* no implica la conservación de las distancias. El grupo de las transformaciones que conservan los ángulos es el *Grupo Conforme*. Dos casos importantes son:

- El *Grupo Ortogonal Conforme* $CO(V, Q)$ asociado a una forma cuadrática

Q definida sobre un espacio vectorial V caracterizado por las transformaciones lineales T tales que

$$\exists \lambda \neq 0 \quad | \quad Q(T\mathbf{v}) = \lambda^2 Q(Q(Tx) = \lambda^2 Q(x)) \quad \forall \mathbf{v} \in V$$

Es el producto de transformaciones ortogonales por el grupo de las dilataciones.

- El *Grupo Conforme de la Esfera* \mathbb{S}^2 o *grupo de Moebius*¹⁹. Este grupo está generado por las inversiones generalizadas en los círculos. En particular, el plano cartesiano \mathbb{R}^2 soporta una estructura compleja en la que representamos cada punto (x, y) como un elemento de la recta compleja \mathbb{C}^1 coordenadas complejas $z = x + iy$ es la más compacta y sencilla de expresar; está dada localmente por

$$a(z) := \frac{az + b}{cz + d} \quad ad - bc \neq 0$$

Nótese que las traslaciones complejas están representadas por $Az = z + b$, las dilataciones por $A(z) = az$ y las inversiones por $A(z) = 1/z$. Las tres conservan los ángulos formados por dos rectas²⁰. Por ello, la expresión general es una composición de los tres tipos de transformaciones básicas.

El grupo especial ortogonal (caracterizado por la conservación de la métrica euclídea es un subgrupo del grupo conforme asociado a la métrica definida por $z \cdot \bar{z}$ donde \bar{z} representa el conjugado de z (esta descripción es válida para cualquier dimensión).

El grupo conforme se utiliza de forma creciente para representar deformaciones continuas de mallas cuadrangulares. Actualmente, se aplica para modelado 3D tanto interactivo como semi-automático de objetos que presentan gran complejidad geométrica (caras, cuerpos, etc). Este tipo de aplicaciones se desarrollan en el módulo 5 (Visión Estéreo Dinámica) en relación con aplicaciones a la industria de contenidos multimedia.

Ejercicio.- Implementar computacionalmente el cálculo de la esfera que pasa por 4 puntos no coplanares (*Indicación* el centro de la esfera es la intersección de las normales a dos triángulos espaciales generados por 2 de las 4 ternas; alternativamente: los planos mediatrices de [al menos tres] segmentos que conectan cada par de puntos se cortan en el centro de la esfera).

1.5. Operaciones sobre espacios proyectivos

Las operaciones básicas para representaciones de tipo proyectivo corresponden a cortar mediante subespacios lineales y proyectar sobre espacios de dimensión más baja (el plano de la cámara ó una línea, p.e.). Otra operación

¹⁹ La geometría de Moebius es la geometría del espacio euclídeo al que se “añade” un punto del infinito, construcción a la que se denomina *compactificación de Alexandrov*; de forma conjuntista la compactificación del espacio cartesiano por un punto da como resultado la esfera

²⁰ Comprobadlo como ejercicio

importantes concierne al producto de dos ó más copias de elementos proyectivos lineales que se describen en términos del producto de Segre y que permite construir superficies sencillas y fácilmente parametrizables ²¹.

Por último, para identificar la forma más apropiada correspondiente a objetos de grado fijo, es necesario identificar el espacio ambiente de dichas formas que está dado por la imagen del llamado embebimiento de Veronese; la imagen de esta aplicación permite describir el conjunto de las cónicas, cúbicas ó cuárticas como espacios proyectivos ²²

1.5.1. Secciones y proyecciones

1.6. Cónicas del plano proyectivo

En la subsección anterior se ha puesto de manifiesto la necesidad de realizar una estimación de cónicas, cuestión que abordamos ahora de forma directa (cónica como lugar de puntos) e indirecta (cónica dual como envolvente de sus tangentes). Para ello, se revisa inicialmente una descripción básica de las cónicas afines y proyectivas

1.6.1. Descripción de cónicas afines y proyectivas

En este apartado se presentan las versiones analítica y sintética de las cónicas en el plano proyectivo \mathbb{P}^2

Una *Cónica afín* en el abierto coordenado $D_+(x_0) := \{\mathbf{p} = (x_0 : x_1 : x_2) \in \mathbf{P}^2 \mid x_0 \neq 0\}$ está dada por

$$ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0 = (x, y, 1)C(x, y, 1)^\top$$

Una *Cónica proyectiva* se obtiene homogeneizando la ecuación anterior, es decir, haciendo $x = \frac{x_1}{x_0}$, $y = \frac{x_2}{x_0}$ se llega a la expresión

$$ax_1^2 + 2bx_1x_2 + cx_2^2 + 2dx_0x_1 + 2ex_0x_2 + fx_0^2 = 0$$

Es frecuente recurrir a una *expresión sintética* $\mathbf{p}^\top C \mathbf{p} = 0$ donde C es una (3×3) -matriz *simétrica* arbitraria. En este contexto, el conjunto de las 3×3 -matrices simétricas M_q es un espacio vectorial V^6 de dimensión 6. Denotamos mediante \mathbf{P}^5 al proyectivizado $\mathbf{P}V^6$ del espacio vectorial de las matrices simétricas que representa las cónicas del plano.

Las cónicas del plano proyectivo real se clasifican por el rango $rk(M_q)$ y la signatura $sign(M_q)$ de la matriz M_q . En el caso proyectivo complejo sólo hay tres tipos de cónicas correspondientes a los posibles valores para $1 \leq rk(M_q) \leq 3$

²¹ Proporcionan una primera aproximación a las superficies B-splines de amplio uso en Informática Gráfica, Análisis Numérico ó Ingeniería Industrial, p.e.

²² Cualquier espacio proyectivo es una variedad compacta y, por tanto, la optimización da lugar a una solución única; es necesario identificar los caminos de longitud mínima (geodésicas) asociados a dichos procesos de optimización

donde M_q es la matriz que representa la cónica q salvo factor de proporcionalidad.

1.6.2. Estimación de una cónica del plano proyectivo

La estimación de la *cónica lugar* se realiza mediante el método de coeficientes indeterminados a partir de una colección minimal de puntos en posición general

Un punto $\mathbf{p}_i = (x_i, y_i, 1)$ pertenece a una cónica \mathcal{C} si, y sólo si

$$ax_i^2 + 2bx_iy_i + cy_i^2 + 2dx_i + 2ey_i + f = 0 = (x_i, y_i, 1)C(x_i, y_i, 1)^\top$$

que impone una condición lineal sobre los coeficientes que reescribimos como

$$(x_i^2, x_iy_i, y_i^2, x_i, y_i, 1)(a, 2b, c, 2d, 2e, f)^\top = 0$$

Como el conjunto de las cónicas del plano es un espacio proyectivo \mathbb{P}^5 , para determinar una cónica salvo factor de proporcionalidad hacen falta 5 condiciones independientes correspondientes a la condición de pasar por 5 puntos del plano en posición general. En términos algebraicos, la cónica \mathcal{C} es el núcleo de la aplicación lineal definida por la 5×6 matriz

$$\begin{pmatrix} x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 & 1 \\ x_2^2 & x_2y_2 & y_2^2 & x_2 & y_2 & 1 \\ x_3^2 & x_3y_3 & y_3^2 & x_3 & y_3 & 1 \\ x_4^2 & x_4y_4 & y_4^2 & x_4 & y_4 & 1 \\ x_5^2 & x_5y_5 & y_5^2 & x_5 & y_5 & 1 \end{pmatrix}$$

Como *conclusión* son necesarios 5 puntos en posición general para determinar una cónica general de rango 3. En Reconstrucción 3D es frecuente utilizar condiciones de incidencia y tangencia (ambas son proyectivamente invariantes) para determinar elementos de forma única; en particular, pasando al dual la descripción anterior, se comprueba fácilmente que una cónica regular se determina de forma única por 5 rectas tangentes. Ello se debe a que la condición de pasar por un punto es dual de la condición de ser tangente a una recta²³.

Ejercicio. - ¿Es contradictoria necesidad de 5 puntos (o 5 tangentes) con la suficiencia de 3 puntos para determinar una circunferencia? (como cónica regular). ¿Cuántos puntos hacen falta para determinar una cónica degenerada de

²³ La combinación de condiciones de incidencia y tangencia puede dar lugar a falta de unicidad en las soluciones; así, p.e. hay dos cónicas que pasen por 4 puntos y sean tangentes a 1 recta y hay cuatro cónicas que pasen por 3 puntos y sean tangentes a 2 recta (análogamente para sus duales); estas observaciones se deben a M.Chasles (hacia 1837) quien lo formuló en términos de las “polares recíprocas”; estos resultados forman parte del origen de la Geometría Enumerativa, una rama de la Geometría Algebraica que analiza el número de soluciones que tienen problemas relacionados con ecuaciones algebraicas, en el supuesto de que dicho número sea finito

rango 2 o 1? Verificad que 4 puntos (resp. 3 puntos) en posición general determinan en general una familia uniparamétrica (resp. biparamétrica) de cónicas del plano proyectivo \mathbb{P}^2 ,

Nota.- Si se desea ajustar una cónica a $N \geq 6$ puntos, es necesario realizar un proceso de *optimización no-lineal* que extiende los procedimientos de regresión lineal presentados en el capítulo 3 del módulo 1 en relación con la estimación de rectas.

Ejercicio.- Diseñar un método para estimar la cónica que mejor aproxima los N puntos (*Indicación.*- Utilizad regresión cuadrática para ajustar una “elipse de inercia” adaptando el método de Newton de los diámetro conjugados)

1.6.3. Envolvente de una cónica por rectas

La *cónica envolvente* C^ν de una cónica plana C se define como la envolvente de las tangentes a la cónica C .

Lema (condición de tangencia).- Una recta ℓ es tangente a la cónica C en \mathbf{p} si y sólo si $\ell = C\mathbf{x}$.

Definición.- La *cónica dual* es la envolvente por rectas tangentes $\mathbf{x} = C^{-1}\ell$.

Lema.- La matriz de la cónica dual es la adjunta C^* de C (verificarlo como ejercicio).

Utilizando la dualidad entre punto y línea en el plano proyectivo y que la dual de la condición de incidencia (pasar por un punto) es una condición de tangencia (ser tangente a una cónica se obtiene el resultado siguiente:

Lema (cónica dual o envolvente).- 5 líneas en posición general determinan una única cónica.

Es importante tener una descripción de las condiciones de tangencia para los *casos degenerados* (cónicas de rango < 3):

- Para una *cónica de rango 2*, una recta ℓ es tangente a la cónica sii

$$\ell_1 \ell_2^\top + \ell_2 \ell_1^\top$$

El núcleo es el punto $\mathbf{p} = \ell_1 \times \ell_2$.

- Una *Cónica de rango 1* se expresa mediante:

$$\mathbf{p}_1 \mathbf{p}_2^\top + \mathbf{p}_2 \mathbf{p}_1^\top$$

. El núcleo es el conjunto de líneas que pasan por alguno de los puntos \mathbf{p}_1 o \mathbf{p}_2

Estas condiciones tienen interés para la estimación de la métrica (en relación con la cónica absoluta) mediante métodos proyectivos. En ausencia de una calibración previa conocida de la cámara, la estimación de la cónica absoluta es clave para la calibración on-line de una cámara ²⁴. Nótese que a menudo

²⁴ Esta aplicación se traduce en las ecuaciones de Kruppa que se obtienen eliminando parámetros de las tangentes trazadas a una cónica desde un punto exterior

(fotos antiguas, grabados realizados de acuerdo con las leyes de perspectiva) es imposible realizar una calibración de la cámara y, en otros casos (condiciones adversas, golpes, alunizaje, p.e.) la cámara ha podido descalibrarse.

2. Transformaciones proyectivas. Estimación

Las transformaciones proyectivas del plano (homografías) de imagen son clave para el pegado de datos lineales procedentes del análisis de varias vistas. Asimismo, permiten simular transformaciones globales asociadas a efectos globales de distorsión asociados a un cambio virtual en la posición del observador tanto a nivel de imagen como de toda la escena. Por ello, estas transformaciones proporcionan los elementos básicos para el pegado de vistas y la simulación de movimientos de cámara.

Esta sección comienza describiendo las transformaciones proyectivas como elementos del proyectivizado $\mathbb{P}GL(n+1; \mathbb{R})$ del grupo lineal general, identificando subgrupos de interés para diferentes marcos geométricos. El método de coeficientes indeterminados proporciona una primera aproximación al problema de la estimación.

Un requisito importante es la *robustez* en la estimación, es decir, pequeñas perturbaciones en los inputs no deben producir cambios cualitativos en la transformación proyectiva estimada. La robustez se satisface de forma inmediata para el proyectivizado del grupo lineal general (los automorfismos forman un abierto dentro del conjunto de endomorfismos de un espacio vectorial), pero la estimación presenta una mayor inestabilidad cuando se analizan subgrupos clásicos del grupo lineal general. En la subsección 2 se realiza una primera aproximación al problema.

A menudo, las descomposiciones de las transformaciones geométricas en alguno de los marcos intermedios (euclídeo, afín, proyectivo) facilitan elementos geométricos que permiten desacoplar el problema de la estimación, simplificando la estimación; este tópico se aborda en la tercera subsección.

La última subsección está dedicada a la estimación de la métrica euclídea que es clave para la Reconstrucción Euclídea, es decir, módulo transformaciones rígidas salvo escala. Los procesos de optimización relacionados con la estimación se abordan más adelante en un contexto más restringido relacionado con la introducción de la Geometría Epipolar.

2.1. Transformaciones proyectivas

En la introducción ya se ha comentado la representación proyectiva de líneas paralelas del espacio cartesiano como líneas convergentes en un punto de fuga \mathbf{V} (vanishing point) también llamado punto del infinito²⁵. Esta representación está justificada por la homogeneidad del espacio proyectivo por la acción del grupo de las transformaciones proyectiva; en otras palabras, cualquier punto del plano proyectivo usado para la representación puede ser un punto de fuga. En esta subsección formalizamos y extendemos este tipo de argumentos.

Las representaciones en perspectiva utilizan haces λ_α de líneas de perspectiva ℓ_i^α a través de cada punto de fuga \mathbf{V}_α . Un haz de líneas a través de un punto

²⁵ Se admite que las líneas paralelas acaban cortándose “en el infinito” para dar un significado a nuestra percepción visual

es una recta proyectiva \mathbb{P}^1 en el plano proyectivo dual $(\mathbb{P}^2)^\nu$. En particular, en un escenario arquitectónico las líneas de visión L_i se proyectan sobre el plano de imagen $\Pi_{\mathbb{C}}$ y se empaquetan en haces de líneas de perspectiva ℓ_i^α .

Los cambios en la visualización asociada a una modificación en la localización virtual de una cámara se representan mediante una transformación sobre cada uno de los haces de líneas de perspectiva (representados por una línea proyectiva). Por ello, la forma más simple de representar un cambio asociado a mapas de perspectiva y simular el cambio en la localización virtual de una cámara es mediante una transformación proyectiva; el caso más simple corresponde a una perspectiva frontal en el que sólo se tiene un punto de fuga a distancia finita. Más adelante se muestra cómo corregir las distorsiones asociadas a esta representación.

Con más generalidad, si deseamos representar transformaciones en el espacio tridimensional ordinario sin recurrir a sus proyecciones, es necesario disponer de modelos para las transformaciones entre haces Λ_α de planos proyectivos Π_i^α que pasan por una recta ℓ_α ; nuevamente, cada haz Λ_α es una línea proyectiva \mathbb{P}^1 , en este caso de planos en el espacio proyectivo dual $(\mathbb{P}^3)^\nu$. Asimismo, el análisis de una familia uniparamétrica de cónicas del plano proyectivo \mathbb{P}^2 (ó el de cuádricas del espacio proyectivo \mathbb{P}^3) se representa mediante las transformaciones de una recta proyectiva. Por todo ello y a pesar de su simplicidad, debemos empezar analizando las transformaciones proyectivas de una recta proyectiva \mathbb{P}^1 , tópico al que se dedica el primer apartado de la subsección; el formalismo algebraico presentado es fácilmente generalizable a dimensión arbitraria y por ello, se presenta con cierto detalle.

En cualquier caso, la condición de regularidad (determinante no nulo) para el grupo lineal general define un abierto del espacio vectorial $M(n \times n; \mathbb{R}) \simeq \mathbb{R}^{n^2}$ de las matrices de tamaño $n \times n$ con coeficientes reales (el abierto se considera con respecto a la topología de los coeficientes). La descripción para el caso complejo es similar, reemplazando el cuerpo real \mathbb{R} por el cuerpo \mathbb{C} de los números complejos (esta observación es relevante para cuestiones relacionadas con la aplicación del análisis de Fourier al procesamiento y Análisis de Imagen ²⁶).

Son transformaciones regulares de \mathbb{P}^n en \mathbb{P}^n , es decir, elementos del proyectivizado $\mathbb{P}GL(n+1; \mathbb{R})$ del grupo lineal general. En esta subsección se sigue una complejidad creciente con los apartados siguientes:

1. Transformaciones de la recta proyectiva \mathbb{P}^1
2. Visualización geométrica de las colineaciones.
3. Transformaciones lineales del plano proyectivo \mathbb{P}^2
4. Primera aproximación a la recuperación de las homografías
5. Transformaciones de líneas y cónicas
6. Invariantes para las colineaciones

²⁶ Ver capítulo 6 del módulo 1 para más detalles

2.1.1. Transformaciones de la recta proyectiva

Las transformaciones regulares $\mathbb{P}^1 \rightarrow \mathbb{P}^1$ o *colineaciones* están representadas por aplicaciones lineales no-degeneradas, es decir, por elementos $\mathbf{A} \in \mathbb{P}GL(2; \mathbb{R})$. La acción de una colineación se expresa analíticamente como

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

que equivale al sistema de dos ecuaciones

$$y_1 = a_{11}x_1 + a_{12}x_2, \quad y_2 = a_{21}x_1 + a_{22}x_2$$

La *descripción afín* es similar a la del caso proyectivo. Para fijar ideas, trabajamos en el abierto coordenado afín $D_2^+ := \{\mathbf{p} \in \mathbb{P}^1 \mid x_2 \neq 0\} \simeq \mathbb{A}^1$ parametrizado por $(x_2/x_1, 1)$ donde $[x_1 : x_2]$ son coordenadas homogéneas de la recta proyectiva \mathbb{P}^1 . La imagen de un abierto afín de \mathbb{P}^1 por una aplicación proyectiva $\varphi^p : \mathbb{P}^1 \rightarrow \mathbb{P}^1$ representada por \mathbf{A} es un abierto afín en la línea proyectiva \mathbb{P}^1 de llegada; mediante un cambio de coordenadas, podemos suponer que dicho abierto afín corresponde a la condición $y_2 \neq 0$. Por ello, la transformación proyectiva induce una transformación afín $\varphi^a : D_+(x_2) \simeq \mathbb{A}^1 \rightarrow D_+(y_2) \simeq \mathbb{A}^1$ donde $D_+(y_2) := \{\underline{y} \mid y_2 \neq 0\}$ que re-escribimos como

$$\frac{y_1}{y_2} = \frac{a_{11}x_1 + a_{12}x_2}{a_{21}x_1 + a_{22}x_2}$$

El cambio de coordenadas afines en el espacio de partida sobre la intersección $D_+(x_1) \cap D_+(x_2)$ se lleva cabo mediante las “funciones de transición” $x_1^{-1}x_2$ (ó su inversa); análogamente en el espacio de llegada, el cambio sobre la intersección $D_+(y_1) \cap D_+(y_2)$ se realiza mediante $x_1^{-1}x_2$ (o su inversa). Nótese que cuanto más pequeños son los denominadores, mayor es la distorsión y que esta dependencia es no-lineal. Por ello, la visualización Geométrica y radiométrica) de este tipo de modelos sólo es válida para una región próxima a la dirección del eje óptico. Para minimizar estos efectos visuales asociados al caso móvil es necesario realizar una “corrección” utilizando parámetros internos de la cámara (calibración interna) que se desarrolla en el capítulo 2 y una corrección mediante transformaciones de perspectiva con coordenadas curvilíneas que se aborda en el capítulo 4. La figura 6 ilustra la necesidad de utilizar coordenadas curvilíneas en escenarios sencillos asimismo creados por el hombre

2.1.2. Visualización geométrica de las colineaciones

Si tomamos $(a_{11} : a_{12} : a_{21} : a_{22})$ como coordenadas proyectivas en \mathbf{P}^3 , las colineaciones de \mathbb{P}^1 están representadas por el abierto del espacio proyectivo \mathbf{P}^3 (que parametriza las entradas a_{ij} de la matriz asociada a la colineación) que es complementario de la cuádrica $a_{11}a_{22} - a_{12}a_{21} = 0$.

Las imágenes de tres puntos $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ sobre una recta ℓ determinan de forma única una colineación, pues usando coordenadas locales sobre la recta, se puede escribir



Fig. 6: Utilidad de las coordenadas curvilíneas en escenarios arquitectónicos

$$\frac{(y - y_1)(y_2 - y_3)}{(y - y_3)(y_2 - y_1)} = \frac{(x - x_1)(x_2 - x_3)}{(x - x_3)(x_2 - x_1)}$$

donde $y := \frac{a_{11}x + a_{12}}{a_{21}x + a_{22}}$. Por consiguiente, los *datos a estimar* son 3 parámetros salvo escala \Rightarrow Dos puntos y sus homólogos. Existen diferentes formas de visualizar las colineaciones. Algunas de las más utilizadas están relacionadas con la versión afín descrita en el apartado anterior y utilizan una reinterpretación en términos de coordenadas complejas.²⁷ Las transformaciones de Moebius se describen en nuestro caso mediante

$$\varphi(z) = \frac{a_{11}z + a_{12}}{a_{21}z + a_{22}} \quad \det(A) \neq 0$$

donde $z = \frac{x_1}{x_2}$ en $D_+(x_2)$. Estas transformaciones generan el grupo conforme sobre la recta compleja \mathbb{C}^1 ; la compactificación de dicha recta añade un “punto del infinito” que da la recta proyectiva compleja $\mathbb{C}\mathbb{P}^1$ que resulta ser equivalente a la esfera de radio unidad. Esta equivalencia se puede visualizar mediante la representación real de la recta compleja dada por pares $(x, y) \in \mathbb{R}^2$ que es equivalente a la esfera $\mathbb{S}^2 - \{\mathbf{N}\}$ mediante la proyección estereográfica, donde $\mathbf{N} = (0, 0, 1)$ es el polo norte. Por consiguiente, la compactificación del plano en una esfera proporciona una reinterpretación geométrica de las transformaciones de Moebius como transformaciones de la esfera \mathbb{S}^2 ; estas transformaciones generan el primer “ejemplo” de grupo de *transformaciones conformes*. Conservan los ángulos y tienen un uso creciente en modelado 3D, incluyendo efectos dinámicos asociados a curvaturas cambiantes²⁸.

²⁷ Esta re-interpretación fue introducida a mediados del siglo XIX por Moebius en relación con cuestiones relacionadas con las funciones de una variable compleja.

²⁸ Estas cuestiones se desarrollan con más detalle en el módulo 5

2.1.3. Transformaciones lineales del plano proyectivo

La proyección de una escena estática sobre el plano de imagen se puede interpretar como una parte dada por un espacio afín \mathbb{A}^2 (dominio coordinado) de un plano proyectivo que es complementaria de una línea del infinito $\ell_\infty \simeq \mathbb{P}^1$; de una manera más abstracta se tiene una descomposición (no canónica), $\mathbb{P}^2 \simeq \mathbb{A}^2 \cup \mathbb{P}^1$. De cara al “pegado” de varias vistas (sobre la intersección de abiertos coordinados afines) o a la generación sintética de nuevas vistas (simulación del movimiento de cámara) tiene sentido considerar las transformaciones globales del plano proyectivo.

Las transformaciones lineales del plano proyectivo \mathbb{P}^2 son aplicaciones lineales no-degeneradas $\in \mathbb{P}GL(3; \mathbb{R})$. Fijadas referencias en los espacios de partida y llegada, la *expresión analítica* de la acción de la transformación lineal se escribe como

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

con $\det(A) \neq 0$. El sistema anterior equivale al sistema de tres ecuaciones

$$x_i = a_{i1}X_1 + a_{i2}X_2 + a_{i3}X_3$$

que reescribimos en el abierto coordinado $D_+(x_3)$ como un par de ecuaciones afines

$$y_1 = \frac{x_1}{x_3} = \frac{a_{11}X_1 + a_{12}X_2 + a_{13}X_3}{a_{31}X_1 + a_{32}X_2 + a_{33}X_3}$$

$$y_2 = \frac{x_2}{x_3} = \frac{a_{21}X_1 + a_{22}X_2 + a_{23}X_3}{a_{31}X_1 + a_{32}X_2 + a_{33}X_3}$$

que nos permite recuperar las homografías y resolver el problema en los pasos siguientes;

1. *Hipótesis*: Suponemos conocidas las coordenadas píxel $(y_1, y_2, 1)$ e $(Y_1, Y_2, 1)$ correspondientes a cada par de puntos homólogos en dos vistas
2. *Objetivo*: Resolver el sistema

$$y_1(a_{31}Y_1 + a_{32}Y_2 + a_{33}) = a_{11}Y_1 + a_{12}Y_2 + a_{13}$$

$$y_2(a_{31}Y_1 + a_{32}Y_2 + a_{33}) = a_{21}Y_1 + a_{22}Y_2 + a_{23}$$

3. Para resolver este sistema es necesario estimar la posición de 4 puntos en posición general $\mathbf{p}_1, \dots, \mathbf{p}_4$ y sus homólogos $\mathbf{p}'_1, \dots, \mathbf{p}'_4$ en dos vistas.
4. *Cuatro puntos en posición general* determinan un cuadrilátero.
5. La geometría de las *transformaciones de cuadriláteros* proporciona el soporte para el estudio de las homografías (transformaciones proyectivas) del plano \mathbb{P}^2

La transformación de un rectángulo en un cuadrilátero arbitrario proporciona una idea aproximada del "sesgo oblicuo" (efecto de perspectiva) producido por la localización del observador en relación con la escena. La "aparición" del rectángulo como cuadrilátero varía cuando el observador se desplaza, dando lugar a "degeneraciones" (desaparición de alguna de las aristas, p.e.) o "desdoblamientos" del cuadrilátero sobre el que se pivota. La extensión de este enfoque a mapas de cuadriláteros (eventualmente curvados) presenta una mayor dificultad y por ello no se aborda en estas notas

2.1.4. Transformaciones de líneas y cónicas

Al comparar elementos homólogos situados en diferentes imágenes, frecuentemente resulta más cómodo y robusto analizar las transformadas de primitivas geométricas en diferentes imágenes; las primitivas geométricas más sencillas son líneas y cónicas.

Para empezar, recordemos que dos *puntos* son *homólogos* si y sólo si existe una Transformación que lleva \mathbf{p}_i en \mathbf{p}'_i por una colineación H , se decir, $\mathbf{p}'_i = H\mathbf{p}_i$.

Análogamente, para *rectas homólogas* se tiene que la transformada ℓ'_i de una línea ℓ_i por una colineación H está dada por $\ell'_i = H^{-T}\ell_i$.

Una propiedad importantes es la *conservación de la incidencia*, es decir,

$$\ell'^T \mathbf{x} = \ell^T H^{-1} H \mathbf{x} = 0$$

Análogamente, la *transformada* C' de una *cónica* C por una colineación H está dada por

$$\mathbf{p}^T C \mathbf{p} = \mathbf{p}'^T (H^{-1})^T C H^{-1} \mathbf{p}'$$

Cuando se utilizan cónicas *las distorsiones en los ejes y el cambio de orientación* de una elipse facilitan la *calibración de la cámara*. Las cuestiones de calibración se abordan en el capítulo siguiente.

2.2. Robustez para el caso geométrico

La *robustez* se entiende como independencia con respecto a "pequeñas variaciones" en los parámetros. La robustez afecta a los objetos y las transformaciones. La robustez depende del marco geométrico elegido.; así, p.e. en el marco euclídeo, ambas cuestiones se ilustran con estimadores robustos para isometrías (transformaciones que conservan la métrica) que son relevantes para la Reconstrucción Euclídea; en este caso, la cuestión crucial a verificar es la conservación de la distancia euclídea.

Con más generalidad, en el caso afín interesa introducir estimadores que conservan la noción de paralelismo o en el caso proyectivo la razón doble entre cantidades. Frecuentemente, interesa estimar las cantidades descritas anteriormente salvo escala o bien interesa introducir otros criterios relativos a la conservación de áreas (o con más generalidad volúmenes) o bien ángulos. Todo ello, se

puede abordar de una manera unificada en el contexto de la Teoría de Invariantes. Para fijar ideas, sólo se consideran invariantes (multi)lineales básicos cuya conservación está asociada a elementos de un grupo fácil de describir (aunque no de estimar). Los objetos más relevantes para la estimación del modelo son las *líneas de perspectiva* a cuya estimación se dedica el primer apartado.

Esta subsección está enfocada a cuestiones básicas relacionadas con invariantes geométricos y, de una forma más específica con la razón doble como el invariante proyectivo fundamental, incluyendo elementos para su estimación. Obviamente, si se fijan otros elementos adicionales (hiperplano del infinito, cuádrica absoluta como representación geométrica de la métrica euclídea) se obtiene una mayor cantidad de invariantes, lo cual facilita la visualización de las transformaciones a realizar. En la subsección siguiente se comentan los invariantes relacionados con transformaciones afines y transformaciones que conservan la métrica o isometrías (que extienden las transformaciones euclídeas usuales).

El enfoque basado en la estimación de invariantes tiene múltiples aplicaciones relacionadas con cuestiones de navegación semi-automática, simulación o animación relacionadas con modelos 3D, proporcionando una realimentación entre la “robustez” del modelo vinculado a una escena rígida y la localización del observador. La robustez geométrica es crucial para garantizar la estabilidad de una escena sobre la que se superponen diferentes contenidos; esta “estabilidad” (módulo transformaciones proyectivas) se describe en la literatura de visión computacional como la “verdad del fondo”. Una vez estimada, más adelante veremos cómo utilizando mapas de profundidad dinámicos podemos superponer una estabilidad estructural para objetos “móviles” en primer plano. Estas cuestiones más avanzadas se abordan en el módulo 5 del CEViC, en relación con algunas aplicaciones de la Visión Estéreo Dinámica.

2.2.1. Invariantes geométricos para el caso lineal

Los invariantes a considerar dependen del marco geométrico elegido

- *Invariantes euclídeos:* La *distancia* es el invariante fundamental. Su conservación implica la invariancia de ángulos, áreas, volúmenes etc El inverso no es cierto.
- *Invariantes afines:* El *paralelismo* es el invariante fundamental; es posible deducirlo a partir de la conservación de la línea del infinito ℓ_∞ , en el plano proyectivo. Como consecuencia se obtiene una invariancia para la ratio de longitudes de longitudes sobre líneas paralelas o colineales (versión sintética del teorema de Thales), a partir de la cual se obtiene invariancia para la razón de áreas o de volúmenes en el caso volumétrico.
- *Invariante proyectivo:* El invariante fundamental es la *razón doble* o cross-ratio. La razón doble se define como

$$\text{Cross}(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4) := \frac{|\overline{\mathbf{p}_1\mathbf{p}_2}| |\overline{\mathbf{p}_3\mathbf{p}_4}|}{|\overline{\mathbf{p}_1\mathbf{p}_3}| |\overline{\mathbf{p}_2\mathbf{p}_4}|}$$

La utilización de estos invariantes permite reconstruir objetos para cada uno de los marcos geométricos de referencia (euclídeo, afín o proyectivo son los más comunes). La elección de cada marco geométrico proporciona una visualización asociada a la localización del observador. La compatibilidad entre diferentes representaciones es consecuencia de la existencia de un modelo proyectivo general que incluye a todos los demás.

Una aportación significativa del enfoque basado en Visión por Computador consiste en que la Reconstrucción se puede realizar directamente sobre la vista, sin tener que recurrir a objetos 3D. Este hecho abarata el coste computacional en la producción de contenidos digitales y no requiere información completa sobre el objeto para su visualización. La clave para poder llevar esto a cabo consiste en una utilización de los diferentes tipos de transformaciones que afectan a los objetos y su representación en imagen incluyendo posibles distorsiones debidas a efectos de perspectiva o bien operaciones típicas de la Geometría Proyectiva (secciones y proyecciones).

2.2.2. Motivación y expresión de la razón doble

La razón simple entre longitudes de segmentos que se corresponden por una transformación proyectiva no se conserva (verificarlo como ejercicio). Sin embargo, la razón doble o cruzada entre las longitudes de pares de segmentos generados por 4 puntos alineados sobre una recta sí se conserva por colineaciones. Esta invariancia ya fue observada por Desargues (1648) si bien su uso generalizado se debe a Chasles et al a partir de la segunda mitad del s.XIX. Para fijar ideas en relación con la Visión Computacional, adoptamos una terminología relacionada con propiedades de las cámaras y su representación geométrica

Supongamos conocidos 4 puntos cuyas proyecciones desde \mathbf{q} son colineales: $\mathbf{a} + \lambda_{\mathbf{p}_i} \mathbf{b}$. Denotemos mediante $(\lambda_{\mathbf{p}_i}, 1)$ a las coordenadas afines de \mathbf{p}_i es decir, $\mathbf{a} = (a_1 : a_2)^\top$, $\mathbf{b} = (b_1 : b_2)^\top$. La *Ecuación de la Proyección Perspectiva* se escribe como

$$\mathbf{p}_i = f \frac{a_1 + \lambda_{\mathbf{p}_i} b_1}{a_2 + \lambda_{\mathbf{p}_i} b_2}$$

siendo f la longitud focal. La *distancia entre dos proyecciones* está dada por:

$$\mathbf{p}_i - \mathbf{p}_j = f \frac{(a_2 b_1 - a_1 b_2)(\lambda_{\mathbf{p}_i} - \lambda_{\mathbf{p}_j})}{(a_2 + \lambda_{\mathbf{p}_i} b_2)(a_2 + \lambda_{\mathbf{p}_j} b_2)}$$

Con esta notación, la *razón doble* se define como:

$$\text{Cross}(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4) := \frac{\mathbf{p}_2 - \mathbf{p}_1}{\mathbf{p}_3 - \mathbf{p}_1} : \frac{\mathbf{p}_2 - \mathbf{p}_4}{\mathbf{p}_3 - \mathbf{p}_4} = \frac{\lambda_{\mathbf{p}_2} - \lambda_{\mathbf{p}_1}}{\lambda_{\mathbf{p}_3} - \lambda_{\mathbf{p}_1}} : \frac{\lambda_{\mathbf{p}_2} - \lambda_{\mathbf{p}_4}}{\lambda_{\mathbf{p}_3} - \lambda_{\mathbf{p}_4}}$$

La posibilidad de recuperar información sobre objetos de una escena de forma independiente con respecto a la posición del observador ya está implícita en resultados bien conocidos en la Geometría Proyectiva de mediados del s.XIX como la *generación proyectiva de las cónicas*: El lugar de los puntos $\mathbf{q} \in \mathbb{P}^2$ t.q. $\text{cross}(\ell_1, \ell_2, \ell_3, \ell_4) = \text{cte.}$ donde $\ell_i = \mathbf{q}\mathbf{p}_i$ para $i = 1, \dots, 4$, es una cónica. El resultado anterior (debido a Chasles, hacia 1850) es una primera muestra del uso

de invariantes para la reconstrucción de objetos situándose en una localización arbitraria con respecto al objeto.

2.2.3. Estimación de la razón doble

Las transformaciones proyectivas son apropiadas para gestionar información global de escenas desde cualquier punto de vista. La estimación de la razón doble no es elemental pues se requiere trabajar con 4 puntos alineados y sus imágenes por transformaciones proyectivas; los puntos habitualmente son vértices de poligonales o esquinas; en el módulo 1 se ha mostrado la presencia de ruido asociado a este tipo de “hechos”. Por ello, resulta más práctico y robusto pasar el punto de vista dual y considerar (al menos) 4 líneas que pasan por un punto que es la condición dual en el plano proyectivo a que 4 puntos estén sobre una recta. En este caso, las 4 líneas a través de un punto son líneas de perspectiva ℓ_i^α y dicho punto es un punto de fuga \mathbf{V}^α . Dichas líneas generan un haz λ^α de líneas de perspectiva.

Las *junturas triples* están (dadas por la confluencia de tres elementos lineales (en forma de flecha \uparrow o de Y); en el módulo 1 se han mostrado herramientas para extraer junturas dobles y triples en relación con la detección de esquinas (filtros de Harris et al). Aunque en el modelo matemático las junturas triples sean estables por perturbación, son el soporte de ruido en imagen lo cual hace difícil una estimación robusta y, frecuentemente, es preciso calcular la intersección de los elementos lineales próximos.

En una juntura cuádruple confluyen 4 elementos lineales por lo que su soporte no sólo contiene ruido sino que es inestable incluso en el modelo teórico. En este caso, hay que identificar previamente los candidatos a las líneas de fuga ℓ_i y agruparlas en haces λ ; para cada haz hay que

1. ajustar la intersección de los elementos lineales ($\ell_i \cap \ell_j = \mathbf{p}_{ij}$ en el caso de líneas p.e.) para que proporcionen el elemento esperado (punto de fuga \mathbf{V}_λ del haz $\lambda = \langle \ell_i, \ell_j \rangle$, p.e.)
2. utilizar procedimientos de optimización con pesos en previsión de distribuciones irregulares para las intersecciones (en torno al punto de fuga, p.e.);
3. retrazar los elementos lineales del haz de acuerdo con la incidencia esperada (líneas de fuga sintéticas $\hat{\ell}_i$ tan “próximas” como sea posible a las obtenidas;
4. calcular las intersecciones con elemento transversal (línea que corta a las 4 del haz);
5. estimar la razón doble correspondiente a las trazas de la línea transversal sobre las 4 líneas seleccionadas del haz λ

2.3. Transformaciones más significativas

En esta subsección se presentan las transformaciones lineales más significativas para la reconstrucción. El *objetivo inicial* es generar de forma sintética nuevas vistas 2D de objetos o de escenas 3D y no tanto generar un objeto 3D propiamente dicho; este último requiere una gestión eficiente de mapas de profundidad que se aborda en capítulos más avanzados de este módulo ²⁹ Por ello, se adopta un enfoque basado inicialmente en transformaciones del plano de imagen (entendido como un plano proyectivo); dichas transformaciones se pueden entender como un subgrupo lineal del (proyectivizado del) grupo lineal general y están estrechamente vinculadas a los diferentes marcos geométricos que se han comentado más arriba.

Desde el punto de vista proyectivo los elementos a transformar son haces de líneas $\lambda^{(\alpha)}$ que pasan por un punto a las que llamaremos *líneas de perspectiva*; si se interpreta dicho punto como un punto de fuga $\mathbf{V}^{(\alpha)}$ la versión en el plano cartesiano es un haz de líneas paralelas. La consideración de dos haces $\lambda^{(\alpha)}$, $\lambda^{(\beta)}$ de líneas proyectivas da lugar a un mapa de cuadriláteros $Q^{(\alpha\beta)} = \lambda^{(\alpha)} \times \lambda^{(\beta)}$ cuya versión cartesiana sería un mapa rectangular; cada cuadrilátero en el proyectivo es la imagen de un rectángulo por una homografía $Aut(\mathbb{P}^2)$ del plano que deja invariante la línea del horizonte $\mathbf{V}^{(\alpha)} \times \mathbf{V}^{(\beta)}$. La *rectificación de una imagen* consiste en convertir los dos haces de líneas de perspectiva en haces de líneas paralelas ortogonales.

Para objetos curvados (esculturas, p.e.) el método anterior proporciona un marco demasiado general y de poca utilidad, pues los haces de líneas rectas pasando por un punto no se adaptan a la geometría curvada del objeto. En este caso, es preferible remapear los haces de líneas paralelas ortogonales del plano cartesiano en haces de curvas que se adaptan a la geometría del objeto. Esta transformación es no-lineal (por lo que no existe un grupo lineal formado por matrices con coeficientes constantes) y se puede realizar de infinitas maneras. Para acotar el problema, es conveniente introducir restricciones adicionales que faciliten el desarrollo de técnicas asociadas a problemas de optimización.

Las transformaciones conformes proporcionan un candidato natural para resolver este problema. Una *transformación conforme* es una transformación proyectiva que conserva ángulos (no necesariamente distancias). En este caso, los ángulos se refieren a los formados por las coordenadas curvilíneas (imágenes por C^r -equivalencias locales de coordenadas cartesianas). Este tipo de transformaciones son especialmente apropiadas para simular efectos de curvatura sobre superficies. El modelado y gestión computacional de este tipo de transformaciones se aborda en el módulo 5 del CEViC en relación con la reconstrucción de esculturas (caso estático) o actores en movimiento (caso dinámico).

De momento y para fijar ideas, comenzamos con el caso básico correspondiente a las transformaciones lineales de interés para la reconstrucción 3D. Para simplificar nos reducimos al caso de transformaciones en el plano, aunque buena parte de los argumentos se extienden a transformaciones en el espacio tridimen-

²⁹ Estos tópicos se introducen en el módulo 5 (Visión Estéreo) que se ofrece como módulo complementario del CEViC.

sional. Se presta más atención a las transformaciones afines por tratarse del caso más general; las transformaciones de semejanza (rígidas salvo factor de escala) se abordan con más detalle en relación con la reconstrucción euclídea que requiere conocimientos previos de calibración.

2.3.1. Transformaciones afines

Una *transformación afín* es una transformación proyectiva que deja invariante un hiperplano del infinito H_∞ . Como el grupo proyectivo actúa de forma transitiva sobre puntos e hiperplanos, podemos suponer que dicho hiperplano es el correspondiente a la anulación de la última coordenada, es decir, tomamos $\{\underline{x} \in \mathbb{P}^n \mid x_n = 0\}$ como H_∞ . Si denotamos mediante $D_+(x_n) := \{\underline{x} \in \mathbb{P}^n \mid x_n \neq 0\}$ al espacio afín complementario del hiperplano $x_n = 0$, se tiene una descomposición inducida por la elección del plano del infinito:

$$\mathbb{P}^n = D_+(x_n) \cup \{x_n = 0\} \simeq \mathbb{A}^n \cup \mathbb{P}^{n-1}$$

en unión disjunta de un espacio afín y un hiperplano del infinito. Las transformaciones afines dejan invariante dicha descomposición. En términos más formales el *grupo afín* es el “producto semidirecto” $\mathbf{A}(n; \mathbb{R}) := GL(n; \mathbb{R}) \cdot \mathbb{R}^n$ del grupo lineal general (automorfismos del espacio vectorial subyacente) y el grupo de traslaciones \mathbb{R}^n . Su descripción como subgrupo de $GL(n+1; \mathbb{R})$ está dada por el conjunto de matrices

$$\mathbf{A}(n; \mathbb{R}) := \left\{ \begin{pmatrix} A & \mathbf{v} \\ 0 & 1 \end{pmatrix} \mid A \in GL(n; \mathbb{R}), \mathbf{v} \in \mathbb{R}^n \right\}$$

La *operación* (ley interna del grupo) está dada por la asignación

$$(\mathbf{A}, \mathbf{v}_1) \cdot (\mathbf{B}, \mathbf{v}_2) := (\mathbf{AB}, \mathbf{v}_1 + \mathbf{Av}_2)$$

La *acción del grupo afín* sobre el espacio afín está dada por

$$\mathbf{A}(n; \mathbb{R}) \times \mathbb{A}^n \rightarrow \mathbb{A}^n \mid (\mathbf{A}, \mathbf{v})(\mathbf{P}, 1) := \mathbf{AP} + (\mathbf{v}, 1)$$

El análisis de las transformaciones afines para el modelo proyectivo de una imagen digital es un caso particular para $n = 2$ y se ha presentado en la sección 4 del capítulo 1 del módulo 1. Por ello, repasamos dichos contenidos con la descomposición similar que se tiene para el espacio afín tridimensional de interés para la reconstrucción afín de la escena:

$$\mathbb{P}^3 = D_+(x_3) \cup \{x_3 = 0\} \simeq \mathbb{A}^3 \cup \mathbb{P}^2$$

donde \mathbb{P}^2 es cualquier plano proyectivo; en la práctica es frecuente tomar dicho plano como un plano del infinito generado por dos líneas independientes del horizonte; cada línea del horizonte $\ell_\infty = \mathbf{V}_\lambda \times \mathbf{V}_{\lambda'}$ pasa por dos puntos de fuga $\mathbf{V}_\lambda, \mathbf{V}_{\lambda'}$. En este caso, los elementos a estimar son la *expresión analítica* de la transformación afín

$$\mathbf{A}(3; \mathbf{R}) := \left\{ A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & t_x \\ a_{21} & a_{22} & a_{23} & t_y \\ a_{31} & a_{32} & a_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \mid \det(A) \neq 0 \right\}$$

Cada par de puntos homólogos da lugar a 3 ecuaciones independientes; esta matriz tiene *once grados de libertad* (doce salvo escala) por lo que bastan 4 puntos y sus homólogos (correspondencia entre tetraedros, p.e.) para determinar una transformación afín espacial. La *propiedad invariante* a estimar es el paralelismo; la perpendicularidad no se conserva (una transformación afín cambia un cubo en un paralelepípedo no necesariamente rectángulo, p.e.). Los elementos geométricos a estimar son el plano del horizonte Π_∞ que pasa por tres puntos de fuga, la ratio de volúmenes, áreas y longitudes sobre elementos geométricos paralelos de la dimensión correspondiente.

2.3.2. Transformaciones de semejanza

Una *transformación de semejanza* es una transformación afín que conserva distancias y ángulos salvo factor de escala y transformaciones rígidas. El *objetivo* es la recuperación de la estructura métrica salvo transformación de semejanza.

La descripción del grupo de las transformaciones de semejanza como subgrupo de $GL(n+1; \mathbb{R})$ se obtiene a partir del grupo euclídeo que está dado por el conjunto de matrices

$$\mathbb{E}(n; \mathbb{R}) := \left\{ \begin{pmatrix} A & \mathbf{v} \\ 0 & 1 \end{pmatrix} \mid A \in SO(n; \mathbb{R}), \mathbf{v} \in \mathbb{R}^n \right\}$$

salvo escala. Como producto semidirecto, la operación (ley interna de grupo) de las transformaciones de semejanza está dada por la asignación

$$(A, \mathbf{t}_1) \cdot (B, \mathbf{t}_2) := (AB, \mathbf{t}_1 + A\mathbf{t}_2)$$

salvo escala. La *acción del grupo de semejanza* está dada por la multiplicación a la izquierda en la forma habitual:

$$\mathbb{E}(n; \mathbb{R}) \times \mathbb{A}^n \rightarrow \mathbb{A}^n \quad | \quad (A, \mathbf{v}) \cdot (P, 1)^T = AP + \mathbf{v}$$

salvo escala.

Las expresiones anteriores se particularizan al caso $n = 2$ de la forma obvia. En particular, la *expresión analítica* de una transformación de semejanza en el plano afín está dada por

$$\left\{ \begin{pmatrix} \lambda \cos \theta & -\lambda \sin \theta & t_x \\ \lambda \sin \theta & \lambda \cos \theta & t_y \\ 0 & 0 & 1 \end{pmatrix} \mid \lambda \neq 0 \right\} = \left\{ \begin{pmatrix} \lambda \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{pmatrix} \mid \lambda \neq 0 \right\}$$

Una transformación de semejanza tiene 4 *grados de libertad*. Por ello, bastan 2 puntos y sus homólogos (correspondencia entre segmentos) para determinar una semejanza plana.

Los *invariantes de las transformaciones de semejanza* son:

- Las longitudes de segmentos (salvo escala para el caso más general de semejanza) y las medidas que se puedan derivar de ellas (áreas, volúmenes).
- Los ángulos entre rectas incidentes.
- Los puntos circulares en la cónica absoluta $\mathbf{I}, \mathbf{J} = (1 : \pm i : 0)$

2.3.3. Isometrías

Una *isometría* es una transformación que conserva distancias, ángulos y áreas (no necesariamente la orientación). Nótese que las reflexiones son isometrías; hay grupos de reflexión no sólo en el espacio euclídeo, sino también sobre esferas o espacios hiperbólicos que resultan de interés para visualizar fenómenos (estáticos o dinámicos) sobre superficies curvadas; un caso particular importante corresponde a la visualización sobre cúpulas semi-esféricas que se comenta en el capítulo 4 en relación con extensiones de la reconstrucción a partir de una sola vista. La utilización de simetrías generadas por reflexiones facilitan el diseño e implementación de fenómenos de propagación elementales. No obstante, en este capítulo introductorio nos restringimos a los casos más sencillos.

El *Grupo de transformaciones* que resulta inicialmente para el análisis de las isometrías es el grupo especial euclídeo

$$SE(2) = \left\{ \begin{pmatrix} \cos \theta & \sin \theta & t_x \\ -\sin \theta & \cos \theta & t_y \\ 0 & 0 & 1 \end{pmatrix} \mid \theta \in [0, 2\pi] \right\} = \left\{ \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{pmatrix} \right\}$$

Los grupos de reflexión son subgrupos discretos Γ de las isometrías; por ello, están generados por un número finito de isometrías. El tratamiento geométrico de la información se realiza en términos de un “dominio fundamental” que se replica sobre la variedad soporte, generando una teselación (descomposición en polígonos) que recubre el objeto. Este hecho facilita el rellenado para regiones donde no se tiene información (geométrica o radiométrica) suficiente.

Como tiene 3 *grados de libertad* bastan 2 puntos y sus homólogos (correspondencia entre segmentos orientados) para determinar una isometría plana.

2.4. Descomposición de transformaciones geométricas

Cada geometría tiene un grupo estructural de transformaciones geométricas que la caracteriza que *nunca* es un espacio cartesiano. Para el caso lineal, es necesario considerar subgrupos del grupo lineal general $GL(n+1; \mathbb{R})$ dando lugar a diferentes geometrías:

1. El caso *euclídeo*: rotaciones y traslaciones
2. El caso *afín*: rotaciones y deformaciones

3. El caso *proyectivo*: desplazamientos de ℓ_∞ y semejanza

En el caso no-lineal se tienen grupos de transformaciones continuas más generales cuyas entradas están dadas por funciones continuas (polinomios de grado bajo para facilitar su implementación). Estas transformaciones permiten "curvar" los objetos, lo cual resulta de utilidad para reconocer "formas próximas" en relación con patrones dadas por plantillas deformables, p.e. (ver módulo 4 del CEViC para más detalles). Para simplificar en este módulo nos restringimos al caso lineal

La *estrategia general* para la *descomposición de transformaciones lineales* utiliza técnicas de desacoplamiento basadas en SVD (Singular Value Decomposition) que reducen la descomposición a un problema de valores propios. En los apartados siguientes se desarrolla esta estrategia para las transformaciones en el marco afín y proyectivo.

Descomposición de transformaciones afines

Una estrategia general para el desacoplamiento de transformaciones afines parte de la siguiente observación: Una transformación afín A descompone en *composición de rotaciones y cambios de escala no-isótropos*:

$$A = \mathbf{R}(\theta) \cdot [\mathbf{R}(-\phi) \cdot \mathbf{D} \cdot \mathbf{R}(\phi)]$$

La estrategia de resolución tiene los pasos siguientes:

1. *Idea*: Utilizar SVD: $A = UDV^\top = (UV^\top)(VDV^\top)$ pues $U, V \in O(3)$.
2. *Interpretación*: Rotación y deformación (transformación de cizalladura).
3. *Área orientada escalar*: $\det(A) = \det(D) = \lambda_1\lambda_2$ donde λ_i son valores propios del endomorfismo A .
4. *Area orientada vectorial*: $A(\mathbf{v}_1 \wedge \mathbf{v}_2) = \lambda_1\lambda_2\mathbf{v}_1 \wedge \mathbf{v}_2$.

2.4.1. Descomposición de transformaciones proyectivas

Una estrategia general para descomponer una transformación proyectiva H tiene los pasos siguientes:

1. $\mathbf{H} = \mathbf{H}_S \mathbf{H}_A \mathbf{H}_P = \begin{pmatrix} \lambda \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{pmatrix} \begin{pmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}^\top & u \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^\top & u \end{pmatrix}$
donde $\mathbf{A} = \lambda \mathbf{R} \mathbf{K} + \mathbf{t} \mathbf{v}^\top$ y \mathbf{K} es una matriz triangular superior con $\det(\mathbf{K}) = 1$.

2. *Interpretación* (Hartley y Zisserman, 2000):

- \mathbf{H}_P "mueve" la línea del infinito (2 DOF).
- \mathbf{H}_A deja fija la línea del infinito pero cambia las propiedades afines (2 DOF).
- \mathbf{H}_S es una transformación de semejanza que no afecta a propiedades afines o proyectivas (4 DOF).

2.5. Estimación de la métrica euclídea

2.5.1. Estimación de la dual de la cónica absoluta

Denotemos mediante ℓ, ℓ' dos líneas en el plano $\Pi_{\mathbb{C}}$ y denotemos mediante α el ángulo que forman. Sean $\mathbf{A} = \ell \cap q_{\infty}$ y $\mathbf{A}' = \ell' \cap q_{\infty}$ los puntos de intersección de ambas líneas con la cónica absoluta q_{∞} . Entonces, los cuatro puntos $\mathbf{A}, \mathbf{A}', \mathbf{I}$ y \mathbf{J} están *alineados* (verificarlo!), donde \mathbf{I}, \mathbf{J} son nuevamente los puntos circulares del infinito. Por ello, tiene sentido calcular su *razón doble* $\{\mathbf{A}, \mathbf{A}'; \mathbf{I}, \mathbf{J}\}$

El ángulo α que forman dichas rectas se calcula mediante la *fórmula de Laguerre*:

$$\alpha = \frac{1}{2i} \log(\{\mathbf{A}, \mathbf{A}'; \mathbf{I}, \mathbf{J}\})$$

Tomando \mathbf{A} y \mathbf{A}' como referencia afín sobre la recta podemos escribir

$$\mathbf{I} = \mathbf{A} + \lambda \mathbf{A}' \quad , \quad \mathbf{J} = \mathbf{A} + \lambda' \mathbf{A}$$

por lo que

$$\{\mathbf{A}, \mathbf{A}'; \mathbf{I}, \mathbf{J}\} = \frac{\lambda}{\lambda'} = e^{2i\alpha}$$

Como $\mathbf{I} \in q_{\infty}$ se verifica que

$$\lambda^2 {}^{\top} \mathbf{A}' \mathbf{Q} \mathbf{A}' + 2\lambda {}^{\top} \mathbf{A}' \mathbf{Q} \mathbf{A} + {}^{\top} \mathbf{A} \mathbf{Q} \mathbf{A} = 0$$

Como $\mathbf{J} \in q_{\infty}$ y $\frac{\lambda}{\lambda'} = e^{2i\alpha}$ se verifica que

$$\lambda^2 e^{4i\alpha} {}^{\top} \mathbf{A}' \mathbf{Q} \mathbf{A}' + 2\lambda e^{2i\alpha} {}^{\top} \mathbf{A}' \mathbf{Q} \mathbf{A} + {}^{\top} \mathbf{A} \mathbf{Q} \mathbf{A} = 0$$

Eliminando t^2 entre ambas ecuaciones se obtiene

$$\lambda^2 {}^{\top} \mathbf{A} \mathbf{Q} \mathbf{A}' \beta(\beta - 1) + {}^{\top} \mathbf{A} \mathbf{Q} \mathbf{A} (\beta^2 - 1) = 0 \quad \text{con} \quad \beta = e^{2i\alpha}$$

por lo que

$$\lambda = -\frac{\beta + 1}{2\beta} \frac{{}^{\top} \mathbf{A} \mathbf{Q} \mathbf{A}}{{}^{\top} \mathbf{A} \mathbf{Q} \mathbf{A}'}$$

3. Modelos geométricos de cámaras

En Visión Computacional, los modelos geométricos de cámara se refieren a los diferentes marcos geométricos asociados a una proyección. La proyección central lineal más general corresponde al marco proyectivo y se representa mediante $\pi_{\mathbf{C}}^p : \mathbb{P}^3 \rightarrow \mathbb{P}^2$. La selección de un hiperplano del infinito (un plano para \mathbb{P}^3 y una recta para \mathbb{A}^3) da lugar a una proyección afín que se representa mediante $\pi_{\mathbf{C}}^a : \mathbb{A}^3 \rightarrow \mathbb{A}^2$. Si ahora se fija una cuádrica absoluta en el espacio proyectivo, se obtiene la versión euclídea de la proyección $\pi_{\mathbf{C}}^e : \mathbb{E}^3 \rightarrow \mathbb{E}^2$.

Fijadas referencias en los espacios de partida y llegada, todas estas proyecciones se representan mediante una 3×4 -matriz con restricciones cada vez más fuertes. De una forma esquemática se puede representar la jerarquía geométrica entre estos “tipos geométricos de cámaras” mediante

$$\begin{array}{ccc} \pi_{\mathbf{C}}^p : \mathbb{P}^3 & \rightarrow & \mathbb{P}^2 \\ & \uparrow & \uparrow \\ \pi_{\mathbf{C}}^a : \mathbb{A}^3 & \rightarrow & \mathbb{A}^2 \\ & \uparrow & \uparrow \\ \pi_{\mathbf{C}}^e : \mathbb{E}^3 & \rightarrow & \mathbb{E}^2 \end{array}$$

donde las flechas horizontales denotan proyecciones con centro en un punto \mathbf{C} (para el modelo simplificado de cámara tipo pinhole) y las flechas verticales \uparrow denotan inclusiones. El objetivo de esta sección es facilitar elementos adicionales para comprender el diagrama anterior. Para ello, se ha adoptado una organización en las subsecciones siguientes:

1. La jerarquía geométrica básica para el caso lineal.
2. La cámara proyectiva
3. Otros tipos geométricos de cámaras
4. Recuperación de la estructura afín
5. Recuperación de la estructura métrica

La cuestión crucial es la estimación y la detección de elementos significativos sencillos en la escena que permitan extrapolar la información (proyectiva, afín, euclídea) a la mayor parte de la escena documentada. El pegado se aborda más adelante

La jerarquía geométrica básica

Los marcos matemáticos para el modelado geométrico lineal están siendo desarrollados desde hace al menos 2500 años. La primera formalización que incluye construcciones de modelos y una deducción explícita de resultados a partir de postulados se desarrolla en los libros de Geometría de Euclides, hace algo más de 2300 años. Esta formalización se utilizó en construcciones griegas

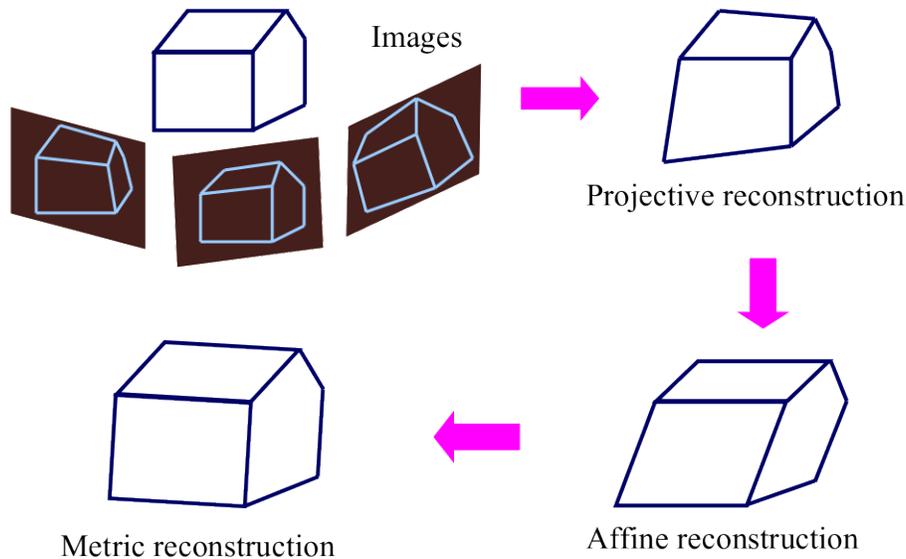


Fig. 7: Autocalibración de la cámara aprovechando los diferentes modelos geométricos

y romanas, llegando incluso a corregir una aplicación estricta de la geometría para adecuarla a la percepción visual como ocurre en el Partenón, p.e.

El marco más general para el modelado estructural es la Teoría Geométrica de Invariantes, contexto desarrollado a partir de la segunda mitad del siglo XIX desde un punto de vista algebraico. Los intentos de adaptación de este marco general a la Visión Computacional a principios de los noventa (1991-93) en relación con la interacción entre Reconstrucción 3D y Reconocimiento no dieron el resultado esperado debido, en buena parte, a la dificultad para estimar los invariantes asociados a este enfoque para el caso polinomial asociado a la Geometría Algebraica. Por ello, nos restringimos al caso de geometrías lineales.

La primera unificación de los aspectos matemáticos del modelado geométrico lineal fue desarrollada por Félix Klein. En su programa de Erlangen (1873), Klein mostró la forma de asociar una geometría clásica al grupo de matrices que dejan invariante las propiedades características de dicha geometría (conservación de distancias o de ángulos, de proporciones o de elementos del infinito, p.e.) y, recíprocamente. De este modo, F.Klein caracterizó cada geometría en términos de su "grupo estructural" dado por matrices que conservan cierta propiedad característica vinculada a una forma cuadrática real (distancia euclídea, p.e.) o compleja (distancia hermítica) o bien a una forma bilineal (geometría simpléctica, p.e.). La geometría epipolar se puede interpretar como la asociada a la conservación de una forma bilineal (la restricción epipolar). Por ello, el enfoque basado en la conservación de formas multilineales invariantes (tensores) es la extensión natural del enfoque de F.Klein.

La jerarquía geométrica es equivalente a la jerarquía entre grupos de transformaciones lineales (Klein, 1873). Esta subsección se organiza en torno a tres tópicos que son significativos para la Reconstrucción 3D (para otros desarrollos ver cualquier curso de Grupos Clásicos; para la conexión con aspectos diferenciales, ver el módulo 2 de mis notas de Geometría Diferencial).

1. *Elementos de Grupos clásicos*
2. *Descomposición Polar*
3. *Grupos y Geometrías*

Elementos de grupos clásicos

Un grupo clásico es un subgrupo continuo de $GL(n; \mathbb{K}) := \{A \in GL(n; \mathbb{K}) \mid \det(A) \neq 0\}$ para $\mathbb{K} = \mathbb{R}$ o \mathbb{C} . Los grupos clásicos más importantes son:

1. *Especial lineal*: $SL(n; \mathbb{K}) := \{A \in GL(n; \mathbb{K}) \mid \det(A) = 1\}$: conservan el volumen orientado.
2. *Ortogonal*: $O(n) := \{A \in GL(n; \mathbb{R}) \mid A^T A = I\}$: conservan la ortogonalidad.
3. *Unitario*: $U(n) := \{A \in GL(n; \mathbb{C}) \mid \bar{A}^T A = I\}$: conservan la ortogonalidad conjugada.
4. *Simpléctico*: $Sp(n) := \{A \in GL(2n; \mathbb{R}) \mid A^T J A = J\}$: \rightarrow cinemática.
5. *Especial ortogonal*: $SO(n) := O(n) \cap SL(n; \mathbb{R})$: rotaciones euclídeas para $m = 2$ y $n = 3$.
6. *Especial unitario*: $SU(n) := U(n) \cap SL(n; \mathbb{C})$: rotaciones hiperbólicas para $m = 2$.
7. *Euclídeo*: $SE(n)$: producto semidirecto de $SO(n)$ y las traslaciones \mathbb{R}^n

Cada grupo clásico G determina una geometría lineal caracterizada por la conservación de una métrica (euclídea, hermítica, riemanniana) o una forma bilineal (grupo simpléctico asociado a la Mecánica Analítica). En particular, la acción $\alpha : G \times X \rightarrow X$ (por multiplicación a la izquierda, p.e.) de un grupo G sobre un espacio X permite trasladar los resultados que se tienen en un punto a un entorno (localmente homogéneo) o a todo el espacio (globalmente homogéneo).

La posibilidad de describir cualquier geometría lineal en términos de un grupo clásico simplifica el análisis (local o global) de objetos/escenas y, al mismo tiempo, proporciona modelos básicos de movimiento (caso de los movimientos rígidos), percepción visual (transformaciones afines o proyectivas) o propagación (transformaciones simplécticas locales). De ahí la importancia de los grupos clásicos en los modelos de reconstrucción lineales a trozos. A pesar de que la

correspondencia estructural entre grupos clásicos y geometrías lineales es conocida en Matemáticas desde el programa de Erlangen (F.Klein, 1873) más allá de la Físico, su utilización en áreas de la Ingeniería diferentes de la Robótica es aún relativamente escaso.

Descomposición polar

La descomposición de matrices en producto de matrices pertenecientes a algunos de los grupos clásicos mencionados más arriba, permite desacoplar fenómenos no-lineales complicados y estimar por separado cada una de las componentes del producto. En este apartado se mencionan resultados de cálculo matricial que son significativos para cuestiones de Geometrías Lineales vinculadas a problemas de Reconstrucción y de una manera más específica con problemas de Optimización. La Descomposición en Valores Singulares (SVD) es una de las técnicas utilizadas con más frecuencia en diferentes aplicaciones.

1. $\forall A \in GL(n; \mathbf{R}) \exists V \in O(n), S \in Sym(n) \mid A = V.S$. Idea: descomponer en rotación y deformación
2. $Sym(n)$ es contractible $\Rightarrow GL(n; \mathbf{R})$ se contrae a $O(n)$.
3. Interpretación: ortonormalización de Gram-Schmid.
4. Problema de optimización: Minimizar $|A - V|^2$.
5. $|A - V|^2 = |A|^2 - 2(A, V) + |V|^2 = |A|^2 - tr(A^T V) + |V|^2$
6. Reducción: Maximizar $tr(A^T V)$ mediante SVD, descomposición polar o cuaterniones.
7. La descomposición polar es única $\forall A \in GL(n)$

Debido a su utilidad y su presencia ubicua, a continuación recordamos algunos hechos básicos de la SVD:

1. Si $A \in GL(n)$ la SVD es una descomposición polar con respecto a los ejes principales de la componentes simétrica S de A .
2. *Inconv*: La SVD no es única si los valores propios no son distintos, a diferencia de la descomposición polar que efectivamente es única.
3. La SVD se aplica a matrices arbitrarias (no necesariamente transformaciones de un espacio).

Grupos y Geometrías

F. Lein en su programa de Erlangen (1873) fue el primero en desarrollar la correspondencia entre grupos clásicos y geometrías lineales. Los "ejemplos" básicos para dichas geometrías corresponden al espacio euclídeo \mathbb{E}^n , afín \mathbb{A}^n y proyectivo \mathbb{P}^n . En todos los casos se tienen los "hechos" siguientes:

1. Cada geometría clásica está definida por un grupo clásico G y recíprocamente.
2. Acción a la izquierda: $G \times V \rightarrow V \mid (A, \mathbf{v}) \mapsto A\mathbf{v}$
3. $G \times f(A^{-1}\underline{x}) \rightarrow \mathbf{K}[\underline{x}] \mid (A, f(\underline{x})) \mapsto f(A^{-1}\underline{x}) \forall A \in G, \forall f \in \mathbf{K}[\underline{x}]$
4. $f \in \mathbf{K}[\underline{x}]$ es G -invariante si $f(A^{-1}\underline{x}) = f(\underline{x})$.
5. Una geometría está caracterizada por un sistema completo de polinomios G -invariantes $\mathbf{K}[\underline{x}]^G$
6. *Problema:* Estimación de polinomios G -invariantes \rightarrow Reconstrucción 3D \rightarrow Reconocimiento?

Grupos y Geometrías \rightarrow Extensión al caso cinemático (S.Lie) \rightarrow Deformaciones (Topología Diferencial)

3.1. Modelos geométricos de proyección

Los modelos geométricos afectan a la visualización de la escena proporcionada por los dispositivos de proyección (entendida como una aplicación que representa el cono proyectante) y a la imagen como resultado de esa proyección. Recíprocamente, la visualización de un objeto 3D a partir de una o varias vistas requiere cortar un cono proyectante variable con el plano de la cámara eventualmente móvil (proyección inversa o retroproyección).

Los modelos relativos a la escena se describen en términos de los diferentes tipos de proyección sobre un plano o sobre una cuádrica; aquí nos centraremos en los modelos planares. No obstante la proyección estereográfica sobre una esfera tiene una gran importancia para diferentes aspectos relacionados con una representación intrínseca de las rectas asociadas a una escena. Asimismo, la proyección sobre un paraboloide es muy útil para recuperar información métrica a partir de la información afín y recíprocamente. Empezamos con el modelo más simple de proyección del plano sobre una recta:

3.1.1. Proyección perspectiva 1D

El haz de líneas contenidas en un plano que pasa por un punto es una recta proyectiva que proporciona el modelo más simple de proyección perspectiva 1D. En este apartado se presenta una descripción algebraica de esta construcción para facilitar la implementación computacional.

La proyección de un punto $\mathbf{P}_0 = (X_0, Y_0)^\top$ del plano desde el origen sobre la línea $y = f$ paralela al eje OX , da lugar al punto \mathbf{p}_0 de coordenadas $(x_0, y_0)^\top = (\frac{X_0}{Y_0}f, f)^\top = f(\frac{X_0}{Y_0}, 1)^\top$. Los términos de la fracción que expresa la coordenada de la proyección sobre la línea $Y = f$ son lineales. La coordenada homogénea (salvo factor de escala) del punto \mathbf{p}_0 se puede expresar en términos de las coordenadas homogéneas $(X_0, Y_0, 1)$ asociada a \mathbf{P}_0 como un producto de dos transformaciones lineales correspondientes a la proyección sobre la línea

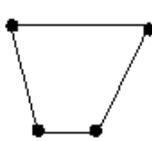
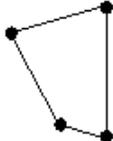
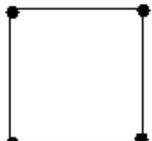
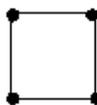
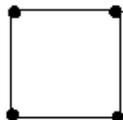
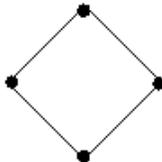
Transformation	Before	After
Projective		
Affine		
Similarity		
Euclidean		

Fig. 8: Jerarquía de transformaciones afines desde las euclídeas (sólo rotación y traslación) hasta las proyectivas donde un cuadrado puede transformarse en un cuadrilátero cualquiera

$Y = f$ y la transformación de perspectiva en la línea imagen (dibujarlo). La expresión analítica es la siguiente:

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/f & 1 \end{pmatrix}}_{\text{Proyección perspectiva}} \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -f \\ 0 & 0 & 1 \end{pmatrix}}_{\text{Traslación a línea imagen}} \begin{pmatrix} X_0 \\ Y_0 \\ 1 \end{pmatrix}$$

(verificad como ejercicio que $x_0 := \frac{x_0}{y_0} = f \frac{X_0}{Y_0}$). En general, la proyección perspectiva 1D es algo más complicada pues los objetos no tienen por qué estar en posición fronto-paralela. Una primera aproximación euclídea al problema consiste en suponer que la línea de proyección está rotada un ángulo θ y se encuentra situada a una distancia f en el nuevo sistema de coordenadas. Verificar como ejercicio que las coordenadas homogéneas del punto imagen están dadas por

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/f & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -f \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_0 \\ Y_0 \\ 1 \end{pmatrix}$$

donde la primera matriz que actúa sobre el punto $\mathbf{P}_0 = (X_0, Y_0, 1)^\top$ corresponde a una rotación de ángulo θ . Nótese que en este caso la coordenada proyectiva sobre la línea está dada por

$$z_0 = \frac{x_0}{y_0} = f \frac{X_0 \cos \theta + Y_0 \sin \theta}{-X_0 \sin \theta + Y_0 \cos \theta} = f \frac{Z_0 \cos \theta + \sin \theta}{-Z_0 \sin \theta + \cos \theta},$$

donde $Z_0 = \frac{X_0}{Y_0}$ que nos muestra la coordenada del punto \mathbf{p}_0 en el modelo de perspectiva 1D como la imagen por un caso particular de una transformación fraccionaria (una transformación conforme en el caso complejo). Recordemos que las transformaciones fraccionarias están inducidas por elementos del grupo especial lineal del plano

$$SL(2) := \{A \in GL(2; \mathbf{K}) \mid \det(A) = ad - bc = 1\}$$

y que $SE(2)$ está incluido en el producto semidirecto de $SL(2)$ por el grupo de traslaciones \mathbf{R}^2 . Esta observación pone de manifiesto la inclusión de la geometría euclídea en el modelo de perspectiva más simple posible, resultado válido en dimensión arbitraria.

3.1.2. Proyecciones planares

La *proyección planar ideal* $\pi_{\mathbf{C}} : \mathbf{R}^3 \rightarrow \mathbf{R}^2$ con centro en un punto $\mathbf{C} = (X_0, Y_0, Z_0)$ correspondiente a una cámara tipo agujero de alfiler (pinhole) se describe mediante la intersección con el plano de imagen del haz de rectas en el espacio ordinario 3D que pasan por \mathbf{C} (este haz de rectas es un plano proyectivo). Interesa disponer de una descripción de la proyección para el marco

proyectivo $\pi_{\mathbf{C}}^p$, afín $\pi_{\mathbf{C}}^a$ y euclídeo $\pi_{\mathbf{C}}^e$. Como es habitual, empezamos realizando una descripción en el espacio cartesiano; a continuación se superpone la estructura geométrica adicional correspondiente a cada uno de los marcos geométricos usuales.

Si el plano de la cámara está situado a una distancia $\pm d$ (depth), todos los puntos que se encuentran en la dirección de $\mathbf{P} = (X, Y, Z)$ se proyectan sobre $\mathbf{p} = (x, y, \pm d)$. Por tanto, la proyección $\pi(\mathbf{P}) = \mathbf{p}$ está definida por la asignación

$$(X, Y, Z) \mapsto (x, y) = \pm t \left(\frac{X}{Z}, \frac{Y}{Z} \right)$$

Razonando de forma similar al caso anterior, se obtiene

$$\begin{pmatrix} x \\ y \\ d \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -f \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & -X_0 \\ 0 & 1 & 0 & -Y_0 \\ 0 & 0 & 1 & -Z_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

donde la transformación situada más a la izquierda es la proyección perspectiva, la segunda es la traslación a la proyección y la tercera la traslación a la lente (dibujarlo y verificar los detalles como ejercicio) En este caso, realizando la multiplicación del segundo miembro vemos que es igual a $(X - X_0, Y - Y_0, \frac{1}{f}[Z - Z_0])^T$ por lo que las coordenadas afines del punto imagen en el abierto $D_+(d)$ se escriben como

$$\frac{x}{d} = f \frac{X - X_0}{Z - Z_0}, \quad \frac{y}{d} = f \frac{Y - Y_0}{Z - Z_0}$$

La introducción de coordenadas homogéneas $[X : Y : Z : W]$ para cada punto $\mathbf{P} \in \mathbf{P}^3$ permite escribir la matriz P de una *proyección perspectiva* $\pi(\mathbf{P}) = \mathbf{p}$ de una forma más compacta como

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1/d \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda X \\ \lambda Y \\ \lambda Z \\ \lambda W \end{pmatrix} = \lambda \begin{pmatrix} X \\ Y \\ Z \\ \frac{d-Z}{d} \end{pmatrix} \mapsto \lambda \begin{pmatrix} x \\ y \end{pmatrix} = \frac{d}{d-Z} \begin{pmatrix} X \\ Y \end{pmatrix}$$

Una *transformación proyectiva* M descompone en un producto $TR_xR_yR_xPSC$ de transformaciones elementales donde P es la matriz de proyección perspectiva que acabamos de describir, T es una traslación por el vector $(-T_1, -T_2, -T_3)$, S es el factor de escala y R_x (resp. R_y) es la rotación con respecto al eje Ox de ángulo Θ (resp., eje Oy de ángulo φ , que escribimos en coordenadas homogéneas como

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -T_1 & -T_2 & -T_3 & 1 \end{pmatrix}, \quad S = \begin{pmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & s_3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

y

$$R_x = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \Theta & \sin \Theta & 0 \\ 0 & -\sin \Theta & \cos \Theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad R_y = \begin{pmatrix} \cos \varphi & 0 & \sin \varphi & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \varphi & 0 & \cos \varphi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Esta descomposición muestra cómo obtener los parámetros de las transformaciones de la cámara en las coordenadas mundo de forma no-iterativa a partir de un conjunto de puntos de calibración cuyas posiciones en el espacio son perfectamente conocidas (Strat, 1984). De este modo, podemos obtener los 11 parámetros de calibración absoluta de la cámara tanto los 6 externos (posición-orientación) como los 5 internos (más adelante volveremos sobre ello).

Cámara proyectiva

Una cámara tipo pinhole se describe geoméricamente en términos del haz de rectas que pasan por un punto ideal de la cámara al que se llama *foco* o centro de la cámara y se denota mediante \mathbf{C} . Los elementos que caracterizan a una cámara de este tipo son los siguientes:

1. *Proyección central* $\pi_{\mathbf{F}} : \mathbf{P} = (X, Y, Z) \mapsto \mathbf{p} = (x, y)$ desde un *foco* \mathbf{F} sobre un plano imagen $\Pi_{\mathbf{F}}$ (plano de la cámara)
2. *Rayo principal*: recta ℓ perpendicular al plano imagen que pasa por el foco:

$$\ell \perp \pi_{\mathbf{F}}, \quad \mathbf{F} \in \ell$$

3. *Longitud focal*: f Distancia del foco al plano imagen
4. *Punto principal*: $\ell \cap \pi_{\mathbf{F}}$ del rayo principal con el plano imagen; frecuentemente se toma dicho punto principal como el origen.
5. *Semejanza de triángulos* $\Rightarrow \mathbf{p} = (f \frac{X}{Z}, f \frac{Y}{Z})$.

3.1.3. Una proyección particular

La expresión analítica de la proyección central $\pi_{\mathbf{F}}$ en *coordenadas euclídeas* con respecto a las referencias apropiadas es:

$$\begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Nótese que si $\mathbf{M}_{\mathbf{C}}$ es la matriz de la proyección entonces

$$\mathbf{C} = \text{Ker}(\mathbf{M}_{\mathbf{C}}) = [0 : 0 : 0 : 1]^{\top}$$

Por otro lado, si el punto principal es (p_x, p_y) entonces las coordenadas del punto imagen dadas como la imagen de las coordenadas espaciales se expresan como

$$\pi_{\mathbf{C}}(\mathbf{P}) \stackrel{(e)}{=} \mathbf{M}_{\mathbf{C}}^e(X, Y, Z, 1)^\top = \left(f \frac{X}{Z} + p_x, f \frac{Y}{Z} + p_y, 1 \right)^\top$$

Nótese que las coordenadas del plano imagen deben ser reconvertidas a coordenadas píxel en las que el origen de coordenadas se supone que está en la esquina superior izquierda de la imagen; esta reconversión se lleva a cabo mediante una traslación y un cambio de sentido en la orientación del eje vertical.

3.1.4. Expresión analítica general

Una proyección central $\pi_{\mathbf{C}} : \mathbb{P}^3 \rightarrow \mathbb{P}^2$ del espacio sobre el plano proyectivo de imagen $\Pi_{\mathbf{C}} \simeq \mathbb{P}^2$ es una aplicación lineal que a cada punto $\mathbf{P} \in \mathbb{P}^3$ le asigna su proyección $\mathbf{p} \in \mathbb{P}^2$. Dicho punto se obtiene como la intersección $\ell_{\mathbf{P}} \cap \Pi_{\mathbf{C}}$ de la línea $\ell_{\mathbf{P}} := \langle \mathbf{C}, \mathbf{P} \rangle$ con el plano de imagen $\Pi_{\mathbf{C}} \simeq \mathbb{P}^2$.

Fijadas referencias en los espacios proyectivos de partida y llegada de la proyección central $\pi_{\mathbf{C}}^p$, entonces la proyección central se expresa mediante una $3 \times r$ -matriz $\mathbf{M} = (m_{ij})$ regular (es decir, de rango máximo igual a 3 en este caso) de la forma

$$\mathbf{M} := \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix}$$

que está definida salvo factor de proporcionalidad. que representamos en forma vectorial más sintética como:

$$\mathbf{M}\mathbf{P} = \mathbf{p}$$

donde $\mathbf{P} = [X_1 : X_2 : X_3 : X_4]^\top$ y $\mathbf{p} = [x_1 : x_2 : x_3]^\top$ representan las coordenadas homogéneas de los puntos (definidas salvo factor de proporcionalidad).

3.1.5. Representación afín

Para analizar la representación de la proyección central $\pi_{\mathbf{C}}^a : \mathbb{A}^3 \rightarrow \mathbb{A}^2$ en el caso afín, podemos suponer que lleva el hiperplano del infinito $X_4 = 0$ en la línea del infinito $x_3 = 0$. Por ello, el espacio afín \mathbb{A}^3 esta caracterizado por la condición $D_+(X_4) := \{\mathbf{P} \in \mathbb{P}^3 \mid X_4 \neq 0\}$, es decir, podemos dividir entre X_4 ; análogamente, \mathbb{A}^2 esta caracterizado por la condición $D_+(x_3) := \{\mathbf{p} \in \mathbb{A}^2 \mid x_3 \neq 0\}$. Eligiendo las referencias estándar en ambos espacios afines, se obtiene una expresión para la matriz de proyección afín de la forma:

$$\mathbf{A}^a := \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

salvo factor de escala, donde $a_{ij} = m_{ij}/m_{34}$ para $i = 1, 2$ y $1 \leq j \leq 4$. La acción de esta transformación sobre un punto afín se representa ahora como

$$\mathbf{x}^a = \mathbf{A}^* \mathbf{X}^a + \mathbf{t}$$

donde $\mathbf{x}^a = (x, y, 1)^\top = (x_1/x_3, x_2/x_3, 1)^\top$ son las coordenadas afines del punto imagen en el plano de la cámara, $\mathbf{X}^a = (x, y, z, 1)^\top = (X_1/X_4, X_2/X_4, X_3/X_4, 1)^\top$ son las coordenadas afines del punto espacial,

$$\mathbf{A}^e = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

y \mathbf{t} es el vector que representa el "centro" de la imagen (proyección del foco sobre el plano de imagen Π_C que, en general, no coincide con el centro físico de la foto).

Ejercicio.- Comprobad analíticamente que la proyección afín conserva el paralelismo.

3.1.6. Representación euclídea

Desde un punto de vista matricial, la representación euclídea se puede visualizar como un caso particular de la representación afín presentada en el apartado anterior con restricciones adicionales para la matriz \mathbf{A}^* cuya primera 2×2 -caja debe ser ortonormal.

3.2. Reinterpretación matricial

En esta subsección se reinterpretan los elementos proyectivos en cámaras afines y métricas. Esta reinterpretación es motivada por las observaciones realizadas en el último apartado de la subsección anterior. De una forma un tanto general, podemos decir que las matrices de proyección y sus descomposiciones contienen información significativa sobre los elementos internos de la cámara y su localización. A continuación, tratamos de precisar esta afirmación.

El primer paso consiste en calcular el centro de proyección. El segundo paso consiste en analizar las columnas de la matriz de proyección (extendiendo el análisis llevado a cabo para la última columna en el caso afín), que es el objeto del primer apartado. A continuación, se obtiene el *eje principal* que proporciona información sobre la línea principal de visión. En el tercer apartado, se analiza la *orientación* de la cámara euclídea en relación con el último apartado analizado en la subsección anterior.

3.2.1. Cálculo del centro de proyección

Al realizar un movimiento (real o virtual) de cámara, el centro de proyección deja de estar en la posición ideal $[0 : 0 : 0 : 1]^\top$. Es necesario mostrar un método que nos permita estimar (al menos idealmente) la posición del centro de proyección a partir de los datos procedentes del análisis de imagen. Los datos más robustos están dados por líneas.. Por ello, el primer paso consiste en estimar

la *imagen de una recta* $(1 - \lambda)\mathbf{C} + \lambda\mathbf{P}$ a través del centro \mathbf{C} de proyección que denotamos mediante

$$\mathbf{M}_{\mathbf{C}}[(1 - \lambda)\mathbf{C} + \lambda\mathbf{P}] = \lambda\mathbf{M}_{\mathbf{C}}(\mathbf{P})$$

La estimación del centro de proyección sigue los pasos siguientes:

1. *Imagen de una recta* $(1 - \lambda)\mathbf{C} + \lambda\mathbf{P}$ a través del centro \mathbf{C} de proyección: $\mathbf{M}_{\mathbf{C}}[(1 - \lambda)\mathbf{C} + \lambda\mathbf{P}] = \lambda\mathbf{M}_{\mathbf{C}}(\mathbf{P})$
2. Sea $P_{123} := \mathbf{M}_{\mathbf{C}}^{(123)}$ la submatriz formada por las 3 primeras columnas de $\mathbf{M}_{\mathbf{C}}$.
3. La *cámara es finita* si $\det(\mathbf{M}_{\mathbf{C}}^{(123)}) \neq 0$.
4. En general, para una *cámara finita*, el centro $\mathbf{C} = \begin{pmatrix} -P_{123}^{-1}\mathbf{c}_4(P) \\ 1 \end{pmatrix}$ donde $\mathbf{c}_4(P)$ es la cuarta columna de $\mathbf{M}_{\mathbf{C}}$
5. Para una *cámara infinita* $\det(P_{123}) = 0$ y $\mathbf{C} = \begin{pmatrix} \mathbf{d} \\ 1 \end{pmatrix}$ donde \mathbf{d} es la solución $\neq 0$ de $P_{123}(\mathbf{d}) = \mathbf{0}$

3.2.2. Columnas de la matriz de proyección

Podemos interpretar las columnas de la matriz de proyección para una cámara afín finita con $\det(A_{123}) \neq 0$ en términos de tres “planos coordenados” asociados a la proyección $\pi_{\mathbf{C}}$ que describimos a continuación de forma esquemática:

1. *Plano principal*: plano a través del centro de la cámara y paralelo al plano imagen

$$\{\mathbf{X} \in \mathbb{P}^3 \mid \mathbf{M}_{\mathbf{C}}(\mathbf{X}) = (x_1 : x_2 : 0)^\top\}$$

- $\mathbf{X} \in \text{Plano principal} \Leftrightarrow P^{3T}(\mathbf{X}) = \mathbf{0}$
 - El plano principal se calcula como $\text{Ker}(P_3) \Rightarrow$ La fila P_3^\top de $\mathbf{M}_{\mathbf{C}}$ representa el plano.
 - El plano principal está acoplado con la geometría de la cámara
2. El plano P_1 esta definido por \mathbf{C} y la línea $x_1 = 0$ en la imagen:
 - Sea \mathbf{X} t.q. $P_1\mathbf{X} = 0 \Rightarrow \mathbf{M}_{\mathbf{C}}(\mathbf{X}) = (0 : x_2 : x_3)^\top$
 3. El plano P_2 esta definido por \mathbf{C} y la línea $x_2 = 0$ en la imagen:
 - Sea \mathbf{X} t.q. $P_2\mathbf{X} = 0 \Rightarrow \mathbf{M}_{\mathbf{C}}(\mathbf{X}) = (x_1 : 0 : x_3)^\top$

La descripción precedente permite expresar los ”planos coordenados” de forma relativa con respecto al sistema interno:

1. Los *planos coordenados* P_1 y P_2 dependen de la visualización de los ejes Ox y Oy en la imagen
2. La línea $P_1 \cap P_2$ une \mathbf{C} y el origen de la imagen. En general, no coincide con el eje de la cámara.
3. *Expresión general del centro de la cámara* en términos de las columnas de la matriz $\mathbf{M}_{\mathbf{C}}$ de proyección $\mathbf{C} = [g_{234} : -g_{134} : g_{124} : -g_{123}]$ donde $g_{ijk} = \det(\mathbf{M}^{ijk})$ (g: Grassmann-Pluecker)

3.2.3. El eje principal

Es una línea a_{pr} a través de \mathbf{C} perpendicular al plano principal P_3 . Si denotamos a la *matriz de proyección* mediante $\mathbf{P} = (\mathbf{M}^{123} | \mathbf{P}^4)$, entonces el *punto principal* está dado por $\mathbf{x}_0 := a_{pr} \cap \Pi_{\mathbf{C}}$. Por ello

- \mathbf{x}_0 se calcula como $\mathbf{M}^{123} \mathbf{m}_3 = \mathbf{x}_0$ donde \mathbf{m}_3 es la fila 3 de \mathbf{M}_{123}
- *vector del eje principal* $\mathbf{v} = \det(\mathbf{M}^{123}) \mathbf{m}_3$ orientado hacia el frente de la cámara.

3.2.4. Cámara euclídea: Orientación

El objetivo de este último apartado es completar y dar un significado geométrico a la descripción matricial asociada a la versión euclídea de la proyección central $\pi_{\mathbf{C}}^e : \mathbb{E}^3 \rightarrow \mathbb{E}^3$

La *idea básica* consiste en utilizar la *RQ-descomposición* de una matriz regular

$$\mathbf{M}^{123} = \mathbf{K}\mathbf{R}$$

en producto de una triangular \mathbf{K} y una ortogonal \mathbf{R} ; para simplificar la notación, escribiremos \mathbf{M} (notación abreviada) \mathbf{M} en lugar de \mathbf{M}^{123} ³⁰. Con esta notación, la *matriz de proyección* se escribe ahora como

$$\mathbf{P} = [\mathbf{M} \mid -\mathbf{M}\mathbf{C}] = \mathbf{K}[\mathbf{R} \mid -\mathbf{R}\mathbf{C}]$$

con $\det(\mathbf{M}) \neq 0$. Para *suprimir la ambigüedad* se toman los valores propios de \mathbf{K} son positivos.

El *método numérico clásico* a implementar consiste en anular de forma recurrente las entradas de \mathbf{M} mediante giros manteniendo las condiciones de ortogonalidad.

Ejemplo: Para anular m_{21} mediante un giro de ángulo θ_3 con respecto al eje Oz resolver $cm_{21} + sa_{22} = 0 \Rightarrow c = -m_{22}$ (donde $c = \cos\theta_3$ y $s = \sin\theta_3$)

Ejercicio (avanzado).- Formular esta descripción utilizando cuaterniones (*Indicación:* Adaptar materiales de [Bay00])

³⁰ Esta descomposición es un caso particular de la descomposición de Iwasawa que aparece en la Teoría de Representación de Grupos de Lie

4. Visualización e interacción de modelos de perspectiva

Cualquier cámara genera de forma automática un modelo de perspectiva; en un modelo ideal de proyección central, los rayos pasan a través de una “pequeña apertura”, inciden sobre un objeto y se almacena en cada píxel la información asociada a dicha intersección. Aplicado un argumento de proporcionalidad entre triángulos semejantes el “tamaño” del objeto almacenado es, en una primera aproximación, inversamente proporcional a la profundidad relativa. Este hecho, introduce un factor de no-linealidad sobre la representación, aunque los elementos utilizados para la representación sean lineales (rayos de luz).

En la práctica, la proyección resulta distorsionada por las características de la lente; por tanto, el modelo de proyección central es sólo una aproximación tosca a la formación geométrica de imágenes. En efecto, la lente no es un punto salvo cuando se cierra casi completamente el diafragma, lo cual tiene como consecuencia la reducción del campo visual. Para una mayor apertura, aparecen distorsiones en la representación de las medidas de los objetos que se obtienen al proyectar sobre el plano de proyección (la retina en el caso humano). Estas distorsiones se corrigen estimando la “calibración” de la cámara, cuestión que se desarrolla en el capítulo 2 de este módulo.

Para fijar ideas en esta sección se adopta un enfoque geométrico muy simplificado basado en diferentes tipos de proyección para representar algunos de los modelos de perspectiva utilizados más frecuentemente. Los diferentes modelos se incluyen en un marco geométrico proyectivo, pues es el único que permite “integrar” las diferentes aproximaciones al problema.

Los modelos de perspectiva básicos son conocidos desde principios del siglo XV. El primer tratado se debe a Leon Battista Alberti, discípulo de Brunelleschi. Tras este tratado, algunos de los artistas más relevantes del Renacimiento mostraron diferentes metodologías para incorporar la profundidad con bellísimas ilustraciones (como las de A. Dürero) que ponen de manifiesto el profundo conocimiento que habían adquirido los artistas desde principios del siglo XV hasta principios del XVII. La primera formulación completa de estos principios desde el punto de vista matemático no llegó hasta el *Le Traité des Propriétés Projectives des Figures* de Poncelet (1822).

4.1. Elementos básicos de perspectiva

Inicialmente supondremos que:

- El centro de proyección coincide con el origen del sistema “mundo” W de coordenadas.
- El eje óptico coincide con la dirección Oz (dado por $x = 0, y = 0$) del sistema de coordenadas mundo.
- El plano de la cámara es ortogonal al eje óptico y pasa por el punto $(0, 0, -f)$ siendo f es la longitud focal.
- La intersección del eje óptico con el plano de imagen es el *punto principal*

Para modelar las transformaciones entre elementos comunes contenidos en diferentes vistas es necesario adoptar un marco proyectivo. El *modelo proyectivo de una vista 2D* más general es un plano proyectivo. De una forma intuitiva, un plano proyectivo es un plano ordinario al que hemos “añadido” todos los “puntos del infinito”. Como cada punto del infinito es el lugar de convergencia de un haz de líneas ℓ paralelas a una dirección ℓ_0 , los elementos a considerar en el plano proyectivo son (proyecciones de) haces de líneas paralelas. Cada una de dichas direcciones esta representada por (la proyección sobre el plano de imagen de) un haz de líneas paralelas que se representa mediante un punto de una esfera \mathbb{S}^2 . Por ello, formalmente el plano proyectivo se define como el conjunto de las direcciones del espacio ordinario 3D:

$$\mathbb{P}^2 := \mathbb{P}\mathbb{R}^3 := (\mathbb{R}^3 - \{0\}) / \sim \simeq \mathbb{S}^2 / \mathbb{Z}_2$$

donde el último miembro denota el cociente de la esfera 2-dimensional \mathbb{S}^2 por la acción antipodal generada por $\mathbb{Z}_2\{\pm 1\}$, es decir, la acción que identifica dos puntos diagonalmente opuestos (ambos representan la misma dirección en el espacio ordinario). En esta subsección se presentan algunos elementos básicos del plano proyectivo que facilitan la representación de modelos de perspectiva de cara a la reconstrucción de objetos o escenas tridimensionales.

4.1.1. Puntos de fuga

Un *punto de fuga* \mathbf{v}_i es un punto de la imagen 2D donde convergen la proyección sobre el plano imagen Π de líneas paralelas en la escena 3D. Por ello, en una vista puede haber hasta un total de 3 puntos de fuga *independientes* (en realidad cualquier punto del plano puede ser un punto de fuga). Las líneas paralelas que convergen en un punto de fuga reciben el nombre de *líneas de perspectiva* (abreviadamente LP).

Los puntos que contienen información de incidencia entre líneas se convierten en vértices para modelos orientados. Dichos vértices son “hechos geométricos” que admiten una etiqueta adicional como *junturas*. Las junturas típicas estables son dobles (en forma de L o de T) o triples (en forma de Y o de \uparrow) correspondientes a esquinas, típicamente. Las junturas se determinan de forma automática como intersección de “aristas largas”. Una “arista larga” se extrae de forma automática imponiendo una restricción de colinealidad a segmentos con extremos próximos. La longitud permitida para una arista larga debe ser un parámetro configurable por el usuario.

En una juntura triple del plano confluyen tres aristas largas que soportan líneas ℓ_i, ℓ_j, ℓ_k en \mathbb{P}^2 que son proyección de líneas L_i, L_j, L_k en \mathbb{P}^3 que son *no paralelas*; en la práctica, las tres líneas se cortan dos a dos en tres puntos muy próximos, por lo que es necesario identificar el “punto de corte” (procedimientos de “voto” y acotación del error) y retrazar las líneas.

El mismo argumento se utiliza para una estimación tosca de los puntos de fuga: Los puntos que son *candidatos a puntos de fuga* son *junturas cuádruples* en las que “de forma ideal” confluyen al menos 4 líneas que soportan “aristas

largas” que son *paralelas en la escena*. El agrupamiento de aristas largas en candidatas a líneas de perspectiva utiliza sólo la restricción de colinealidad. En la práctica, 4 líneas de perspectiva correspondientes a un mismo haz no pasan por el mismo punto de fuga; por ello, es necesario modificar la imagen para obtener un ”verdadero” punto de fuga. Una vez determinadas las aristas largas, es necesario ”empaquetarlas” en haces λ_i de líneas ℓ_{i_j} ; cada haz de al menos 4 líneas determina un punto de intersección que es candidato a punto de fuga (si es intersección de un número menor de líneas, el punto puede ser una esquina).

Una vez identificado el punto de fuga \mathbf{v}_i , se retrazan las candidatas a líneas de perspectiva LP como auténticas líneas de perspectiva, es decir, proyección de líneas paralelas en la escena. El conjunto de líneas de perspectiva contenidas en el plano de imagen a través del punto de fuga \mathbf{V}_i forman una *recta proyectiva* $\lambda_i \simeq \mathbb{P}^1$; el ajuste de haces en una imagen es una transformación proyectiva $\lambda_i \rightarrow \lambda_i$ entre dos rectas proyectivas.

Una vez actualizadas, las líneas de perspectiva ℓ_{i_j} se pueden elevar a líneas paralelas $L_i \subset \mathbb{P}^3$ en el modelo $3D$ de la escena. El haz de líneas de perspectiva $\lambda_i = (\ell_{i_j})_{j \in J}$ se puede representar de dos formas: como la proyección de un haz de líneas $\Lambda_i = (L_{i_j})_{j \in J}$ en $3D$ sobre el plano Π_C de la cámara C o bien como la sección de un haz de planos por un plano transversal (especificar cuál). La segunda descripción permite conectar el enfoque basado en ajuste de haces de líneas en \mathbb{P}^2 con ajuste de haces de planos en \mathbb{P}^3 . Si $\pi_{\mathbf{F}} : \mathbb{P}^3 \rightarrow \mathbb{P}^2$ denota la proyección desde el foco \mathbf{F} de la cámara C , entonces $\pi_{\mathbf{F}}(L_{i_j}) = \ell_{i_j}$ para cada $j \in J$.

4.1.2. Modelo de líneas de perspectiva

Una *línea de perspectiva en una vista 2D* es una línea $\ell \subset \mathbb{P}^2$ que pasa por un punto de fuga $\mathbf{v}_i \in \mathbb{P}^2$. Por ello, se expresa como $\mathbf{v}_i + \mathbf{p}$ donde $\mathbf{p} \in \ell$.

Para una cámara ideal (tipo pinhole) \mathbf{C} con plano de cámara $\Pi_{\mathbf{C}}$ (imagen de la proyección central $\pi_{\mathbf{C}}$ desde el punto ideal \mathbf{C}), el conjunto de líneas de perspectiva que pasan por el punto de fuga \mathbf{v}_i y son proyección sobre el plano de la cámara $\Pi_{\mathbf{C}}$ de líneas $3D$ paralelas entre sí forman un haz uniparamétrico que se denota mediante λ_i .

4.1.3. Líneas del horizonte

Una *línea del horizonte* ℓ_{∞} es una recta que pasa por dos puntos de fuga, es decir, $\ell_{\infty} = \overline{\mathbf{v}_i \mathbf{v}_j}$ para $0 \leq i < j \leq 2$ que denotamos mediante $\mathbf{v}_i + \mathbf{v}_j$. Por tanto, en un plano proyectivo hay tres líneas del horizonte independientes (de hecho hay infinitas, pero independientes sólo 3). Todos los puntos de cualquier línea del horizonte son puntos de fuga.

Ejercicio.- Razonar las afirmaciones siguientes: a) En el plano euclídeo no hay líneas del horizonte; b) en el plano afín sólo hay una línea del horizonte (la que le falta para ser un plano proyectivo); c) en el plano proyectivo cualquier línea puede ser una línea del horizonte.

4.1.4. Planos de perspectiva

Un plano de perspectiva $\Lambda_i + \Lambda_j$ es el generado por dos haces de líneas de perspectiva Λ_i, Λ_j que pasan por dos puntos de fuga $\mathbf{v}_i, \mathbf{v}_j$ y un punto \mathbf{p}_k . Le denotamos mediante $\pi_{i,j,k} = \Lambda_i + \Lambda_j$.

Dos planos de perspectiva con un punto de fuga común \mathbf{v}_i se cortan a lo largo de una línea de perspectiva que pasa por \mathbf{v}_i dada por $\Lambda_i \cap \Lambda_j$, donde se han identificado los haces Λ_i, Λ_j con el plano que determinan.

4.1.5. Planos de referencia

Pueden estar dados por un plano de perspectiva o bien por el plano determinado por tres puntos de fuga $\mathbf{v}_i, \mathbf{v}_j, \mathbf{v}_k$ que denotamos mediante $\mathbf{v}_i + \mathbf{v}_j + \mathbf{v}_k$ o bien mediante $\langle \mathbf{v}_i, \mathbf{v}_j, \mathbf{v}_k \rangle$.

En Reconstrucción 3D afín, siempre hay un plano del infinito que permanece invariante. La determinación de este plano de referencia, simplifica la Reconstrucción 3D y el cálculo de las transformaciones que facilitan su visualización.

Ejercicio.- Comparad esta descripción con la presentada más arriba en términos de los elementos de la cámara y realizad una valoración, mostrando la elevada casuística y mayor inestabilidad de este método basado en modelos de perspectiva en relación con el descrito en la sección anterior.

4.2. Modelos de perspectiva

Los modelos de perspectiva vienen siendo utilizados desde comienzos del s.XV y utilizan aspectos básicos de las Geometría Proyectiva relacionados con las dos operaciones básicas como son proyectar y cortar. Son cruciales para la Reconstrucción 3D a partir de una sola vista y para la visualización de escenas que se desarrollan en el capítulo 4. En esta subsección se presentan los elementos básicos de cualquier modelo de perspectiva y se ilustran los tipos más frecuentes que resultan de utilidad para tareas de visualización que se abordan en la subsección siguiente.

La Geometría Descriptiva se sistematiza a finales del s.XVIII incorporando elementos curvados a representaciones volumétricas. La introducción del sistema diédrico para la representación proporciona elementos adicionales para el abatimiento de elementos que permiten relacionar vistas contenidas a partir de su proyección sobre planos ortogonales. Sin embargo, esta forma de representación presenta serias dificultades para una visualización continua que permita interpolar entre vistas próximas o bien entre las representaciones en perspectiva asociadas a la extracción de elementos geométricos contenidos en las vistas.

El *problema fundamental a resolver* es la estimación y generación automática de elementos característicos como las líneas de perspectiva (LP), los puntos de fuga \mathbf{V}_i y las líneas del horizonte (LH) obtenidas conectando puntos de fuga $\langle \mathbf{V}_i, \mathbf{V}_j \rangle$ para $i \neq j$. Dependiendo de la localización de los puntos de fuga en el plano de imagen se obtienen diferentes tipos de perspectiva a los que se dedican diferentes apartados de esta subsección.



Fig. 9: Imagen generada mediante computador de un modelo de Piranesi

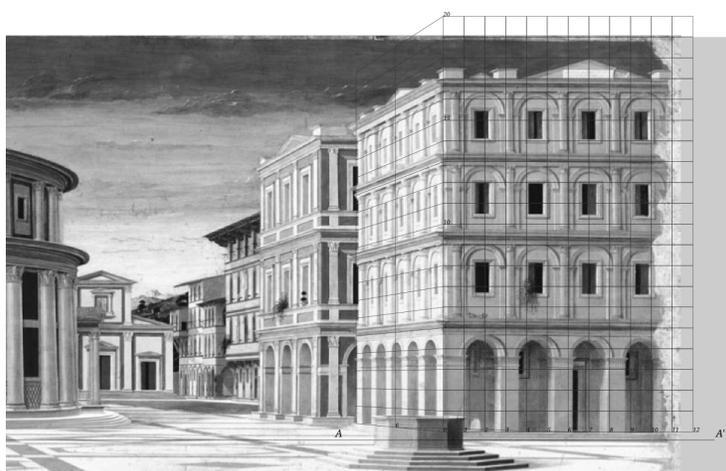


Fig. 10: Estimación geométrica de las líneas de perspectiva en una imagen

4.2.1. Elementos básicos y tipos de perspectiva

La *idea básica a formalizar* en cualquier modelo de perspectiva consiste en que las Líneas paralelas en la escena a l_0 convergen en *punto de fuga*. V_ℓ . Por ello, la localización de los puntos de fuga y el control de sus transformaciones facilita la visualización de los modelos de perspectiva.

Los *Elementos característicos* de un modelo a perspectiva son:

- *Cámara* en el origen $(0, 0, 0)$
- *Plano imagen* paralelo al *plano neutral* Oxz y pasa por $(0, f, 0)$
- *Línea principal*: Dirección principal que parte del "ojo" C
- *Punto principal*: Punto de fuga h de la línea principal (en el plano imagen)

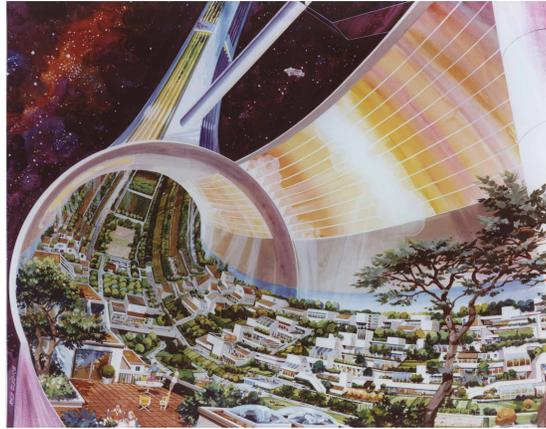


Fig. 11: Ejemplo de imagen con perspectiva paralela

El *problema inicial* consiste en estimar \mathbf{h} . Una primera aproximación consiste en estimar \mathbf{h} como una "suma ponderada" $\mathbf{h} = \sum_{i=1}^3 \lambda_i \mathbf{V}_i$, donde λ_i son las coordenadas baricéntricas de \mathbf{h} con respecto a los puntos de fuga \mathbf{V}_i

Suponemos conocidos i puntos de fuga \mathbf{V}_i con $i \leq 3$ e introducimos restricciones sobre paralelismo:

1. $N = 1$ *Perspectiva Frontal*
2. $N = 2$ *Perspectiva Angular*
3. $N = 3$ *Perspectiva Oblicua*

Excluimos inicialmente la *perspectiva paralela* (tres puntos de fuga en el infinito) por no compatible con los tipos de cámara usuales, ni con la percepción humana.

4.2.2. Perspectiva Frontal

Los elementos característicos de la perspectiva frontal son:

1. Dos ejes principales paralelos al plano imagen
2. Las líneas paralelas en la dirección de profundidad convergen en el punto de fuga $V_2 = P$ (punto principal)
3. El punto principal $\mathbf{h} = \mathbf{V}_2$ (las líneas de fuga son paralelas al rayo principal)
4. Las direcciones correspondientes a $\mathbf{V}_1, \mathbf{V}_3$ forman un sistema ortonormal paralelo al plano imagen.
5. La *escala* es constante en cada plano de profundidad.

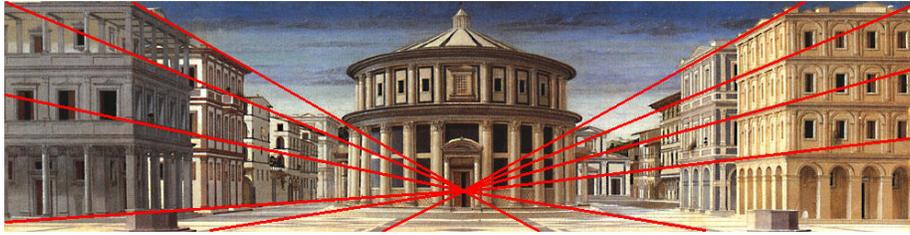


Fig. 12: Ejemplo de imágenes con perspectiva frontal

6. La *longitud focal* f no se puede determinar.

4.2.3. Perspectiva Angular

Los elementos característicos de la perspectiva angular son:

1. *Dos puntos de fuga* \mathbf{V}_1 y \mathbf{V}_2 son *finitos*
2. $\overline{\mathbf{V}_1\mathbf{V}_2}$ es la *línea del horizonte* ℓ_{12} .
3. *Circunferencia principal* C centrada en $\frac{1}{2}[\mathbf{V}_1 + \mathbf{V}_2]$ y diámetro $d(\mathbf{V}_1, \mathbf{V}_2)$.
4. *Estación cameral*: $\mathbf{e} \in C$ sobre perpendicular a ℓ_{12} a través de \mathbf{h}
5. El *punto principal* $\mathbf{h} = \ell_{12} \cap \ell_{12}^\perp$ es el punto de $\overline{\mathbf{V}_1\mathbf{V}_2}$ más próximo al centro de la imagen.
6. La *longitud focal* f se determina a partir de $\mathbf{V}_1, \mathbf{V}_2$ y \mathbf{h} .
7. El *factor de escala* o el *ángulo* proporciona información métrica exacta a partir de los parámetros de la cámara.
8. *Longitud focal* $f = \sqrt{\lambda_1\lambda_2} |\mathbf{V}_2 - \mathbf{V}_1|$
9. \mathbf{V}_3 es *perpendicular* a $\mathbf{V}_2 - \mathbf{V}_1$ y *proporcional* a f .

4.2.4. Perspectiva Oblicua

Los elementos característicos de la perspectiva oblicua son:

1. *Ninguno de los ejes principales* del objeto es paralelo al plano imagen.
2. Las *líneas paralelas* convergen en tres puntos de fuga $\mathbf{V}_1, \mathbf{V}_2$ y \mathbf{V}_3 finitos.
3. El *punto principal* \mathbf{h} es el *ortocentro* del triángulo con vértices \mathbf{V}_i .
4. $(\lambda_1 : \lambda_2 : \lambda_3) = (tg \beta_1, tg \beta_2, tg \beta_3)$ siendo β_i los ángulos internos del triángulo

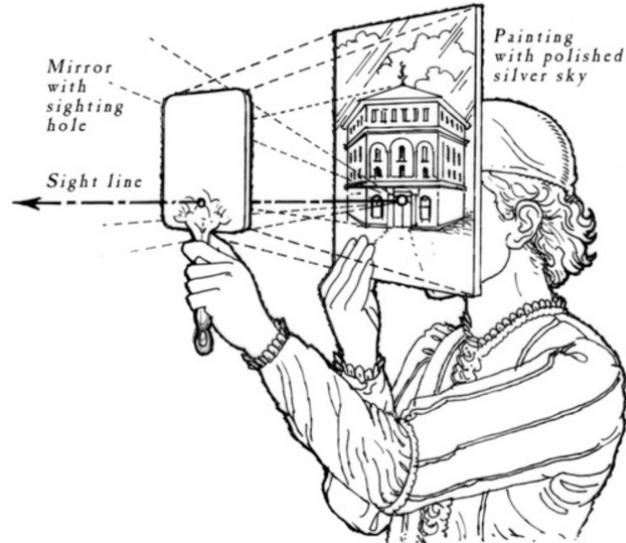


Fig. 13: Ejemplo de imágenes con perspectiva angular de Brunelleschi (1430)

5. La *longitud focal* f se calcula a partir de \mathbf{V}_1 , \mathbf{V}_2 y \mathbf{h} mediante:

$$f = \frac{\sqrt{\lambda_1 \lambda_2}}{\sqrt{\lambda_1 + \lambda_2}} |\mathbf{V}_2 - \mathbf{V}_1|$$

6. El *factor de escala* o el ángulo proporciona información métrica exacta a partir de los parámetros de la cámara.

Este tipo de perspectiva se utiliza con frecuencia para realzar la expresividad de algunas representaciones arquitectónicas o de personajes en aplicaciones de infografía ó videojuegos.

4.3. Visualización de modelos

La mayor utilidad actual de la Visualización basada en modelos de perspectiva está orientada hacia la creación de representaciones con apariencia 3D a partir de imágenes digitales. Los principios utilizados son conocidos desde hace casi seiscientos años. En el Renacimiento se crea la ilusión de una volumetría mediante mallas asociadas a una discretización de las líneas de perspectiva. La aplicación de transformaciones curvilíneas a este tipo de representaciones ha dado lugar a representaciones de “perspectiva curvada” que son aplicables no sólo a objetos arquitectónicos con complicadas geometrías (cubiertas del Museo Guggenheim, p.e.), sino al modelado de animales o personajes. Para ello es necesario desarrollar:

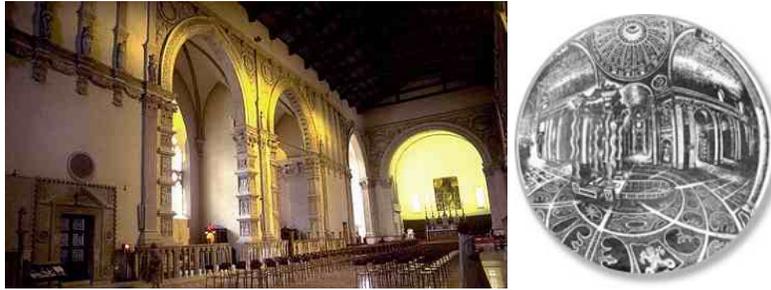


Fig. 14: a) Templo de Malatesta (Alberti), b) perspectiva oblicua del cubo (Stevin)

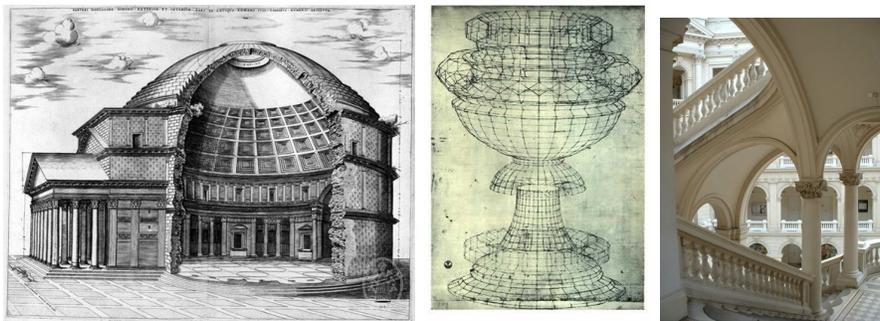


Fig. 15: a) Modelado curvado mediante levantamientos del Panteón (Agripa), b) Cáliz de Paolo Uccello y c) Encabalgamiento de representaciones en la escalera de la Universidad de Varsovia

1. Una representación basada en *perspectiva curvilínea* para objetos curvados que está dada por transformaciones no-lineales definidas sobre líneas de perspectiva
2. Construir *Mallas curvilíneas* mediante algún tipo de transformación de mallas planares cuadrangulares que sean adaptables sobre el objeto. Ello requiere fijar el tipo de curvas (incluyendo el grado para curvas racionales algebraicas, p.e.) e identificar puntos de control sobre el objeto curvado que puedan facilitar la generación automática de las transformaciones.
3. Estimar y corregir los efectos de *encabalgamiento de objetos* sobre modelos de perspectiva curvilínea de manera invariante sobre el marco proyectivo. Ello requiere adaptar el modelo de secciones y proyecciones presentado para el caso lineal.

4.3.1. Visualización de perspectivas

La *Visualización de perspectivas* requiere mostrar las transformaciones que permiten transformar una representación en otra. Un elemento importante es la organización del modelo en torno a la línea de visión como “eje principal” o punto privilegiado del haz de rectas del espacio ordinario desde el que se proyecta. El modelo para dicho haz es un plano proyectivo \mathbb{P}^2 . Por tanto, la representación más sencilla de la transformación requerida está dada por una homografía plana, es decir, por una transformación de \mathbb{P}^2 en \mathbb{P}^2 ; dicha homografía está determinada por un elemento de $GL(3; \mathbb{R})$ salvo escala.

A menudo se cuenta con modelos toscos (procedentes de diseño manual o bien de análisis de imagen) para la fase de pre-visualización que puedan ser refinados en etapas posteriores. La *relación entre vistas planares* de un objeto volumétrico se puede generar mediante *homografías*. Sin embargo, las *homografías* dan lugar a *distorsiones* en zonas próximas a las líneas del infinito ℓ_∞ . Para corregir este efecto hay dos opciones

- *Opción 1: recortar las imágenes* en las “zonas próximas” a ℓ_∞
- *Opción 2: navegar modelo hiperbólico* del plano proyectivo \mathbb{P}^2

4.3.2. Visualización asociado al movimiento virtual de cámara

Una vez elaborado un modelo tosco del objeto o de la escena, es posible simular el movimiento relativo de la escena con respecto al observador virtual que se genera de forma interactiva. A (la simulación de) este proceso se le llama *movimiento a partir de la estructura/forma* (MFS/mfs) y se aborda con más detalle en el módulo 3 relacionado con el análisis de movimiento; es claro que, al menos para los mamíferos superiores, existe una realimentación entre SFM (resp. sfm) y MFS (resp. mfs) que debe ser modelada en términos de $\dots \rightarrow M \rightarrow S \rightarrow M \rightarrow \dots$ y que se engloba dentro del ciclo percepción-acción (PAC) que se comenta en varios apartados de este Curso.

De una manera intuitiva, el movimiento de ratón realizado por el usuario debe traducirse en un movimiento virtual de cámara (rígido o asociado a algún tipo de perspectiva), dando lugar a nuevas vistas del objeto o de la escena. La visualización de este efecto se puede realizar en términos de imagen o bien en términos de la escena.

- *Visualización en términos de imagen*: Existen restricciones topológicas (orden, adyacencia, localización relativa) y proyectivas (epipolar básicamente) que afectan a los elementos homólogos que, actualmente, es posible estimar en tiempo real.
- *Visualización en términos de escena*: Requieren una rápida actualización de mapas de vóxeles coloreados, operación que es necesario realizar a diferentes niveles de resolución utilizando variantes de conos proyectantes

asociados a siluetas y actualizando la información en función de movimiento propio o del de los agentes móviles en la escena;; estas cuestiones se abordan en los módulos 3 y 5.

A partir del capítulo 2 de este módulo se desarrollan sobre todo los métodos de Reconstrucción 3D que afectan a la visualización en términos de imagen.

4.3.3. Visualizando la Geometría Descriptiva

Los modelos de perspectiva lineales proporcionan un soporte para visualizar modelos desde diferentes puntos de vista. La Reconstrucción 3D permite visualizar dichos modelos usando transformaciones proyectivas en el espacio ambiente y corrigiendo las distorsiones asociadas a las transformaciones afines que se manifiestan como falta de ortogonalidad en los elementos de referencia.

Esta estrategia se desarrolla en el capítulo 4 del módulo 2 (Reconstrucción). Para obtener una visualización más completa, es necesario incorporar herramientas de la Geometría Descriptiva incluyendo primitivas geométricas con parámetros ajustables.

4.3.4. Visualizaciones de modelos hiperbólicos

La modificación de la apariencia continua por navegación en un espacio hiperbólico se visualiza más fácilmente cuando se dispone de una teselación sobre algún modelo del plano hiperbólico. El modelo más fácil de visualizar está dado por el *disco de Poincaré*, un disco plano en el que la semicircunferencia del borde se obtiene identificando los puntos diametralmente opuestos por la aplicación antipodal. El disco de Poincaré admite teselaciones regulares por polígonos regulares que facilitan la visualización tal y como muestra la figura 18.

La deformación aparente de las células básicas ilustra el cambio aparente en la forma a medida que nos aproximamos al borde; esta percepción es “engañosa” pues el espacio que representamos es homogéneo.

Los métodos de Reconstrucción 3D presentados en los capítulos siguientes muestran procedimientos alternativos centrados en el objeto que son independientes de la localización del observador.

Ejercicio.- Mostrad ejemplos correspondientes al primer procedimiento.

4.4. Navegación interactiva

4.4.1. Navegación 2D

La navegación 2D realiza la simulación de un vuelo sobre una superficie con un desplazamiento virtual de la cámara, modificando la altura al tiempo que realiza traslaciones y rotaciones con una velocidad variable. Un esquema simplificado consiste en los pasos siguientes:

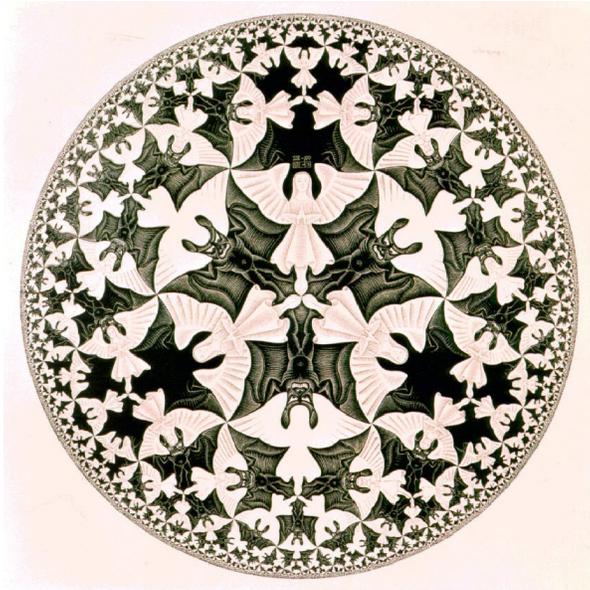


Fig. 16: Grabado de Escher con la teselación dada por Ángeles y Demonios

1. Discretizar la superficie S en términos de una colección \mathcal{S} de puntos \mathbf{p}_i ó sitios pertenecientes a una superficie S .
2. Construir una red triangular irregular (TIN) \mathcal{T} asociada a \mathcal{S} ³¹
3. Introducir una función de altura variable h sobre el vector normal unitario \mathbf{u}_{ijk} en el baricentro de cada triángulo $T_{ijk} \in \mathcal{T}$.
4. Para un desplazamiento del ratón, generar una secuencia de movimiento sobre el mapa de alturas interpolando entre los extremos de las alturas correspondientes.
5. Visualizar la TIN mediante interpolación afín entre vistas consecutivas.

Nótese que en este caso no hay obstáculos a la navegación.

4.4.2. Modelos centrados en el observador

Este tipo de modelos utilizan una perspectiva central ó cónica asociada a modelos de perspectiva actualizables en función de la localización de la cámara virtual. El esquema es similar al del apartado anterior, pero cambiando las vistas de la superficie planar por una representación en perspectiva de los objetos presentes en la escena.

³¹ En un gran número de aplicaciones software se dispone de la triangulación de Delaunay $\mathcal{D}(\mathcal{S})$ asociada a \mathcal{S} que tiene propiedades óptimas

Para simplificar se supone que la escena está dada por un mapa de prismas con diferentes alturas con texturas superpuestas. El objetivo es facilitar una navegación en el mapa de prismas, sin colisionar con ninguno de ellos. Denotemos mediante B^α a los objetos que son obstáculos para la navegación y sea $\mathcal{B} = \cup_{\alpha \in A} B^\alpha$ la unión de todos los obstáculos. Los pasos adicionales a dar son los siguientes:

1. Realizar una triangulación del espacio libre \mathcal{F} y del espacio ocupado \mathcal{B} .
2. Introducir restricciones relativas a altura mínima para el mapa de alturas asociado al espacio ocupado \mathcal{B} .
3. Para un desplazamiento arbitrario del ratón identificar posibles eventos relacionados con colisiones, saltos ó navegación aérea en el entorno asociado al espacio libre \mathcal{F} y al espacio ocupado \mathcal{B} .
4. Interpolan sobre el mapa de alturas asociado a las normales unitarias del mapa de alturas sobre los baricentros de los triángulos.
5. Visualizar el resultado final.

4.4.3. Navegación en tercera persona

Este tipo de navegación requiere insertar un personaje al que sigue la cámara. El personaje se almacena y se gestiona como una capa independiente con respecto a la escena. Para empezar, es necesario restringir el tipo de movimientos que están permitidos para el personaje, simplificar la representación y resolver las modificaciones en las apariencias vinculadas al movimiento del personaje en la escena.

Las modificaciones en las apariencias se abordan en términos de transformaciones casi-conformes (se permite variación acotada para los ángulos que forman parte de los elementos de referencia) ó bien utilizando una implementación en tiempo real del método de las características de Cauchy (para representaciones simplificadas tipo armadura). La información geométrica se actualiza en términos de transformaciones o de resolución de sistemas dinámicos dependientes de parámetros.

4.4.4. Valoración y resumen

Las transformaciones conformes son las más generales que conservan ángulos y generan un grupo al que se llama *grupo conforme* que contiene al subgrupo ortogonal conforme. Los subgrupos de semejanza y especial ortogonal forman subgrupos encajados (cada vez más pequeños) dentro de una “jerarquía natural” entre las geometrías. Esta jerarquía es de utilidad para facilitar la generación semi-automática de transformaciones que permitan comparar o realizar la síntesis de datos entre geometrías adyacentes (correspondientes a subgrupos encajados).

En particular, las *transformaciones euclídeas* conservan la distancia $d(\mathbf{P}, \mathbf{Q})$, el ángulo entre dos rectas, la orientación relativa de unos elementos con respecto a otros, áreas y volúmenes. La conservación de estas últimas “cantidades” no implica que se conserve la distancia euclídea, pero proporciona criterios más flexibles para comparar datos en relación con diferentes modelos de perspectiva en relación con problemas de reconocimiento o bien de síntesis de nuevas vistas; un caso típico

En la práctica es importante controlar las deformaciones desde el punto de vista discreto. Una estrategia típica consiste en llevar a cabo dicho control sobre elementos de mallas triangulares o cuadrangulares. En particular, ello afecta a restricciones que afectan a los ángulos de las células básicas (triángulos o cuadriláteros, habitualmente) eventualmente curvilíneos.

Recordemos que la información sobre ángulos de cada triángulo es proyectivamente equivalente (es decir, salvo factor de escala) a la información sobre distancias gracias al teorema de los senos:

$$\frac{\text{sen } \alpha}{a} = \frac{\text{sen } \beta}{b} = \frac{\text{sen } \gamma}{c}$$

En particular, la conservación de distancias implica la conservación de ángulos, pero el recíproco no es cierto. La proyección estereográfica muestra una equivalencia conforme entre la esfera menos un punto y el plano; en particular, aplica la malla cuadrangular curvilínea formada por meridianos y paralelos en una malla cuadrangular plana. Estamos interesados en distribuciones eventualmente irregulares de puntos y mallas triangulares o cuadrangulares asociadas.

La disponibilidad de 3-uplas de puntos próximos no-alineados, permite recuperar propiedades métricas regionales de la imagen; el carácter métrico se evalúa en términos de circunferencias asociadas a ternas de puntos. El espacio de Hough proporciona un soporte para parametrizar circunferencias. El mapa de circunferencias se proyecta sobre un mapa de elipses para una localización diferente de la cámara. La estimación de la deformación de una circunferencia en una elipse es clave para procesos de auto-calibración que se abordan en el capítulo siguiente.

Nótese que la disponibilidad de 4-uplas de puntos próximos no-coplanarios, permite recuperar propiedades volumétricas asociadas a descomposiciones tetraedrales de uso común en Informática Gráfica o en los Métodos de Elementos Finitos (FEM). El carácter no-ordenado de las nubes de puntos subyacentes a descomposiciones tetraedrales hace muy difícil la resolución de la puesta en correspondencia, incluyendo un número muy elevado de posibles transiciones con respecto a las células contiguas. Por ello, resulta bastante más práctico utilizar descomposiciones basadas en mapas cuadrangulares planares y su elevación a paralelepípedos, incluyendo eventualmente la posibilidad de modelos de perspectiva curvados para adaptarse a objetos con una volumetría complicada tanto para el caso estático (esculturas, p.e.) como dinámico (actores en movimiento, p.e.)

La utilización de esferas asociadas a 4-tuplas de puntos en posición general facilita una representación para la segmentación volumétrica de la escena 3D y

su gestión en términos del grupo conforme de la esfera. Este marco facilita el desarrollo de herramientas para la deformación de objetos.

Como *inconveniente* del modelado basado en este tipo de transformaciones, cabe destacar que el modelado euclídeo es *muy* sensible al ruido. Por ello, salvo que haya requerimientos específicos relativos a la precisión, es preferible trabajar bajo condiciones más flexibles identificando la variante del grupo conforme que resulta más significativa para el problema propuesto ... o bien extendiendo el marco geométrico al afín o al proyectivo (este enfoque se desarrolla sobre todo en el capítulo 3).

5. Conclusiones y retos

Los elementos fundamentales para cualquier representación en perspectiva son las líneas de perspectiva y los puntos de fuga. Las líneas de perspectiva se agrupan en haces que pasan a través de cada punto de fuga. Los modelos de perspectiva resultan muy útiles para visualizar escenarios y permiten la generación de una apariencia volumétrica de la escena en términos de un mapa de profundidad. Frecuentemente, este tipo de representación recibe el nombre de reconstrucción $2,5D$ haciendo referencia a que sólo ofrece la apariencia volumétrica para la parte visible del objeto

Existen diferentes modelos de perspectiva dependiendo del número de puntos de fuga independientes que se encuentran a distancia finita.

5.1. Retos

En los apartados siguientes se comentan algunos retos que se consideran significativos para el desarrollo de posibles prácticas adicionales.

5.1.1. Estimación robusta de modelos de perspectiva para el caso móvil

La no-linealidad de los modelos de perspectiva y la necesidad de aplicaciones que puedan proporcionar resultados robustos en tiempo real sugiere el desarrollo de aproximaciones lineales para la estimación de los modelos de perspectiva. Las dos estrategias más conocidas son los modelos de perspectiva débil y de para-perspectiva. Una estimación de estos modelos se desarrolla en el capítulo 4.

Las transformaciones entre diferentes modelos de perspectiva se realizan en términos del modelo proyectivo subyacente. Estas transformaciones habitualmente no tienen en cuenta los modelos de estimación utilizados para estimar los modelos de perspectiva.

Las aproximaciones lineales a los modelos de perspectiva presentan errores que es necesario acotar. Es importante acotar la variación del error por cambios coordenados. El enfoque matemático usual parte de una cantidad mínima de dominios coordenados afines que recubren el espacio proyectivo. Sin embargo, los cambios coordenados asociados (definidos mediante $x_i x_j^{-1}$ dan lugar a distorsiones importantes (muy significativos cuando $x_j \rightarrow 9$). Por ello, en la práctica interesa trabajar con un mayor número de cartas con cambios "más graduales" que faciliten una mejor gestión del error. Es necesario desarrollar una estrategia para un número finito de refinamientos sucesivos utilizando la discretización de coordenadas locales afines.

5.1.2. Corrección de la distorsión asociada a modelos de perspectiva

Los modelos de perspectiva generados a partir de la información capturada por una cámara sólo son válidos para un "pequeño entorno" de la línea de visión. La utilización de lentes de gran angular proporciona una representación distorsionada de la escena pero con un control mucho más efectivo de la misma,

pues "minimiza" el número de pegados a realizar. Las distorsiones asociadas a la utilización de un gran angular se pueden controlar en términos de una proyección estereográfica que conservan los ángulos. Por ello, la realimentación entre modelos de proyección esférica y modelos de perspectiva ofrece posibilidades más amplias que las asociadas a modelos de perspectiva más frecuentes.

5.1.3. Superposición de información radiométrica

La información radiométrica se incorpora habitualmente mediante la combinación de dos tipos de técnicas que afectan a rectas del espacio proyectivo. De una manera muy simplificada, estamos interesados en dos tipos de rectas; las que pasan por el foco de una cámara \mathbf{C} o una fuente de luz \mathbf{S}_L o bien las que se encuentran sobre un plano tangente a la superficie (para representar el comportamiento óptico de la materia en relación con la luz).

5.1.4. Visualización de objetos complejos o imaginados

1. *Geometrías imaginadas*: modelos no-euclídeos sugeridos en grabados antiguos (ver figura 9)
2. *Modelos complejos de perspectiva* con un número elevado ($N \geq 4$) de puntos de fuga. La gestión de los haces de líneas de perspectiva correspondientes presenta una mayor complejidad para la que es necesario desarrollar herramientas relacionadas con *arrangements* degenerados de rectas. Actualmente, no existen herramientas computacionales para gestionar este tipo de problemas.
3. *Modelos curvilíneos* para el mapeado de modelos de perspectiva sobre mapas de terreno o bien sobre objetos volumétricos complicados.
 - *Representaciones semi-esféricas* con aplicaciones a proyecciones esféricas sobre entornos FullDome.
 - *Representaciones cilíndricas* con aplicaciones a la generación semi-automática de vistas panorámicas.
 - *Representaciones hiperbólicas* con aplicaciones a la visualización de objetos volumétricos (caras humanas, p.e.) o de objetos con la topología del disco (representaciones simplificadas para cuerpos de mamíferos, p.e.) o bien escenas con geometrías "imposibles" como las ilustradas en la figura 18

El modelo global del plano proyectivo \mathbb{P}^2 está dado por el cociente $\mathbb{S}^2/\mathbb{Z}_2$ de la esfera \mathbb{S}^2 por la aplicación antipodal que identifica puntos diametralmente opuestos. La proyección de la semiesfera sobre el plano da lugar a un disco \mathbb{P}^2 cuyo borde es una semicircunferencia $\mathbb{P}^1 = \mathbb{S}^1/\mathbb{Z}^2$ que representa una línea del infinito; este modelo recibe el nombre de *disco de Poincaré*.

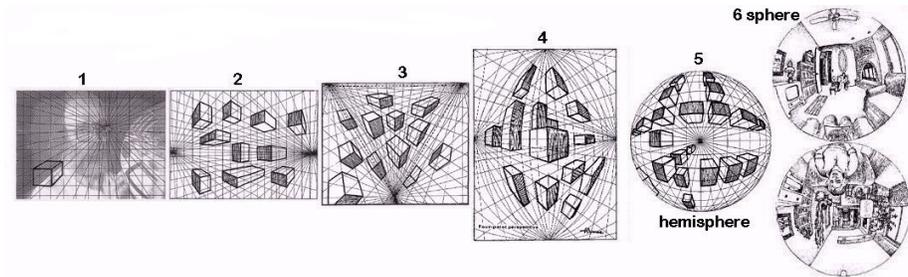


Fig. 17: Ejemplo de imágenes de perspectiva con hasta 6 puntos de fuga

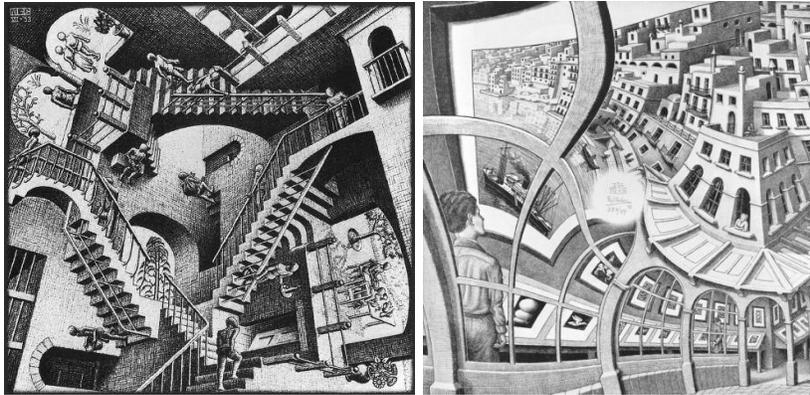


Fig. 18: Perspectivas relativas de Escher con implicaciones paradójicas y orientaciones contradictorias

La representación basada en el disco de Poincaré es preferible a la representación basada en el semiplano superior pues las transformaciones definidas sobre el primero son “conformes” de una manera natural, es decir, *conservan los ángulos*. Esta propiedad es importante para deformaciones aparentes o reales de objetos, pues la Geometría se reconoce a pesar de dichas deformaciones. Tienen una gran cantidad de “buenas propiedades” y se utilizan de forma cada vez mayor para la generación de personajes virtuales o, con más generalidad, en la producción de contenidos multimedia.

El plano proyectivo es una variedad homogénea que tiene curvatura total constante negativa, es decir, es una variedad hiperbólica. Por ello, el disco de Poincaré proporciona una *representación hiperbólica* del plano. Este modelo tiene múltiples aplicaciones incluyendo la visualización de objetos volumétricos (caras humanas, p.e.) o de objetos con la topología del disco (representaciones simplificadas para cuerpos de mamíferos, p.e.) o bien escenas con geometrías “imposibles” como las ilustradas en la figura

5.2. Problemas