

Visión tridimensional

Javier Sánchez Pérez
Facultad de Informática
Universidad de Las Palmas de Gran Canaria
email: jsanchez@dis.ulpgc.es
<http://serdis.dis.ulpgc.es/~jsanchez>

Contents

1	Introducción	3
2	Geometría proyectiva y transformaciones 2D	3
2.1	El plano proyectivo 2D	3
2.1.1	Puntos y líneas	3
2.1.2	Puntos ideales y la línea en el infinito	4
2.1.3	Cónicas y cónicas duales	4
2.2	Transformaciones proyectivas	5
2.2.1	Transformaciones de líneas y cónicas	5
2.3	Jerarquía de transformaciones	5
2.3.1	Clase I: Isometrías	5
2.3.2	Clase II: Transformaciones de similaridad	6
2.3.3	Clase III: Transformaciones afines	6
2.3.4	Clase IV: Transformaciones proyectivas	6
2.4	Geometría proyectiva 1D (\mathbb{P}^1)	7
2.5	Recuperación de propiedades afines y métricas a partir de las imágenes	7
2.5.1	Línea en el infinito	7
2.5.2	Recuperación de propiedades afines	8
2.5.3	Puntos circulares y su dual	8
2.5.4	Ángulos en el plano proyectivo	8
2.5.5	Recuperación de propiedades métricas	9
2.6	Otras propiedades de las cónicas	9
2.6.1	La relación pole-polar	9
3	Geometría proyectiva y transformaciones 3D	10
3.1	Puntos y transformaciones proyectivas	10
3.2	Planos, líneas y cuádricas	10
3.3	Plano en el infinito	11
3.4	La cónica absoluta	11
3.5	La cuádrica dual absoluta	12

4	Estimación de transformaciones proyectivas	12
4.1	Algoritmo DLT	12
4.2	Estimación robusta	13
4.2.1	RANSAC	13
4.3	Cálculo automático de una homografía	14
5	Geometría de una vista simple - Modelo de la cámara	15
5.1	Cámara finita	15
5.2	Cámara proyectiva	17
5.3	Cámaras en el infinito	17
5.3.1	Cámara afines	18
6	Cálculo de la matriz \mathbf{P}	18
6.1	Ecuaciones básicas	18
7	Geometría epipolar y matriz fundamental	19
7.1	Geometría epipolar	19
7.2	La matriz fundamental	19
7.2.1	Deducción geométrica	19
7.2.2	Deducción algebraica	20
7.2.3	Condición de correspondencia	20
7.2.4	Propiedades de la matriz fundamental	20
7.2.5	Homografía de la línea epipolar	20
7.3	Matrices fundamentales especiales	21
7.3.1	Traslación pura	21
7.4	Cálculo de las matrices de las cámaras	21
7.5	Matriz esencial	21
8	Reconstrucción 3D de cámaras	22
8.1	Método de reconstrucción	22
8.2	Ambigüedad en la reconstrucción	22
8.3	Teorema de la reconstrucción proyectiva	22
8.4	Reconstrucción estratificada	23
9	Cálculo de la matriz \mathbf{F}	23
9.1	Ecuaciones básicas	23
9.2	El algoritmo de 8 puntos normalizados	24
9.3	Distancia geométrica	24
9.4	Cálculo automático de \mathbf{F}	25
10	Cálculo de la estructura	25
10.1	Métodos lineales de triangulación	25
10.2	Función de coste de error geométrico	26
10.3	Solución óptima	26
11	Geometría epipolar afín	28
11.1	La matriz \mathbf{F} afín	28
11.2	Cálculo de \mathbf{F}_A a partir de puntos en correspondencia	28

1 Introducción

El objetivo de este curso es el de profundizar en temas relacionados con la visión tridimensional y, en concreto, con la reconstrucción 3D. De forma básica, el campo de la visión tridimensional trata de obtener, a partir de imágenes 2D, información relacionada con la escena original de donde se tomaron dichas imágenes.

Las imágenes 2D se suelen capturar a través de cámaras digitales o videocámaras. Estas cámaras suelen proyectar los objetos 3D en un plano a través de una transformación proyectiva. Esta transformación se puede representar por medio de una matriz que facilita el tratamiento de muchos problemas gracias al álgebra lineal.

Una parte fundamental del curso será, por tanto, el estudio de la *geometría proyectiva*. Veremos muchos conceptos que tienen un representación geométrica real y, lo más importante, propondremos modelos matemáticos que nos permitirán representarlos y trabajar con ellos.

Otra parte importante es el *modelado de las cámaras*. Para poder conocer la geometría de una escena real, es necesario que conozcamos cómo las cámaras transforman los objetos y los proyectan en la imagen. Las cámaras se pueden modelar a través de un conjunto de parámetros, intrínsecos y extrínsecos, que tienen un significado físico concreto. Los parámetros intrínsecos son los que tienen que ver con la propia lente de la cámara, mientras que los extrínsecos tienen que ver con la posición y orientación de la cámara en la escena.

Por otra parte, estudiaremos el problema de la *calibración*. Estos son métodos que nos ayudan a estimar la configuración del sistema de cámaras y del propio modelo de las cámaras a partir de las imágenes que se toman de la escena. Este problema es importante de cara a estimar de forma precisa los datos necesarios para poder realizar posteriormente la reconstrucción de la escena o para relacionar geoméricamente las cámaras.

Posteriormente extrapolaremos el caso de la visión a partir de una cámara al caso en que tenemos dos cámaras. Esto es lo que se denomina la *visión estereoscópica*. La visión estereoscópica es la que posee el ser humano y un gran número de animales. El interés de este tipo de visión es que nos permite percibir la profundidad de los objetos y realizar un amplio conjunto de tareas que sin ella serían bastante complejas. Nos interesaremos por los modelos matemáticos que nos permitirán representar y trabajar con dos cámaras que visualizan una escena común y estudiaremos sus características más interesantes.

Finalmente trataremos el problema de la *reconstrucción tridimensional*. A partir de dos cámaras y de ciertas correspondencias entre sus imágenes, es posible obtener los puntos 3D originales. Estudiaremos formas distintas de reconstruir la escena a partir de un conjunto de correspondencias.

Existen muchos problemas en visión 3D que no requieren realizar la reconstrucción final. Muchas veces sólo nos basta con conocer la geometría de las cámaras y cómo están relacionadas para realizar ciertas medidas que de igual forma se pueden realizar entre los puntos de la escena. Por esto, la reconstrucción es, muchas veces, un paso opcional.

2 Geometría proyectiva y transformaciones 2D

2.1 El plano proyectivo 2D

2.1.1 Puntos y líneas

Un punto se puede representar en un plano por medio de un par (x, y) en \mathbb{R}^2 .

Una línea en el plano se representa por medio de la ecuación $ax + by + c = 0$. $\mathbf{l} = (a, b, c)^T$. Los vectores $(a, b, c)^T$ y $k(a, b, c)^T$ representan la misma línea. Vector *homogéneo* es una clase

de equivalencia de vectores que cumplen esta relación. El conjunto de clases de equivalencia de vectores en $\mathbb{R}^3 - (0, 0, 0)^T$ forma el *espacio proyectivo* \mathbb{P}^2

Un punto en coordenadas homogéneas se representa como $\mathbf{x} = (x, y, 1)^T$. Los puntos incluidos en una recta son aquellos que cumplen $\mathbf{x} \cdot \mathbf{l} = 0$. Se cumple además que $\mathbf{x} \cdot \mathbf{l} = k(\mathbf{x} \cdot \mathbf{l}) = 0$, por lo tanto los puntos en el espacio proyectivo \mathbb{P}^2 se representan como $\mathbf{x} = (kx, ky, k) = (x_1, x_2, x_3)$, y su correspondiente en \mathbb{R}^2 como $\mathbf{x} = (x_1/x_3, x_2/x_3)$.

Theorem 1 *La intersección de dos líneas \mathbf{l} y \mathbf{l}' es el punto $\mathbf{x} = \mathbf{l} \times \mathbf{l}'$.*

$$\mathbf{l} \cdot (\mathbf{l} \times \mathbf{l}') = \mathbf{l}' \cdot (\mathbf{l} \times \mathbf{l}') = 0 = \mathbf{l}'^T \cdot \mathbf{x} = \mathbf{l}^T \cdot \mathbf{x}$$

Example 2 *Encontrar el punto de intersección entre las rectas $x = 1$ e $y = 1$.*

La representación homogénea de la primera línea es $\mathbf{l} = (-1, 0, 1)^T$ y la de la segunda es $\mathbf{l}' = (0, -1, 0)$, por lo que la intersección es $\mathbf{x} = \mathbf{l} \times \mathbf{l}' = (1, 1, 1)^T$.

Remark 3 *La línea que une dos puntos \mathbf{x} y \mathbf{x}' es $\mathbf{l} = \mathbf{x} \times \mathbf{x}'$.*

2.1.2 Puntos ideales y la línea en el infinito

La intersección de dos líneas paralelas $\mathbf{l} = (a, b, c)^T$ y $\mathbf{l}' = (a, b, c')^T$ da como resultado un punto $\mathbf{x} = \mathbf{l} \times \mathbf{l}' = (b, -a, 0)$ que se denomina punto ideal o punto en el infinito. Si obtenemos el correspondiente punto en coordenadas inhomogéneas, obtenemos $(b/0, -a/0)^T$. Gracias a las coordenadas homogéneas podemos representar puntos de este estilo.

Luego los puntos en el infinito son aquellos que se pueden representar como $(x_1, x_2, 0)$ en \mathbb{P}^2 . Todos los puntos en el infinito están incluidos en una misma línea, *línea en el infinito*, que se representa como $\mathbf{l}_\infty = (0, 0, 1)^T$.

Cualquier línea y sus paralelas, $\mathbf{l} = k(a, b, c)^T$, intersectan la línea en el infinito en $\mathbf{x} = (b, -a, 0)$. En coordenadas inhomogéneas, el vector $(b, -a)$ es ortogonal a la normal (a, b) , con lo que representa la dirección de la línea. La línea en el infinito se puede ver como el conjunto de las direcciones de las líneas del plano.

Remark 4 *Principio de dualidad: A cualquier teorema sobre geometría proyectiva en 2D le corresponde un teorema dual que se deriva de intercambiar los roles entre puntos y líneas.*

2.1.3 Cónicas y cónicas duales

La ecuación de una cónica en coordenadas inhomogéneas es:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0.$$

Si homogeneizamos esta ecuación atendiendo a las relaciones $x \rightarrow x_1/x_3$ e $y \rightarrow x_2/x_3$ entonces tenemos:

$$ax_1^2 + bx_1x_2 + cx_2^2 + dx_1x_3 + ex_2x_3 + fx_3^2 = 0$$

que si lo expresamos a través de notación matricial tenemos $\mathbf{x}^T \mathbf{C} \mathbf{x} = 0$, con

$$\mathbf{C} = \begin{pmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{pmatrix}.$$

Esta matriz tiene cinco grados de libertad, con lo que necesitaríamos cinco puntos distintos para poder especificarla.

Theorem 5 La línea l tangente a la cónica C en el punto \mathbf{x} viene dada por $l = C\mathbf{x}$

La línea $l = C\mathbf{x}$ pasa por \mathbf{x} ya que $\mathbf{x}^T l = \mathbf{x}^T C\mathbf{x} = 0$. Si la línea tiene otro punto de intersección, \mathbf{y} , en la cónica entonces se tiene que cumplir que $\mathbf{x}^T C\mathbf{y} = \mathbf{y}^T l = 0$. De esto se deduce que $(\mathbf{x} + \lambda\mathbf{y})^T C(\mathbf{x} + \lambda\mathbf{y}) = 0$ se tiene que cumplir para cualquier λ , lo que significa que la línea y la cónica coinciden (cónica degenerada).

dtbpFUX2.0418in1.3612in0ptPuntos que satisfacen $\mathbf{x}^T C\mathbf{x} = 0$.Plot

Cónica dual: La cónica dual nos da el conjunto de líneas que son tangentes a la cónica y viene dado por $l^T C^* l = 0$, con C^* la matriz adjunta de C . Si C es invertible entonces $C^* = C^{-1}$.

ftbpFU146.5pt83.8125pt0ptLíneas que satisfacen $l^T C^* l = 0$.Figure

2.2 Transformaciones proyectivas

Una *proyectividad* es un mapeo invertible h de \mathbb{P}^2 sobre sí mismo de tal forma que tres puntos $\mathbf{x}_1, \mathbf{x}_2$ y \mathbf{x}_3 reposan sobre la misma línea sí y solo sí $h(\mathbf{x}_1), h(\mathbf{x}_2)$ y $h(\mathbf{x}_3)$ también reposan sobre una misma línea.

Una proyectividad se suele llamar también *colineación*, *transformación proyectiva* u *homografía*.

Una proyectividad se puede representar a través de una matriz \mathbf{H} de 3×3 no singular de forma que la relación entre puntos originales y transformados, es una aplicación lineal, $\mathbf{x}' = \mathbf{H}\mathbf{x}$.

2.2.1 Transformaciones de líneas y cónicas

Las líneas, bajo una transformación proyectiva, se transforman de la siguiente manera. Tenemos dos líneas $l^T \mathbf{x} = 0$ y $l'^T \mathbf{x}' = 0$, y sabemos que los puntos se corresponden a través de $\mathbf{x}' = \mathbf{H}\mathbf{x}$, luego las líneas se relacionan como $l'^T \mathbf{x}' = l^T \mathbf{H}^{-1} \mathbf{H}\mathbf{x}$, con lo que la línea transformada es

$$l' = \mathbf{H}^{-T} l.$$

Las cónicas se transforman como $\mathbf{x}^T C\mathbf{x} = \mathbf{x}'^T \mathbf{H}^{-T} C\mathbf{H}^{-1} \mathbf{x}'$, con lo que

$$C' = \mathbf{H}^{-T} C\mathbf{H}^{-1}.$$

2.3 Jerarquía de transformaciones

2.3.1 Clase I: Isometrías

Son transformaciones que preservan la distancia Euclídea.

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} \epsilon \cos \theta & -\sin \theta & t_x \\ \epsilon \sin \theta & \cos \theta & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

donde $\epsilon = \pm 1$. Este tipo de transformaciones sirve para modelar movimientos rígidos. Se puede representar como una rotación, un desplazamiento y reflexión:

$$\mathbf{x}' = \mathbf{H}_E \mathbf{x} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \mathbf{x}.$$

Este tipo de transformaciones tiene tres grados de libertad, con lo que se puede calcular a partir de dos puntos en correspondencia.

Los invariantes que se mantienen con estas transformaciones son: la distancia entre dos puntos, el ángulo entre dos líneas, y el área.

2.3.2 Clase II: Transformaciones de similaridad

Es igual que una isometría compuesta con un escalado isotrópico.

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} s \cos \theta & -s \sin \theta & t_x \\ s \sin \theta & s \cos \theta & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$\mathbf{x}' = \mathbf{H}_S \mathbf{x} = \begin{pmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \mathbf{x}.$$

Esta transformación tiene como característica que preserva la forma de los objetos, teniendo cuatro grados de libertad, con lo que se pueden calcular a partir de dos puntos en correspondencia.

Los invariantes para este tipo de transformaciones son: ángulo entre dos líneas, el *ratio* entre dos distancias y el *ratio* entre dos áreas.

Cuando hablamos de *estructura métrica* nos referimos a que la estructura está definida hasta una similaridad.

2.3.3 Clase III: Transformaciones afines

Una transformación afín se representa como una matriz:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$\mathbf{x}' = \mathbf{H}_A \mathbf{x} = \begin{pmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \mathbf{x}.$$

Una transformación afín tiene seis grados de libertad, con lo que se puede especificar a través de tres puntos en correspondencia. Este tipo de transformaciones se puede ver como la composición de una rotación y de un escalado no isotrópico. La matriz afín se puede descomponer siempre como:

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

$$\text{con } \mathbf{D} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

Los invariantes que se mantienen a través de este tipo de transformaciones son: líneas paralelas (se mantienen los puntos en el infinito), ratio de longitudes de segmentos de líneas paralelas y ratio de áreas.

2.3.4 Clase IV: Transformaciones proyectivas

La forma de una transformación proyectiva viene dada por la siguiente matriz:

$$\mathbf{x}' = \mathbf{H}_P \mathbf{x} = \begin{pmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^T & v \end{pmatrix} \mathbf{x}.$$

con $\mathbf{v} = (v_1, v_2)^T$. La matriz tiene ocho grados de libertad, con lo que se necesitan al menos 4 puntos en correspondencia.

El invariante proyectivo más importante es el *cross ratio* (ratio de ratios) de cuatro puntos colineales.

Las transformaciones proyectivas nos permiten trabajar con puntos en el infinito de igual forma que con cualquier otro punto.

El número de invariantes funcionalmente independientes para cada transformación es igual o mayor que el número de grados de libertad de la configuración menos el número de grados de libertad de la transformación.

Example 6 *Un cuadrado tiene 8 grados de libertad (2 por cada punto), por lo que posee 4 invariantes para transformaciones de similaridad, 2 para afines y cero para proyectivas.*

2.4 Geometría proyectiva 1D (\mathbb{P}^1)

Un punto en una línea se representa como $\mathbf{x} = (x_1, x_2)^T$. Cuando el segundo componente es igual a cero entonces se trata de un punto en el infinito. Una transformación proyectiva se representa como una matriz homogénea de 2×2 . Esta matriz tiene tres grados de libertad.

El cross ratio. Es el invariante básico proyectivo de \mathbb{P}^1 . Dados 4 puntos el cross ratio se define como:

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

con $|\mathbf{x}_1 \mathbf{x}_2|$ el determinante de los dos puntos. Algunos comentarios sobre el cross ratio:

1. El cross ratio es independiente del escalar que multiplica al punto ya que este se cancela en la división.
2. Si la segunda componente de los puntos es igual a 1 entonces el determinante entre dos puntos es igual a la distancia con signo entre esos dos puntos..
3. La definición del cross ratio también es válido para puntos ideales.
4. El valor del cross ratio es invariante bajo cualquier transformación proyectiva. Si $\mathbf{x}' = \mathbf{H}_{2 \times 2} \mathbf{x}$ entonces $\text{Cross}(\mathbf{x}'_1, \mathbf{x}'_2, \mathbf{x}'_3, \mathbf{x}'_4) = \text{Cross}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$.

Líneas Concurrentes. Las líneas que intersectan en un punto de un plano también tienen la geometría de \mathbb{P}^1 . El cross ratio de las líneas es igual al cross ratio de los puntos y es independiente de la posición de la línea l. ftbpFU162pt125.375pt0ptLínea Concurrentes.Figure

2.5 Recuperación de propiedades afines y métricas a partir de las imágenes

Si queremos realizar algún tipo de estimación acerca de longitudes de líneas o de ángulos en la imagen, entonces tenemos que eliminar la distorsión debida a la proyección de la escena en la imagen. Los grados de libertad de una transformación proyectiva es de 8 mientras que los de una transformación de similaridad es de 4, con lo que si queremos determinar algunas propiedades métricas de la imagen, entonces tenemos que eliminar los 4 grados de libertad de diferencia. Estos grados de libertad se pueden asociar a objetos geométricos concretos: La línea en el infito (\mathbf{l}_∞) que representa 2 dof y los dos *puntos circulares* (2 dof) que se encuentran sobre esa línea.

2.5.1 Línea en el infinito

En una transformación proyectiva, los puntos y líneas en el infinito se pueden transformar en puntos y líneas finitas. Si la transformación fuera una afinidad, entonces los puntos y líneas en el infinito se quedarían en el infinito.

La línea en el infinito \mathbf{l}_∞ se mantiene fija bajo una transformación proyectiva si y solo si la transformación es una afinidad.

2.5.2 Recuperación de propiedades afines

Una manera de recuperar las propiedades afines es localizar \mathbf{l}_∞ en la imagen y transformarla a su posición canónica $\mathbf{l}_\infty = (0, 0, 1)^T$. La matriz que logra esta transformación se puede aplicar a toda la imagen con el fin de rectificarla y poder tomar medidas afines. Si la imagen de la línea en el infinito es $\mathbf{l} = (l_1, l_2, l_3)$, entonces una transformación para eliminar el efecto proyectivo es:

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & l_3 \end{pmatrix} \mathbf{H}_A$$

ftbpFU235.875pt94.8125pt0ptRectificación afin.Figure

2.5.3 Puntos circulares y su dual

Bajo cualquier transformación de similituditud existen dos puntos en \mathbf{l}_∞ que se mantienen fijos. Estos son los puntos circulares \mathbf{I} , \mathbf{J} con coordenadas canónicas

$$\mathbf{I} = \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} \quad \mathbf{J} = \begin{pmatrix} 1 \\ -i \\ 0 \end{pmatrix}$$

Se ve fácilmente que son fijos bajo este tipo de transformaciones ya que si les aplicamos una transformación de similituditud cualquiera los puntos no cambian

$$\begin{aligned} \mathbf{I}' &= \mathbf{H}_S \mathbf{I} \\ &= \begin{pmatrix} s \cos \theta & -s \sin \theta & t_x \\ s \sin \theta & s \cos \theta & t_y \\ 0 & 0 & 1 \end{pmatrix} \mathbf{I} \\ &= s e^{-i\theta} \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix} = \mathbf{I}. \end{aligned}$$

Se llama puntos circulares porque todos los círculos intersectan a \mathbf{l}_∞ en esos dos puntos. La ecuación de un círculo en coordenadas homogéneas es

$$x_1^2 + x_2^2 + dx_1x_3 + ex_2x_3 + fx_3^2 = 0.$$

Esta ecuación intersecta a la línea en el infinito en los puntos ideales para los que $x_3 = 0$, con lo que nos queda

$$x_1^2 + x_2^2 = 0$$

que tiene por soluciones precisamente los puntos \mathbf{I} , \mathbf{J} .

Si identificamos los puntos circulares, entonces podemos recuperar las propiedades de similituditud de la imagen.

2.5.4 Ángulos en el plano proyectivo

El ángulo entre dos líneas $\mathbf{l} = (l_1, l_2, l_3)$ y $\mathbf{m} = (m_1, m_2, m_3)$ es

$$\cos \theta = \frac{l_1 m_1 + l_2 m_2}{\sqrt{(l_1^2 + l_2^2)(m_1^2 + m_2^2)}}$$

Esto no se puede aplicar después de una transformación afín o proyectiva. Existe una expresión análoga que sí es invariante a transformaciones proyectivas:

$$\cos \theta = \frac{\mathbf{l} \mathbf{C}_\infty^* \mathbf{m}}{\sqrt{(\mathbf{l}^T \mathbf{C}_\infty^* \mathbf{l})(\mathbf{m}^T \mathbf{C}_\infty^* \mathbf{m})}}$$

donde \mathbf{C}_∞^* es la cónica dual a los puntos circulares. Gracias a esta fórmula, si se puede identificar \mathbf{C}_∞^* en el plano proyectivo, entonces podemos calcular ángulos.

Dos líneas son ortogonales si se cumple que $\mathbf{l} \mathbf{C}_\infty^* \mathbf{m} = 0$.

También se pueden medir ratios de longitud si la cónica dual en el infinito es identificada.

2.5.5 Recuperación de propiedades métricas

Se pueden recuperar las propiedades métricas de una imagen si conseguimos transformar los puntos circulares a su forma canónica. Si conocemos los puntos circulares en una imagen podemos encontrar una transformación proyectiva, \mathbf{H} , que los convierta a $(1, \pm i, 0)$ en \mathbf{l}_∞ .

La cónica dual en el infinito \mathbf{C}_∞^* nos permite realizar una rectificación métrica, identificando los componentes afines y proyectivos de la transformación y dejando sólo las de similaridad. Si suponemos que un conjunto de puntos euclídeos se transforman a través de \mathbf{H} , entonces la cónica en el infinito \mathbf{C}_∞^* se transforma como $\mathbf{C}_\infty^{*'} = \mathbf{H} \mathbf{C}_\infty^* \mathbf{H}^T$. Descomponiendo la transformación tenemos

$$\begin{aligned} \mathbf{C}_\infty^{*'} &= \mathbf{H}_P \mathbf{H}_A \mathbf{H}_S \mathbf{C}_\infty^* (\mathbf{H}_P \mathbf{H}_A \mathbf{H}_S)^T \\ &= (\mathbf{H}_P \mathbf{H}_A) (\mathbf{H}_S \mathbf{C}_\infty^* \mathbf{H}_S^T) (\mathbf{H}_A^T \mathbf{H}_P^T) \\ &= (\mathbf{H}_P \mathbf{H}_A) \mathbf{C}_\infty^* (\mathbf{H}_A^T \mathbf{H}_P^T) \\ &= \begin{pmatrix} \mathbf{K} \mathbf{K}^T & \mathbf{K}^T \mathbf{v} \\ \mathbf{v}^T \mathbf{K} & 0 \end{pmatrix}. \end{aligned}$$

\mathbf{C}_∞^* es invariante frente a transformaciones de similaridad, con lo que la componente de similaridad está indefinida.

Por lo tanto, si podemos identificar \mathbf{C}_∞^* en la imagen, entonces la distorsión proyectiva se puede rectificar hasta una similaridad.

Una forma sencilla de encontrar esta transformación \mathbf{H} es por medio del SVD. El SVD de $\mathbf{C}_\infty^{*'}$ es

$$\mathbf{C}_\infty^{*'} = \mathbf{U} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{U}^T$$

con lo que la transformación para rectificar la imagen es $\mathbf{H} = \mathbf{U}$.

Para localizar la cónica dual en el infinito podemos especificar 5 pares de líneas ortogonales.

2.6 Otras propiedades de las cónicas

2.6.1 La relación pole-polar

ftbpFU164.0625pt142.0625pt0ptRelación pole-polarfi:polepolarFigure

Un punto \mathbf{x} y una cónica \mathbf{C} definen una recta $\mathbf{l} = \mathbf{C} \mathbf{x}$. Esta recta se denomina el *polar* de \mathbf{x} con respecto a \mathbf{C} , y el punto \mathbf{x} es el *pole* de \mathbf{l} con respecto a \mathbf{C} .

La línea polar intersecta a la cónica en dos puntos que coinciden con las rectas tangentes a la cónica que intersectan en \mathbf{x} (ver figura ??). Si el punto \mathbf{x} está en \mathbf{C} , entonces la línea polar coincide con la línea tangente a la cónica en el punto.

Example 7 Si suponemos un círculo de radio r centrado en el eje de las x en el punto a , su ecuación será $(x - a)^2 + y^2 = r^2$. La cónica que representa esta ecuación es

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & -a \\ 0 & 1 & 0 \\ -a & 0 & a^2 - r^2 \end{pmatrix}.$$

La línea polar del origen es $\mathbf{l} = \mathbf{C}(0, 0, 1)^T = (-a, 0, a^2 - r^2)^T$. Esto es una línea vertical en $x = (a^2 - r^2)/a$. Si $r = a$ entonces el origen está incluido en el borde del círculo y la línea polar es el eje de las y (que es tangente al círculo en el origen).

Puntos conjugados: Si un punto \mathbf{y} está en la línea $\mathbf{l} = \mathbf{C}\mathbf{x}$, entonces $\mathbf{y}^T \mathbf{l} = \mathbf{y}^T \mathbf{C}\mathbf{x} = 0$. Cualquier par de puntos \mathbf{x}, \mathbf{y} que satisfacen $\mathbf{y}^T \mathbf{C}\mathbf{x} = 0$ son puntos conjugados. Esta relación es simétrica, lo que quiere decir que si \mathbf{x} está en la línea polar de \mathbf{y} , entonces \mathbf{y} está en la polar de \mathbf{x} .

Existe una relación dual con las líneas conjugadas: dos líneas \mathbf{l}, \mathbf{m} son conjugadas si $\mathbf{l}^T \mathbf{C}^* \mathbf{m} = 0$.

3 Geometría proyectiva y transformaciones 3D

3.1 Puntos y transformaciones proyectivas

Un punto en \mathbb{P}^3 se representa en coordenadas homogéneas como $\mathbf{X} = (x_1, x_2, x_3, x_4)$. Este punto en el espacio \mathbb{R}^3 representa el punto $\mathbf{X} = (x_1/x_4, x_2/x_4, x_3/x_4)$.

Aquellos puntos que cumple que $x_4 = 0$ representan puntos en el infinito.

Una transformación proyectiva en \mathbb{P}^3 es una transformación lineal en coordenadas homogéneas representada por una matriz de 4×4 : $\mathbf{X}' = \mathbf{H}\mathbf{X}$. Esta matriz tiene 15 grados de libertad (16 menos uno por el escalado).

3.2 Planos, líneas y cuádricas

En \mathbb{P}^3 los puntos y los planos son duales. La líneas son duales entre sí.

Planos: Un plano se puede escribir como $\pi_1 x + \pi_2 y + \pi_3 z + \pi_4 = 0$, con lo que un plano tiene 3 grados de libertad. Si homogeneizamos esta ecuación obtenemos $\pi_1 x_1 + \pi_2 x_2 + \pi_3 x_3 + \pi_4 x_4 = 0$, o lo que es lo mismo $\pi^T \mathbf{X} = 0$. Las tres primeras coordenadas del plano representan la normal del plano.

Tres puntos definen un plano: si tenemos tres puntos \mathbf{X}_i que reposan sobre un plano π entonces tenemos un sistema de ecuaciones de la forma

$$\begin{pmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \\ \mathbf{X}_3^T \end{pmatrix} \pi = \mathbf{0}.$$

Otra forma de obtener un plano es a través del determinante de la matriz $\mathbf{M} = (\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$. Si el $\det \mathbf{M} = 0$, entonces significa que el punto \mathbf{X} es una combinación lineal de los otros tres, y, por tanto, reposa en el mismo plano que ellos

$$\det \mathbf{M} = x_1 D_{234} - x_2 D_{134} + x_3 D_{124} - x_4 D_{123},$$

luego el plano viene dado por

$$\pi = (D_{234}, -D_{134}, D_{124}, -D_{123})^T$$

Tres planos definen un punto: Se puede obtener el punto de intersección de tres planos a través del sistema de ecuaciones

$$\begin{pmatrix} \pi_1^T \\ \pi_2^T \\ \pi_3^T \end{pmatrix} \mathbf{X} = \mathbf{0}.$$

También podemos obtener el punto de forma análoga a como se obtiene un plano a partir de tres puntos descrito anteriormente.

Bajo una transformación proyectiva, $\mathbf{X}' = \mathbf{H}\mathbf{X}$, un plano se transforma como $\pi' = \mathbf{H}^{-1}\pi$.

Líneas: Una forma de representar las líneas en el espacio es a través de dos puntos, \mathbf{A}, \mathbf{B} . Podemos utilizar notación matricial para representarlas:

$$\mathbf{W} = \begin{pmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{pmatrix}.$$

Los puntos de la línea se forman como $(\lambda, \mu)\mathbf{W} = \lambda\mathbf{A} + \mu\mathbf{B}$. El espacio nulo por la derecha es el conjunto de planos que tienen esa línea como generadora: $\mathbf{W}\mathbf{P} = \mathbf{0}$, con $\mathbf{A}^T\mathbf{P} = 0$ y $\mathbf{B}^T\mathbf{P} = 0$, lo que significa que los dos puntos están en el plano.

Análogamente, si tenemos dos planos \mathbf{P}, \mathbf{Q} , entonces la matriz

$$\mathbf{W}^* = \begin{pmatrix} \mathbf{P}^T \\ \mathbf{Q}^T \end{pmatrix}$$

representa el conjunto de planos, $\lambda\mathbf{P} + \mu\mathbf{Q}^T$, que tienen la línea común de ambos como eje y el espacio nulo de la derecha, es el conjunto de puntos que reposan sobre la recta: $\mathbf{W}^*\mathbf{X} = \mathbf{0}$.

3.3 Plano en el infinito

El plano en el infinito en el espacio es similar a la línea en el infinito en 2D. Su forma canónica es $\pi_\infty = (0, 0, 0, 1)^T$.

Este nos permite identificar propiedades afines: dos planos son paralelos sí y solo sí su línea de intersección está en π_∞ ; una línea es paralela a otra línea, o a un plano, si el punto de intersección está en π_∞ .

El plano en el infinito π_∞ es un plano fijo bajo una transformación proyectiva \mathbf{H} , sí y solo sí \mathbf{H} es una afinidad.

3.4 La cónica absoluta

La cónica absoluta, Ω_∞ , es una cónica en π_∞ . Los puntos en Ω_∞ satisfacen

$$\left. \begin{array}{l} x_1^2 + x_2^2 + x_3^2 \\ x_4 \end{array} \right\} = 0$$

La cónica absoluta, Ω_∞ , es una cónica fija bajo una transformación proyectiva \mathbf{H} , sí y solo sí \mathbf{H} es una similitud.

Propiedades interesantes de la cónica absoluta:

1. Los puntos de la cónica absoluta permanecen en ella después de una transformación de similitud (sólo globalmente, no punto a punto).
2. Todos los círculos la intersectan en dos puntos.

3. Todas las esferas intersectan a π_∞ en Ω_∞ .

Propiedades métricas:

$$\cos \theta = \frac{\mathbf{l}\Omega_\infty\mathbf{m}}{\sqrt{(\mathbf{l}^T\Omega_\infty\mathbf{l})(\mathbf{m}^T\Omega_\infty\mathbf{m})}}$$

3.5 La cuádrica dual absoluta

La dual de la cónica absoluta es una cuádrica dual que se denota por \mathbf{Q}_∞^* y tiene la forma

$$\mathbf{Q}_\infty^* = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{pmatrix}$$

La cuádrica absoluta, \mathbf{Q}_∞^* , es una cónica fija bajo una transformación proyectiva \mathbf{H} , sí y solo sí \mathbf{H} es una similaridad.

El plano en el infinito π_∞ es el vector nulo de \mathbf{Q}_∞^* .

El ángulo entre dos planos viene dado por

$$\cos \theta = \frac{\pi_1^T \mathbf{Q}_\infty^* \pi_2}{\sqrt{(\pi_1^T \mathbf{Q}_\infty^* \pi_1)(\pi_2^T \mathbf{Q}_\infty^* \pi_2)}}$$

4 Estimación de transformaciones proyectivas

4.1 Algoritmo DLT

Vamos a desarrollar un algoritmo para calcular una matriz \mathbf{H} a partir de un conjunto de 4 puntos 2D en correspondencia, $\mathbf{x}'_i = \mathbf{H}\mathbf{x}_i$.

Si trabajamos en coordenadas homogéneas, sabemos que los vectores están multiplicados por un escalar, luego para resolver el anterior sistema nos fijamos en la dirección de los vectores, que tienen que cumplir que $\mathbf{x}'_i \times \mathbf{H}\mathbf{x}_i = \mathbf{0}$.

Si desarrollamos este producto y extraemos las incógnitas (elementos de \mathbf{H}), entonces obtenemos el siguiente sistema de ecuaciones

$$\begin{pmatrix} \mathbf{0}^T & -w'_i\mathbf{x}_i & y'_i\mathbf{x}_i \\ w'_i\mathbf{x}_i & \mathbf{0}^T & -x'_i\mathbf{x}_i \\ -y'_i\mathbf{x}_i & x'_i\mathbf{x}_i & \mathbf{0}^T \end{pmatrix} \begin{pmatrix} \mathbf{h}^1 \\ \mathbf{h}^2 \\ \mathbf{h}^3 \end{pmatrix} = \mathbf{0}.$$

Estas ecuaciones tienen la forma $\mathbf{A}_i\mathbf{h} = \mathbf{0}$, donde la matriz es de dimensión 3×9 y el vector tiene 9 elementos.

Este sistema tiene en realidad dos ecuaciones linealmente independientes, por lo que cada punto en correspondencia ofrece dos ecuaciones al sistema. Como la matriz \mathbf{H} tiene ocho grados de libertad, entonces necesitamos 4 puntos en correspondencia. Generalmente, los puntos que se toman para formar el sistema suelen contener cierto error, o muchas veces se tienen más de 4 puntos, con lo que es necesario aplicar un método de minimización por mínimos-cuadrados en el que se minimiza $\|\mathbf{A}\mathbf{h}\|$ con la restricción $\|\mathbf{h}\| = 1$.

Algorithm 8 *Algoritmo DLT: Dados $n \geq 4$ puntos 2D en correspondencia $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$ determinar la homografía \mathbf{H} tal que $\mathbf{x}'_i = \mathbf{H}\mathbf{x}_i$*

1. Para cada correspondencia $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ calcular la matriz \mathbf{A}_i (solo las dos primeras ecuaciones)
2. Ensamblar las n 2×9 matrices \mathbf{A}_i en una sola \mathbf{A} de $2n \times 9$
3. Obtener el SVD de \mathbf{A} . El vector singular correspondiente al valor singular menor es la solución para \mathbf{h}
4. Recolocar los elementos de \mathbf{h} para obtener \mathbf{H}

Este algoritmo depende de lo bien condicionado que esté la matriz del sistema, \mathbf{A} . Si ésta está mal condicionada, entonces la solución del mismo será errónea. Es por esto que a este algoritmo se le suele aplicar una normalización antes de calcular la homografía. Esta normalización consiste en aplicar un escalado y una traslación a los puntos en correspondencia antes de formar la matriz y resolver el sistema. El algoritmo quedaría de la siguiente manera:

Algorithm 9 *Algoritmo DLT normalizado: Dados $n \geq 4$ puntos 2D en correspondencia $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$ determinar la homografía \mathbf{H} tal que $\mathbf{x}'_i = \mathbf{H}\mathbf{x}_i$*

1. **Normalización \mathbf{x} :** transformar los puntos \mathbf{x} a través de una transformación de similaridad \mathbf{T} , que consiste en un escalado y traslación, de forma que se obtenga un conjunto de puntos $\tilde{\mathbf{x}}$ tal que su centroide recaiga en el origen y su distancia media sea de $\sqrt{2}$.
2. **Normalización \mathbf{x}' :** realizar la misma operación que en el paso anterior para los puntos \mathbf{x}' . En este paso se obtiene una transformación de similaridad \mathbf{T}' y los puntos transformados $\tilde{\mathbf{x}}'$
3. **DLT:** aplicar el algoritmo anterior a las correspondencias $\tilde{\mathbf{x}}_i \leftrightarrow \tilde{\mathbf{x}}'_i$ para obtener la matriz $\tilde{\mathbf{H}}$
4. **Denormalización:** deshacer las transformaciones de normalización de forma $\mathbf{H} = \mathbf{T}'^{-1}\tilde{\mathbf{H}}\mathbf{T}$

Esta normalización es necesaria y no se debe considerar un paso excepcional.

4.2 Estimación robusta

Cuando queremos realizar algoritmos en los que todos los pasos, desde la puesta en correspondencia, hasta el cálculo de las matrices, se calculan de forma totalmente automática, entonces es posible que nos encontremos con *outliers*. Los outliers son puntos que se a través de una algoritmo se han estimado que están en correspondencia pero que realmente no lo están. Esto tiene el inconveniente de que si intentamos solucionar el sistema de ecuaciones con puntos de este estilo, entonces obtendremos soluciones completamente erróneas.

Existen métodos (robustos) que tienen en cuenta este hecho e intentan identificar los outliers con el fin de que no se tengan en cuenta. Se clasifican los conjuntos de puntos entre outliers e inliers.

4.2.1 RANSAC

El método RANSAC (RANdom SAmple Consensus) es el más conocido y se basa en ir cogiendo pequeños conjuntos de puntos (el menor número permitido), en calcular el modelo a partir de esos puntos, y en estimar cuántos puntos del conjunto completo están dentro de una distancia mínima con respecto a ese modelo. Este proceso se repite hasta obtener un conjunto de puntos grandes que cumplan con el modelo.

Algorithm 10 *Obtención robusta de un modelo asociado a un conjunto de datos, S , con outliers*

1. Se selecciona aleatoriamente una muestra de s puntos a partir de S y se calcula el modelo para este subconjunto
2. Se determina el conjunto de puntos S_i que están dentro de una distancia umbral t del modelo. El conjunto S_i es el consenso de la muestra y define los inliers de S .
3. Si el tamaño de S_i (número de inliers) es más grande que un umbral T , se recalcula el modelo utilizando todos sus puntos y se termina
4. Si el tamaño de S_i es más pequeño que el umbral T , se selecciona un nuevo subconjunto y se repite lo de arriba
5. Después de N intentos el consenso S_i mayor se selecciona y el modelo se recalcula utilizando todos los puntos en el subconjunto S_i

En este algoritmo tenemos tres umbrales (T, N, t) . El parámetro t depende del tipo de modelo que estemos estimando, p.e. si el modelo es una recta, entonces el umbral se compara con la distancia de un punto a la recta $d(\mathbf{p}, \mathbf{l}) < t$. Para el parámetro T se suele suponer un porcentaje de puntos que puedan ser outliers. Entonces $T = (1 - \varepsilon)n$, lo que representa el número de puntos que nosotros consideramos el máximo de inliers.

El parámetro N es más complejo puesto que puede incrementar considerablemente el tiempo de ejecución del algoritmo. Por un lado no puede ser muy pequeño porque la muestra final seleccionada puede no ser la más conveniente, mientras que por el otro, si es muy grande el algoritmo puede tardar demasiado tiempo.

Se puede adaptar N iterativamente, atendiendo a la probabilidad de tener en cuenta a los inliers, de la siguiente manera:

1. $N = \infty$, $sample_count = 0$
2. Mientras $N > sample_count$ hacer
 - (a) Elegir una muestra y contar el número de inliers
 - (b) Colocar $\varepsilon = 1 - (n_inliers)/(puntos_totales)$
 - (c) Igualar $N = \log(1 - p)/\log(1 - (1 - \varepsilon)^s)$ con $p = 0.99$
 - (d) Incrementar $sample_count$ en 1
3. Terminar

4.3 Cálculo automático de una homografía

En el siguiente algoritmo se propone una manera de calcular de forma automática una homografía entre dos imágenes.

Algorithm 11 *Cálculo de una homografía 2D entre dos imágenes*

1. **Puntos de interés:** se calculan puntos de interés en ambas imágenes
2. **Correspondencias potenciales:** se ponen en correspondencia dichos puntos de interés atendiendo a su proximidad y parecido
3. **Estimación robusta por RANSAC:** repetir para N muestras

- (a) Seleccionar aleatoriamente una muestra de 4 correspondencias y calcular la homografía \mathbf{H}
- (b) Se calcula la distancia d_{\perp} para cada correspondencia potencial
- (c) Se calcula el número de inliers consistentes con \mathbf{H} para el número de correspondencias para las cuales $d_{\perp} < t = \sqrt{5.99}\sigma$ pixels

Se elige la \mathbf{H} con el mayor número de inliers.

4. **Estimación óptima:** se recalcula \mathbf{H} para todas las correspondencias clasificadas como inliers minimizando la función de coste ML utilizando Levenberg-Marquardt
5. **Correspondencia guiada:** se seleccionan nuevos puntos de interés en correspondencia utilizando \mathbf{H}

5 Geometría de una vista simple - Modelo de la cámara

En esta sección estudiaremos cómo se modela una cámara. Básicamente podemos clasificar las cámaras en dos conjuntos: Cámaras con centro finito, que son aquellas cuyo centro está a una distancia finita del plano proyectivo y que se puede modelar a través de una matriz de proyección convencional; Cámaras con centro en el infinito, denominada cámara afín, que son aquellos que proyectan los puntos en el plano proyectivo por medio de líneas paralelas.

5.1 Cámara finita

El modelo pinhole: Los puntos del espacio se proyectan a través de un punto en el plano proyectivo (ver figura ??).

ftbpFU168.3125pt108.625pt0ptEl modelo pinholefi:pinholeFigure

La proyección del espacio \mathbb{R}^3 al espacio \mathbb{R}^2 se realiza como $(x, y, z)^T \mapsto (fx/z, fy/z)^T$. El centro de proyección se denomina el *centro de cámara*. La línea perpendicular al plano de la imagen que pasa por el centro de la cámara se denomina *eje o rayo principal* y el punto de intersección entre éste y el plano de la imagen se llama el *punto principal*. El *plano principal* es el que pasa por el centro de la cámara y es paralelo al plano de la imagen.

Utilizando coordenadas homogéneas, la proyección se expresa como

$$\begin{aligned} \mathbf{x} &= \begin{pmatrix} fx \\ fy \\ z \end{pmatrix} = \begin{pmatrix} f & 0 \\ f & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \\ &= \mathbf{P}\mathbf{X} \\ &\mathbf{x} = \mathbf{P}\mathbf{X} \end{aligned}$$

También se puede representar como $\text{diag}(f, f, 1)[\mathbf{I}|\mathbf{0}]$.

Si el origen del plano de la imagen está desplazado del origen con respecto al centro de la cámara por (p_x, p_y) , entonces los puntos se transforman como $(x, y, z)^T \mapsto (fx/z + p_x, fy/z + p_y)^T$, que si lo representamos en coordenadas homogéneas tenemos

$$\begin{pmatrix} fx + zp_x \\ fy + zp_y \\ z \end{pmatrix} = \begin{pmatrix} f & p_x & 0 \\ f & p_y & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

Si llamamos

$$\mathbf{K} = \begin{pmatrix} f & p_x \\ & f & p_y \\ & & & 1 \end{pmatrix}$$

entonces $\mathbf{x} = \mathbf{K}[\mathbf{I}|\mathbf{0}]\mathbf{X}$. Esta matriz \mathbf{K} se denomina *matriz de calibración de la cámara*. Si además suponemos que el eje de coordenadas de la cámara no está situado en el eje de coordenadas universal, entonces tenemos que expresar los puntos 3D de la escena en el eje de referencia de la cámara. Los dos ejes de coordenadas están relacionados a través de una rotación y una traslación (ver figura ??).

ftbpFU207.4375pt201.3125pt0ptRelación entre el eje de coordenadas de la cámara y el eje de coordenadas universal:RtFigure

La relación entre un punto de la cámara y el correspondiente universal viene dado por $\mathbf{X}_{cam} = \mathbf{R}(\mathbf{X} - \mathbf{C})$

$$\begin{aligned} \mathbf{X}_{cam} &= \begin{pmatrix} \mathbf{R} & -\mathbf{RC} \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{R} & -\mathbf{RC} \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{X} \end{aligned}$$

Si ponemos este resultado con el que habíamos obtenido anteriormente, tenemos que $\mathbf{P} = \mathbf{KR}[\mathbf{I} - \tilde{\mathbf{C}}]$. Esta matriz tiene 9 grados de libertad (3 para \mathbf{K} , 3 para \mathbf{R} y 3 para \mathbf{C}). Los parámetros de \mathbf{K} se denominan parámetros *internos o intrínsecos* de la cámara y los de \mathbf{R} y \mathbf{C} se llaman parámetros *externos o extrínsecos*.

cámaras CCD: En este tipo de cámaras es posible que los pixels de la imagen no sean cuadrados. En este caso, la escala para cada dirección es distinta. Esto se modela utilizando dos factores m_x y m_y , uno para cada dirección obteniendo

$$\mathbf{K} = \begin{pmatrix} \alpha_x & & x_0 \\ & \alpha_y & y_0 \\ & & & 1 \end{pmatrix}$$

donde $\alpha_x = fm_x$ y $\alpha_y = fm_y$ representan la distancia focal de la imagen en términos de la dimensión del pixel. Además, (x_0, y_0) es el punto principal en términos de la dimensión del pixel, con $x_0 = m_x p_x$, $y_0 = m_y p_y$. Este tipo de cámaras tiene por tanto, 10 grados de libertad.

cámara proyectiva finita: Podemos aumentar la generalidad de esta transformación si añadimos otro parámetro en la matriz

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

El parámetro s se denomina *skew*. Gracias a éste podemos modelar el caso en que los ejes de la cámara no sean perfectamente ortogonales, aunque en la mayoría de los casos éste va a ser igual a cero. En este caso tenemos 11 grados de libertad.

5.2 Cámara proyectiva

Podemos representar la matriz de proyección como $\mathbf{P} = [\mathbf{M}|\mathbf{p}_4]$. Si \mathbf{M} no es singular entonces se trata de una cámara finita.

El centro de la cámara cumple la siguiente relación $\mathbf{PC} = \mathbf{0}$. Si se trata de una cámara finita (\mathbf{M} no es singular), entonces $\mathbf{C} = \begin{pmatrix} -\mathbf{M}^{-1}\mathbf{p}_4 \\ 1 \end{pmatrix}$. Si la cámara es infinita (\mathbf{M} es singular), entonces $\mathbf{C} = \begin{pmatrix} \mathbf{d} \\ 0 \end{pmatrix}$ donde \mathbf{d} es el vector nulo de $\mathbf{M}\mathbf{d} = \mathbf{0}$.

Los vectores columnas de la matriz \mathbf{P} , $(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3)$ son los puntos en el infinito de los ejes X, Y, Z del sistema de coordenadas universal. Estos puntos son la proyección de las direcciones de los ejes. el eje X tiene la dirección $\mathbf{D} = (1, 0, 0, 0)$ cuya imagen es $\mathbf{p}_1 = \mathbf{PD}$.

Los vectores fila se corresponden con los planos del sistema universal

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} = \begin{pmatrix} \mathbf{P}^{1T} \\ \mathbf{P}^{2T} \\ \mathbf{P}^{3T} \end{pmatrix}$$

El *plano principal* es aquel que incluye al centro de la cámara y es paralelo al plano de la imagen. Este contiene al conjunto de puntos \mathbf{X} que se proyectan a la línea en el infinito de la imagen, $\mathbf{PX} = (x, y, 0)$. Por lo tanto, un punto reside en el plano principal sí y solo sí $\mathbf{P}^{3T}\mathbf{X} = 0$. O lo que es lo mismo, \mathbf{P}^3 representa el plano principal.

ftbpF182.3125pt169.75pt0ptFigure

Los puntos en el plano \mathbf{P}^1 satisfacen $\mathbf{P}^{1T}\mathbf{X} = 0$ y se proyectan como $\mathbf{PX} = (0, y, w)^T$. El centro de la cámara se encuentra en la intersección de los tres planos.

El *punto principal* es la intersección del eje principal (línea que pasa por \mathbf{C} y es perpendicular a \mathbf{P}^3) con el plano de la imagen. Para calcularlo podemos tomar la normal al plano $\mathbf{P}^3 : (p_1, p_2, p_3)$. Este vector se puede representar alternativamente con el punto en el infinito $\hat{\mathbf{P}}^3 = (p_1, p_2, p_3, 0)^T$, luego proyectando este punto a través de la matriz de proyección obtenemos

$$\mathbf{x}_0 = \mathbf{P}\hat{\mathbf{P}}^3 = [\mathbf{M}|\mathbf{p}_4] \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ 0 \end{pmatrix} = \mathbf{M} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$$

y como $(p_1, p_2, p_3)^T$ es igual a la tercera fila de \mathbf{M} , \mathbf{m}^{3T} , entonces nos queda que $\mathbf{x}_0 = \mathbf{M}\mathbf{m}^3$.

Para encontrar el rayo de proyección que pasa por el centro de la cámara e intersecta al plano de la imagen en un punto \mathbf{x} podemos utilizar dos puntos. El primero es el centro de la cámara \mathbf{C} y el otro puede ser el dado por $\mathbf{P}^+\mathbf{x}$, con \mathbf{P}^+ la pseudoinversa de \mathbf{P} , $\mathbf{P}^+ = \mathbf{P}^T(\mathbf{P}\mathbf{P}^T)^{-1}$. El rayo formado por estos dos puntos sería $\mathbf{X}(\lambda) = \mathbf{P}^+\mathbf{x} + \lambda\mathbf{C}$.

Si tuviésemos una cámara finita, podríamos tomar los puntos $\tilde{\mathbf{C}} = -\mathbf{M}^{-1}\mathbf{p}_4$ y el vector director $\mathbf{D} = ((\mathbf{M}^{-1}\mathbf{x})^T, 0)$, con lo que la recta se representaría como $\mathbf{X}(\lambda) = \begin{pmatrix} \mathbf{M}^{-1}(\lambda\mathbf{x} - \mathbf{p}_4) \\ 1 \end{pmatrix}$.

5.3 Cámaras en el infinito

Son aquellas cuyo centro está en el plano en el infinito. Esto significa que la matriz \mathbf{M} es singular. El centro de la cámara se puede encontrar igual que antes como $\mathbf{PC} = \mathbf{0}$.

Las cámaras en el infinito se pueden clasificar como cámaras afines o no-afines.

5.3.1 Cámara afines

Cámaras afines son aquellas que tiene la tercera fila de la matriz \mathbf{P} de la forma $(0, 0, 0, 1)$. Se denominan afines porque los puntos en el infinito se corresponden con puntos en el infinito en la imagen.

$$\mathbf{P}_\infty = \mathbf{K}\mathbf{R}[\mathbf{I} | -\tilde{\mathbf{C}}] = \mathbf{K} \begin{pmatrix} \mathbf{r}^{1T} & -\mathbf{r}^{1T}\tilde{\mathbf{C}} \\ \mathbf{r}^{2T} & -\mathbf{r}^{2T}\tilde{\mathbf{C}} \\ \mathbf{r}^{3T} & -\mathbf{r}^{3T}\tilde{\mathbf{C}} \end{pmatrix}$$

La matriz \mathbf{K} tiene la forma siguiente $\mathbf{K} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 0 \end{pmatrix}$ de forma que la proyección en x e y

son idénticas a las coordenadas originales del punto en el espacio y la matriz $\mathbf{P} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 0 & 1 \end{pmatrix}$ y de forma general,

$$\mathbf{P}_\infty = \mathbf{K}\mathbf{R}[\mathbf{I} | -\tilde{\mathbf{C}}] = \mathbf{K} \begin{pmatrix} \mathbf{r}^{1T} & -\mathbf{r}^{1T}\mathbf{C} \\ \mathbf{r}^{2T} & -\mathbf{r}^{2T}\mathbf{C} \\ \mathbf{0}^T & d_0 \end{pmatrix}$$

6 Cálculo de la matriz P

6.1 Ecuaciones básicas

Dado un conjunto de puntos 3D puestos en correspondencia con puntos en 2D, $\mathbf{X}_i \leftrightarrow \mathbf{x}_i$, el problema consiste en encontrar la matriz \mathbf{P} que cumpla $\mathbf{x}_i = \mathbf{P}\mathbf{X}_i$ para todos los puntos. Debido a que la relación es válida para cualquier escalar que multiplique a la relación, entonces nos interesa que la dirección de los vectores sea la misma, por lo que podemos utilizar el producto escalar

$$\mathbf{x}_i \times \mathbf{P}\mathbf{X}_i = 0$$

Nuestras incógnitas son los parámetros de la matriz \mathbf{P} .

$$\begin{pmatrix} \mathbf{0}^T & -w_i\mathbf{X}_i^T & y_i\mathbf{X}_i^T \\ w_i\mathbf{X}_i^T & \mathbf{0}^T & -x_i\mathbf{X}_i^T \\ -y_i\mathbf{X}_i^T & x_i\mathbf{X}_i^T & \mathbf{0}^T \end{pmatrix} \begin{pmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \\ \mathbf{P}^3 \end{pmatrix} = \mathbf{0}$$

Podemos seleccionar dos de las tres ecuaciones anteriores debido que dependen linealmente entre sí. Para cada punto obtenemos dos ecuaciones con 12 incógnitas cada una. Este sistema se puede resolver como un sistema de ecuaciones lineal de la forma $\mathbf{A}\mathbf{p} = 0$, donde \mathbf{p} contiene todos los elementos de la matriz de proyección.

Se necesitan en total 11 ecuaciones para determinar los 11 grados de libertad de la matriz \mathbf{P} . Por lo tanto necesitamos al menos 5 puntos y medio para poder calcularla.

Si los datos de las correspondencias no son exactos o tenemos más de los puntos necesarios (≥ 6), entonces necesitamos aplicar algún método de minimización. Para esto se minimiza la norma de $\|\mathbf{A}\mathbf{p}\|$ con algún tipo de restricción de la forma $\|\mathbf{p}\| = 1$.

Es importante realizar la normalización de los datos antes de calcular la matriz de forma que el sistema esté bien condicionado. Generalmente se suele trasladar los puntos de forma que su

centroide caiga en el centro del sistema de coordenadas y se escalan de forma que su distancia media al centroide diste $\sqrt{2}$.

Podemos minimizar el error geométrico dado por $\min_{\mathbf{P}} \sum_i d(\mathbf{x}_i, \mathbf{P}\mathbf{X}_i)^2$. Esta minimización requiere el uso de técnicas iterativas como Levenberg-Marquardt.

Muchas veces se suele utilizar algún tipo de patrón, del estilo de la figura ??, para calibrar una cámara.

ftbpFU222.4375pt167.25pt0ptCalibrador de cuadradosfi:calibradorFigure

Se seleccionan las esquinas de cada cuadrado y se ponen en correspondencia con sus posiciones en 3D (conocidas). Para seleccionar las esquinas, generalmente se detectan primero los contornos de la imagen, luego se detectan las líneas, para luego encontrar las intersecciones de las mismas.

Algorithm 12 *Cálculo de la matriz \mathbf{P} a partir de un conjunto de correspondencias, $\{\mathbf{X}_i \leftrightarrow \mathbf{x}_i\}$*

1. *Solución lineal: Se calcula \mathbf{P} usando un método lineal*

(a) *Normalización: Normalizar los puntos 2D y 3D de la forma $\tilde{\mathbf{x}}_i = \mathbf{T}\mathbf{x}_i$ y $\tilde{\mathbf{X}}_i = \mathbf{U}\mathbf{X}_i$*

(b) *DLT: Formar la matriz \mathbf{A} de $2n \times 12$ con todas las correspondencias. La solución para $\mathbf{A}\mathbf{p} = 0$ con $\|\mathbf{p}\| = 1$ viene dado por el vector singular unitario de \mathbf{A} correspondiente a su valor singular más pequeño.*

2. *Minimización del error geométrico: Utilizando la aproximación lineal calcular $\min_{\mathbf{P}} \sum_i d(\tilde{\mathbf{x}}_i, \tilde{\mathbf{P}}\tilde{\mathbf{X}}_i)^2$ utilizando un método iterativo.*

3. *Denormalización: La matriz de proyección original es igual a $\mathbf{P} = \mathbf{T}^{-1}\tilde{\mathbf{P}}\mathbf{U}$*

7 Geometría epipolar y matriz fundamental

7.1 Geometría epipolar

ftbpF212pt121.625pt0ptFigure

Epipolo: El punto de intersección (e) de la línea que une los dos centros de cámara con el plano de la imagen.

Plano epipolar: Plano que contiene a la línea base (línea que une los dos centros de cámara y contiene a los epipolos). Existe una familia de planos epipolares (pencil of planes).

Línea epipolar: La intersección de un plano epipolar con el plano de la imagen.

7.2 La matriz fundamental

Es la representación algebraica de la geometría epipolar. Para cada punto en una imagen existe una línea en la otra imagen en la que está incluido el punto en correspondencia. Luego existe una correspondencia entre un punto de una imagen y una línea en la otra. Este correspondencia se representa a través de una matriz \mathbf{F} , la matriz fundamental, de dimensión 3×3 .

7.2.1 Deducción geométrica

ftbpF212pt123.4375pt0ptFigure

La línea epipolar que pasa por \mathbf{x}' y \mathbf{e}' se puede escribir como $\mathbf{l}' = \mathbf{x}' \times \mathbf{e}' = [\mathbf{e}']_{\times} \mathbf{x}'$. Como $\mathbf{x}' = \mathbf{H}_{\pi} \mathbf{x}$, entonces $\mathbf{l}' = [\mathbf{e}']_{\times} \mathbf{H}_{\pi} \mathbf{x} = \mathbf{F} \mathbf{x}$, por lo que $\mathbf{F} = [\mathbf{e}']_{\times} \mathbf{H}_{\pi}$.

La matriz fundamental tiene rango 2 ya que $[\mathbf{e}']_{\times}$ tiene rango 2 y \mathbf{H}_{π} rango 3.

7.2.2 Deducción algebraica

Otra forma de deducir la matriz es a través de la línea que une el punto de la imagen con el punto 3D

$$\mathbf{X}(\lambda) = \mathbf{P}^+ \mathbf{x} + \lambda \mathbf{C}$$

$$\begin{aligned} \mathbf{l}' &= \mathbf{P}' \mathbf{C} \times \mathbf{P}' \mathbf{P}^+ \mathbf{x} = [\mathbf{e}']_{\times} \mathbf{P}' \mathbf{P}^+ \mathbf{x} = \mathbf{F} \mathbf{x} \\ \mathbf{F} &= [\mathbf{e}']_{\times} \mathbf{P}' \mathbf{P}^+ \end{aligned}$$

Luego $\mathbf{H}_{\pi} = \mathbf{P}' \mathbf{P}^+$

7.2.3 Condición de correspondencia

Para cualquier par de puntos en correspondencia entre las dos imágenes, la matriz fundamental cumple

$$\mathbf{x}'^T \mathbf{F} \mathbf{x} = 0$$

Podemos, por tanto, calcular la matriz \mathbf{F} a partir de puntos en correspondencia entre las imágenes.

7.2.4 Propiedades de la matriz fundamental

Traspuesta: Si \mathbf{F} es la matriz fundamental asociada al par de cámaras $(\mathbf{P}, \mathbf{P}')$, entonces \mathbf{F}^T es la asociada a $(\mathbf{P}', \mathbf{P})$.

Líneas epipolares: La línea epipolar correspondiente a un punto \mathbf{x} es $\mathbf{l}' = \mathbf{F} \mathbf{x}$. La línea epipolar para un punto \mathbf{x}' es $\mathbf{l} = \mathbf{F}^T \mathbf{x}'$.

Epipolos: Para cualquier punto \mathbf{x} que no sea el epipolo (\mathbf{e}), la línea epipolar $\mathbf{l}' = \mathbf{F} \mathbf{x}$ contiene el epipolo \mathbf{e}' . Esto satisface $\mathbf{e}'^T \mathbf{F} \mathbf{x} = (\mathbf{e}'^T \mathbf{F}) \mathbf{x} = 0$ para cualquier \mathbf{x} , lo que implica que $\mathbf{e}'^T \mathbf{F} = 0$ y, análogamente, $\mathbf{F} \mathbf{e} = 0$.

Grados de libertad: \mathbf{F} tiene 7 grados de libertad ya que $\det \mathbf{F} = 0$.

\mathbf{F} es una *correlación*.

7.2.5 Homografía de la línea epipolar

El conjunto de líneas epipolares intersectan en el epipolo (ver figura ??).ftbpFU208.1875pt110.5625pt0ptLineas epipolaresfi:lineas_eepipolaresFigure

Como se puede ver en la figura, las líneas epipolar están relacionadas a través de una transformación proyectiva, con lo que existe una homografía entre las líneas eipolares de una imagen y de la otra. Esta homografía tiene 3 grados de libertad. Los grados de libertad de la matriz fundamental se pueden contar como 2 para \mathbf{e} otros 2 para \mathbf{e}' y 3 para la homografía.

Si suponemos que hay dos líneas epipolares correspondientes, \mathbf{l} y \mathbf{l}' , y \mathbf{k} es una línea cualquiera que no pasa por \mathbf{e} , entonces \mathbf{l} y \mathbf{l}' están relacionadas por $\mathbf{l}' = \mathbf{F} [\mathbf{k}]_{\times} \mathbf{l}$ o, simétricamente, por $\mathbf{l} = \mathbf{F}^T [\mathbf{k}']_{\times} \mathbf{l}'$.

Una elección conveniente para \mathbf{k} es la línea \mathbf{e} , ya que $\mathbf{k}^T \mathbf{e} = \mathbf{e}^T \mathbf{e} \neq 0$, por lo que no pasa por el epipolo. La homografía de la línea epipolar se puede escribir como

$$\mathbf{l}' = \mathbf{F} [\mathbf{e}]_{\times} \mathbf{l} \quad \mathbf{l} = \mathbf{F}^T [\mathbf{e}']_{\times} \mathbf{l}'$$

7.3 Matrices fundamentales especiales

7.3.1 Traslación pura

Cuando los parámetros intrínsecos de las cámaras son iguales y una de las cámaras está desplazada con respecto a la otra por un vector \mathbf{t} sin rotación. Podemos suponer que las dos cámaras vienen dadas por $\mathbf{P} = \mathbf{K}[\mathbf{I}|\mathbf{0}]$ y $\mathbf{P}' = \mathbf{K}[\mathbf{I}|\mathbf{t}]$. La matriz fundamental quedaría como

$$\mathbf{F} = [\mathbf{e}']_{\times} \mathbf{K} \mathbf{K}^{-1} = [\mathbf{e}']_{\times}$$

Si el desplazamiento de la segunda cámara se realiza en la dirección del eje x , entonces $\mathbf{e}' = (1, 0, 0)^T$, así que

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

La relación entre puntos en correspondencia se reduce a $y = y'$, o lo que es lo mismo, las líneas epipolares son horizontales y paralelas al eje x .

7.4 Cálculo de las matrices de las cámaras

Una de las propiedades más significantes de la matriz \mathbf{F} es que nos permite calcular las matrices de las cámaras.

Si H es una matriz de 4×4 que representa una transformación proyectiva del espacio 3D, entonces las matrices fundamentales que corresponden al par de matrices de cámaras $(\mathbf{P}, \mathbf{P}')$ y $(\mathbf{PH}, \mathbf{P}'H)$ son la misma. De esta forma, aunque las matrices de dos cámaras identifican unívocamente una matriz fundamental, lo inverso no es cierto. Una matriz fundamental define un conjunto de matrices de cámaras que están relacionadas por una transformación proyectiva.

Forma canónica de las matrices de las cámaras: Podemos suponer que una de las cámaras está normalizada en el sentido de que sus parámetros intrínsecos son iguales a la unidad y que está situada en el origen de coordenadas apuntando en la dirección del eje z . La matriz para esta cámara es de la forma $\mathbf{P} = [\mathbf{I}|\mathbf{0}]$. Si representamos la otra matriz de la forma $\mathbf{P}' = [\mathbf{M}|\mathbf{m}]$, entonces la matriz fundamental correspondiente a ellas es $\mathbf{F} = [\mathbf{m}]_{\times} \mathbf{M}$.

Se pueden seleccionar las matrices de cámara a partir de la matriz fundamental de la siguiente forma: $\mathbf{P} = [\mathbf{I}|\mathbf{0}]$ y $\mathbf{P}' = [[\mathbf{e}']_{\times} \mathbf{F} | \mathbf{e}']$.

La fórmula general para un par de matrices de cámara canónicas viene dada por:

$$\mathbf{P} = [\mathbf{I}|\mathbf{0}] \quad \mathbf{P}' = [[\mathbf{e}']_{\times} \mathbf{F} + \mathbf{e}' \mathbf{v}^T | \lambda \mathbf{e}']$$

7.5 Matriz esencial

La matriz esencial es la especialización de la matriz fundamental al caso en que las coordenadas de la imagen están normalizadas. Coordenadas normalizadas significa que la matriz \mathbf{K} es igual a la identidad. Una matriz de cámara se puede descomponer como $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$. Si conocemos la matriz de calibración los puntos proyectados, $\mathbf{x} = \mathbf{P}\mathbf{X} = \mathbf{K}[\mathbf{R}|\mathbf{t}]\mathbf{X}$, se pueden representar como $\hat{\mathbf{x}} = \mathbf{K}^{-1}\mathbf{x}$, con lo que tenemos $\hat{\mathbf{x}} = [\mathbf{R}|\mathbf{t}]\mathbf{X}$. Estos puntos están en coordenadas normalizadas. Si consideramos el par de matrices $\mathbf{P} = [\mathbf{I}|\mathbf{0}]$ y $\mathbf{P}' = [\mathbf{R}|\mathbf{t}]$, la matriz fundamental que se obtiene se denomina *esencial* y tiene la forma

$$\mathbf{E} = [\mathbf{t}]_{\times} \mathbf{R}$$

Se cumple la relación $\hat{\mathbf{x}}'^T \mathbf{E} \hat{\mathbf{x}} = 0$. Si sustituimos por los puntos originales, obtenemos $\mathbf{x}'^T (\mathbf{K}'^{-T} \mathbf{E} \mathbf{K}^{-1}) \mathbf{x} = 0$ con lo que la relación con la matriz fundamental es

$$\mathbf{E} = \mathbf{K}'^T \mathbf{F} \mathbf{K}.$$

La matriz esencial tiene 5 grados de libertad

8 Reconstrucción 3D de cámaras

8.1 Método de reconstrucción

Una forma de reconstruir una escena a partir de dos vistas se puede realizar siguiendo los pasos

1. Calcular la matriz fundamental a partir de puntos en correspondencia
2. Calcular las matrices de las cámaras a partir de la matriz fundamental
3. Para cada par de puntos en correspondencia se calcula el punto en el espacio

8.2 Ambigüedad en la reconstrucción

A partir de dos vistas no podemos situar la posición y orientación de la escena con respecto a un eje de coordenadas universal. Esto significa que la reconstrucción está supeditada a una transformación Euclídea desconocida.

Tampoco es posible determinar el escalado general de la escena. Para cámaras calibradas es posible realizar la reconstrucción hasta una transformación de similaridad, donde no conocemos ni la rotación, ni el desplazamiento, ni el escalado de la escena.

Si las cámaras no están calibradas, entonces la ambigüedad en la reconstrucción sería posible hasta una transformación proyectiva. Si las cámaras se relacionan a través de un movimiento traslacional, entonces la ambigüedad es hasta una transformación afín y si no se conocen las distancias focales de las cámaras, entonces sería hasta una transformación de similaridad.

Dependiendo de qué tipo de ambigüedad se trate, hablamos de *reconstrucción proyectiva*, *reconstrucción afín*, *reconstrucción de similaridad* y *reconstrucción Euclídea*.

8.3 Teorema de la reconstrucción proyectiva

Theorem 13 *Suponemos que tenemos un conjunto de correspondencias $\mathbf{x}_i \longleftrightarrow \mathbf{x}'_i$ entre puntos de dos imágenes y que determinamos unívocamente la matriz fundamental \mathbf{F} , entonces para cualquier par de reconstrucciones $(\mathbf{P}_1, \mathbf{P}'_1, \{\mathbf{X}_{1i}\})$ y $(\mathbf{P}_2, \mathbf{P}'_2, \{\mathbf{X}_{2i}\})$ existe una matriz no singular \mathbf{H} que cumple que $\mathbf{P}_2 = \mathbf{P}_1 \mathbf{H}^{-1}$, $\mathbf{P}'_2 = \mathbf{P}'_1 \mathbf{H}^{-1}$ y $\mathbf{X}_{2i} = \mathbf{H} \mathbf{X}_{1i}$.*

Este teorema es importante porque es la base para realizar una reconstrucción proyectiva sólo a partir de un conjunto de puntos en correspondencia. Este tipo de reconstrucción es suficiente para realizar distintos tipos de aplicaciones como localizar la intersección entre líneas y planos. Además también es interesante porque a partir de ésta se pueden obtener la reconstrucción afín y posteriormente la métrica.

8.4 Reconstrucción estratificada

Esta reconstrucción consiste en primero estimar una reconstrucción proyectiva, luego refinarla a una transformación afín y posteriormente a una métrica.

Para refinar la reconstrucción proyectiva a una *afín*, tenemos que identificar el plano en el infinito, π_∞ . En una transformación afín, el plano en el infinito se mantiene invariante. Las coordenadas del plano en el infinito son $\pi_\infty = (0, 0, 0, 1)^T$, así que lo que tenemos que hacer es encontrar una transformación proyectiva que me transforme el plano en el infinito detectado al plano con las coordenadas anteriores, o lo que es lo mismo, encontrar un \mathbf{H} tal que $\mathbf{H}^{-T}\pi = (0, 0, 0, 1)^T$. Esta transformación viene dada por

$$\mathbf{H} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \pi^T & \end{pmatrix}.$$

Con una reconstrucción afín podemos realizar cálculos como el punto medio entre dos puntos, el centroide para un conjunto de estos u otros cálculos que tengan que ver con los invariantes para este tipo de transformaciones.

¿Cómo identificar el plano en el infinito? Para definir el plano en el infinito podemos utilizar tres puntos en el infinito. Si el desplazamiento entre las cámaras es una traslación pura, entonces los puntos en el infinito se corresponderán con las mismas coordenadas en las dos imágenes, luego bastaría con seleccionar tres coordenadas iguales en las dos imágenes, obtener los puntos \mathbf{X}_i en la escena y obtener a partir de ellos el plano en el infinito.

También se puede identificar el plano en el infinito a través de líneas paralelas en la escena. Si podemos reconocer en la imagen las líneas que son paralelas originalmente en la escena, entonces podemos calcular los puntos de fuga como la intersección de dichas líneas en la imagen. Si conseguimos tres conjuntos de líneas paralelas distintos, entonces podremos obtener tres puntos de fuga (puntos en el infinito) y, por tanto, obtener el plano en el infinito.

Otra forma de indentificar el plano en el infinito es a través de ratios de distancia en una línea. Si conocemos dos intervalos en una línea con un ratio de distancia conocido, entonces el punto en el infinito para esa línea se puede determinar.

Para refinar a una reconstrucción métrica, tenemos que identificar la cónica absoluta, Ω_∞ .

9 Cálculo de la matriz F

9.1 Ecuaciones básicas

La matriz fundamental se define a partir de la ecuación $\mathbf{x}'^T \mathbf{F} \mathbf{x} = 0$. Si tenemos un conjunto mínimo de puntos en correspondencia (al menos 7), entonces podemos calcular la matriz. Cada par de puntos, $\mathbf{x} = (x, y, 1)$ y $\mathbf{x}' = (x', y', 1)$ da lugar a una ecuación lineal. De esta forma, si tomamos un conjunto de puntos y representamos en un vector \mathbf{f} los elementos de la matriz \mathbf{F} , entonces tendremos que resolver un sistema de la forma $\mathbf{A} \mathbf{f} = 0$. Si los datos no son exactos o tenemos más correspondencias de las que necesitamos, se busca una solución por mínimos cuadrados. Esta solución es el vector singular que se corresponde con el valor singular de \mathbf{A} más pequeño, esto es, la última columna de \mathbf{V} en el SVD $\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{V}^T$.

La matriz \mathbf{F} tiene rango 2, con lo que hay que forzar que la solución obtenida por mínimos cuadrados se convierta en una matriz cuyo $\det \mathbf{F} = 0$. Para esto, lo que se suele hacer es buscar una matriz \mathbf{F}' que minimice $\|\mathbf{F} - \mathbf{F}'\|$ tal que $\det \mathbf{F}' = 0$. Esto se puede calcular de nuevo utilizando el SVD.

9.2 El algoritmo de 8 puntos normalizados

Algorithm 14 *Algoritmo de los 8 puntos*

Dados 8 ó más puntos en correspondencia $\{\mathbf{x}_i \longleftrightarrow \mathbf{x}'_i\}$ calcular la matriz \mathbf{F} tal que $\mathbf{x}'_i{}^T \mathbf{F} \mathbf{x}_i = 0$.

1. Normalización: Se transforman las coordenadas de imagen $\hat{\mathbf{x}}_i = \mathbf{T} \mathbf{x}_i$, $\hat{\mathbf{x}}'_i = \mathbf{T} \mathbf{x}'_i$
2. Encontrar la matriz fundamental:
 - (a) Solución lineal: Se calcula la matriz $\hat{\mathbf{F}}$ a partir del SVD de la matriz $\hat{\mathbf{A}}$ compuesta por las correspondencias de los puntos $\{\hat{\mathbf{x}}_i \longleftrightarrow \hat{\mathbf{x}}'_i\}$.
 - (b) Forzar restricción: Reemplazar $\hat{\mathbf{F}}$ por $\hat{\mathbf{F}}'$ tal que $\det \hat{\mathbf{F}}' = 0$ utilizando el SVD.
3. Denormalizar: La matriz fundamental original es $\mathbf{F} = \mathbf{T}'^T \hat{\mathbf{F}}' \mathbf{T}$ para los puntos en correspondencia originales.

9.3 Distancia geométrica

El algoritmo de los 8 puntos normalizados es simple y rápido, pero tiene el inconveniente que no es óptimo en el sentido de que no le da igual importancia a todas las entradas de \mathbf{F} . Otra forma de calcular la matriz fundamental es a través de métodos no lineales.

La estimación del Máximo Parecido (Maximum Likelihood) de la matriz fundamental depende de la asunción de un modelo de error. Podemos suponer que los puntos en correspondencia contienen ruido Gausiano. En este caso, la estimación ML es la que minimiza la distancia geométrica

$$\sum_i d(\mathbf{x}_i, \hat{\mathbf{x}}_i) + d(\mathbf{x}'_i, \hat{\mathbf{x}}'_i)$$

donde $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ son los puntos que se han medio y $\hat{\mathbf{x}}_i, \hat{\mathbf{x}}'_i$ son los que se suponen las verdaderas correspondencias que satisfacen $\hat{\mathbf{x}}'_i{}^T \mathbf{F} \hat{\mathbf{x}}_i = 0$. Para minimizar este funcional, podemos definir dos matrices $\mathbf{P} = [\mathbf{I}|\mathbf{0}]$ y $\mathbf{P}' = [\mathbf{M}|\mathbf{t}]$. Se obtienen los puntos 3D, \mathbf{X}_i , asociados a las correspondencias de forma que $\hat{\mathbf{x}}_i = \mathbf{P} \mathbf{X}_i$, $\hat{\mathbf{x}}'_i = \mathbf{P}' \mathbf{X}_i$. Luego se varía el valor de \mathbf{P}' y de \mathbf{X}_i para minimizar la expresión. La matriz fundamental se obtiene como $\mathbf{F} = [\mathbf{t}]_{\times} \mathbf{M}$. Para encontrar el mínimo se puede utilizar el algoritmo de Levenberg-Marquardt. Los métodos de minimización iterativos necesitan por lo general una aproximación inicial, con lo que podemos utilizar el resultado del algoritmo de 8 puntos como estimación inicial.

Algorithm 15 *Algoritmo de minimización del error de reproyección*

1. Se calcula un estimación lineal inicial de $\hat{\mathbf{F}}$ (algoritmo de los 8 puntos)
2. Se calculan las correspondencias estimadas $\{\hat{\mathbf{x}}_i, \hat{\mathbf{x}}'_i\}$:
 - (a) Se eligen las matrices de cámara $\mathbf{P} = [\mathbf{I}|\mathbf{0}]$ y $\mathbf{P}' = [[\mathbf{e}']_{\times} \hat{\mathbf{F}} | \mathbf{e}']$ donde \mathbf{e}' se obtiene de $\hat{\mathbf{F}}$
 - (b) A partir de las correspondencias $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$ y de $\hat{\mathbf{F}}$ se obtienen los puntos 3D $\{\mathbf{X}_i\}$
 - (c) Las correspondencias consistentes con $\hat{\mathbf{F}}$ se obtienen como $\hat{\mathbf{x}}_i = \mathbf{P} \mathbf{X}_i$, $\hat{\mathbf{x}}'_i = \mathbf{P}' \mathbf{X}_i$
3. Se minimiza la función coste

$$\sum_i d(\mathbf{x}_i, \hat{\mathbf{x}}_i) + d(\mathbf{x}'_i, \hat{\mathbf{x}}'_i)$$

sobre $\hat{\mathbf{F}}$ y \mathbf{X}_i utilizando un método como el de Levenberg-Marquardt.

Se necesita parametrizar la matriz fundamental para forzar que su rango sea igual a 2.

Una forma de parametrizarla es $\mathbf{F} = [\mathbf{t}]_{\times} \mathbf{M}$ ya que $[\mathbf{t}]_{\times}$ es singular y hace que \mathbf{F} también lo sea. Otra forma de parametrizarla es especificando dos columnas y representar la tercera como una combinación lineal de las dos primeras

$$\mathbf{F} = \begin{pmatrix} a & b & \alpha a + \beta b \\ c & d & \alpha c + \beta d \\ e & f & \alpha e + \beta f \end{pmatrix}$$

9.4 Cálculo automático de \mathbf{F}

Es interesante poder calcular la matriz fundamental únicamente a partir de dos imágenes. Para esto se utiliza un algoritmo de estimación robusta basado en el RANSAC (RANdom SAmple Consensus).

Algorithm 16 *Cálculo de la matriz fundamental entre dos imágenes*

1. *Puntos de interés: Se calculan puntos en las dos imágenes (método de Harris)*
2. *Correspondencias potenciales: Se encuentran posibles correspondencias entre las dos imágenes*
3. *RANSAC: Para N muestras se repite*
 - (a) *Se seleccionan aleatoriamente 7 correspondencias y se calcula la matriz fundamental*
 - (b) *Se calcula la distancia d_{\perp} para cada correspondencia potencial*
 - (c) *Se calculan el número de inliers consistentes con \mathbf{F} tal que $d_{\perp} < t$*
 - (d) *Si existen tres soluciones para \mathbf{F} se selecciona aquella con mayor número de inliers*

Se selecciona la \mathbf{F} con mayor número de inliers

4. *Estimación no-lineal: Se recalcula \mathbf{F} a partir de todas las correspondencias potenciales clasificadas como inliers, utilizando Levenberg-Marquardt*
5. *Correspondencia guiada: Se determinan más puntos en correspondencia utilizando ahora la matriz \mathbf{F}*

10 Cálculo de la estructura

Si tenemos las matrices de las cámaras y un conjunto de puntos en correspondencia, entonces es posible obtener los puntos 3D originales. Uno de los problemas con los que nos encontramos es que los rayos que unen las correspondencias con los centros de las cámaras generalmente no intersectan en un punto, sino que determinan rectas convergentes que no se llegan a tocar. Esto es así debido a que la precisión de las imágenes es finita y a que las correspondencias no son exactas.

10.1 Métodos lineales de triangulación

El método más simple para reconstruir un punto es el de la triangulación. Este consiste en determinar los dos rayos que unen cada correspondencia con el foco de la cámara y después encontrar el punto cuya distancia a dichas rectas sea mínima.

El método lineal es análogo al método DLT para el cálculo de las homografías. Las proyecciones $\mathbf{x} = \mathbf{P}\mathbf{X}$ y $\mathbf{x}' = \mathbf{P}'\mathbf{X}$ se pueden combinar de forma que obtengamos un sistema de la forma $\mathbf{A}\mathbf{X} = \mathbf{0}$.

Para eliminar el factor de escala utilizamos el producto escalar de la forma $\mathbf{x} \times \mathbf{P}\mathbf{X} = \mathbf{0}$ y obtenemos lo siguiente

$$\begin{aligned} x(\mathbf{p}^{3T}\mathbf{X}) - (\mathbf{p}^{1T}\mathbf{X}) &= 0 \\ y(\mathbf{p}^{3T}\mathbf{X}) - (\mathbf{p}^{2T}\mathbf{X}) &= 0 \\ x(\mathbf{p}^{2T}\mathbf{X}) - y(\mathbf{p}^{1T}\mathbf{X}) &= 0. \end{aligned}$$

Estas ecuaciones son lineales en los componentes de \mathbf{X} y una de ellas depende linealmente de las otras dos. Si introducimos la información del otro punto, entonces podemos obtener la matriz

$$\mathbf{A} = \begin{pmatrix} x\mathbf{p}^{3T} - \mathbf{p}^{1T} \\ y\mathbf{p}^{3T} - \mathbf{p}^{2T} \\ x'\mathbf{p}'^{3T} - \mathbf{p}'^{1T} \\ y'\mathbf{p}'^{3T} - \mathbf{p}'^{2T} \end{pmatrix}$$

Para resolver el sistema podemos utilizar el DLT.

10.2 Función de coste de error geométrico

Este método consiste en encontrar las verdaderas correspondencias $\hat{\mathbf{x}}$ y $\hat{\mathbf{x}'}$ cumpliendo la restricción epipolar de la forma

$$\begin{aligned} \mathcal{C}(\mathbf{x}, \mathbf{x}') &= d(\mathbf{x}, \hat{\mathbf{x}})^2 + d(\mathbf{x}', \hat{\mathbf{x}'})^2 \text{ restringido a} \\ \hat{\mathbf{x}'^T \mathbf{F} \hat{\mathbf{x}} &= 0. \end{aligned}$$

Esto es equivalente a minimizar el error de reproyección de los puntos 3D estimados. Si suponemos un error Gaussiano, entonces los puntos $\hat{\mathbf{x}}$ y $\hat{\mathbf{x}'}$ son las Estimaciones de Máximo Parecido (MLE) de las correspondencias entre las imágenes.

10.3 Solución óptima

Este método encuentra el mínimo global de la función de coste anterior utilizando un algoritmo no lineal.

$$\mathbf{F}(1, 0, f)^T = (1, 0, f')\mathbf{F} = \mathbf{0}$$

$$\mathbf{F} = \begin{pmatrix} ff'd & -f'c & -f'd \\ -fb & a & b \\ -fd & c & d \end{pmatrix}$$

Suponemos que la línea epipolar pasa por el punto $(0, t, 1)^T$ y por el epipolo $(1, 0, f)^T$. Esta línea se representa como $\mathbf{l}(t) = (0, t, 1) \times (1, 0, f) = (tf, 1, -t)$. La distancia del origen a esta recta es

$$d^2(\mathbf{0}, \mathbf{l}(t)) = \frac{t^2}{1 + (tf)^2}$$

La línea epipolar en la otra imagen es $\mathbf{l}'(t) = \mathbf{F}(0, t, 1)^T = (-f'(ct + d), at + b, ct + d)^T$. La distancia del origen a esta línea es

$$d^2(\mathbf{0}, \mathbf{l}'(t)) = \frac{(ct + d)^2}{(at + b)^2 + f'^2(ct + d)^2}$$

$$s(t) = \frac{t^2}{1 + f^2 t^2} + \frac{(ct + d)^2}{(at + b)^2 + f'^2 (ct + d)^2}$$

$$s'(t) = \frac{2t}{(1 + f^2 t^2)^2} - \frac{2(ad - bc)(at + b)(ct + d)}{\left((at + b)^2 + f'^2 (ct + d)^2\right)^2}$$

$$g(t) = t \left((at + b)^2 + f'^2 (ct + d)^2 \right)^2 - (1 + f^2 t^2)^2 (ad - bc)(at + b)(ct + d)$$

$$= 0$$

Algorithm 17 *Reconstrucción óptima*

1. Se obtienen las matrices de traslación

$$\mathbf{T} = \begin{pmatrix} 1 & -x \\ & 1 & -y \\ & & 1 \end{pmatrix} \text{ y } \mathbf{T}' = \begin{pmatrix} 1 & x \\ & 1 & y \\ & & 1 \end{pmatrix}$$

para trasladar los puntos al origen

2. Se reemplaza la matriz \mathbf{F} por $\mathbf{T}'^{-T} \mathbf{F} \mathbf{T}^{-1}$
3. Se calculan los epipolos como $\mathbf{F} \mathbf{e} = 0$ y $\mathbf{e}'^T \mathbf{F} = 0$. Se normalizan de forma que $e_1^2 + e_2^2 = 1$
4. Se crean las matrices de rotación

$$\mathbf{R} = \begin{pmatrix} e_1 & e_2 & \\ -e_2 & e_1 & \\ & & 1 \end{pmatrix} \text{ y } \mathbf{R}' = \begin{pmatrix} e'_1 & e'_2 & \\ -e'_2 & e'_1 & \\ & & 1 \end{pmatrix}$$

de forma que $\mathbf{R} \mathbf{e} = (1, 0, e_3)^T$ y $\mathbf{R}' \mathbf{e}' = (1, 0, e'_3)^T$

5. Se reemplaza la matriz \mathbf{F} por $\mathbf{R}' \mathbf{F} \mathbf{R}^T$
6. Se igualan $f = e_3$, $f' = e'_3$, $a = \mathbf{F}_{22}$, $b = \mathbf{F}_{23}$, $c = \mathbf{F}_{32}$ y $d = \mathbf{F}_{33}$
7. Se forma el polinomio $g(t)$ y se buscan las 6 raíces
8. Se evalúa la función de coste $s(t)$ en las raíces y para el valor de $t = \infty$. Se selecciona el valor de t que dé el menor valor
9. Se obtienen las dos líneas $\mathbf{l} = (tf, 1, -t)$ y \mathbf{l}' para el valor del mínimo y se encuentran los valores \hat{x} y \hat{x}' sobre dichas líneas que estén más próximos al origen. Para una línea cualquiera (λ, μ, ν) el punto más próximo al origen viene dado por $(-\lambda\nu, -\mu\nu, \lambda^2 + \mu^2)$
10. Se deshacen las transformaciones anteriores sustituyendo \hat{x} por $\mathbf{T}^{-1} \mathbf{R}^T \hat{x}$ y \hat{x}' por $\mathbf{T}'^{-1} \mathbf{R}'^T \hat{x}'$
11. El punto 3D $\hat{\mathbf{X}}$ se calcula a través del método de triangulación homogéneo

11 Geometría epipolar afín

Cuando trabajamos con cámaras afines, la geometría epipolar se simplifica considerablemente, permitiendo realizar ciertos algoritmos de forma más sencilla. Las cámaras afines tienen su centro en el infinito y la proyección se realiza perpendicularmente al plano de la imagen.

Para este tipo de cámaras, las líneas epipolares son paralelas, así como los planos epipolares. Los epipolos también se encuentran en el infinito.

11.1 La matriz \mathbf{F} afín

La matriz fundamental se obtiene a partir de las matrices de las cámaras que tienen la tercera fila igual a $(0, 0, 0, 1)$. En particular, la matriz \mathbf{F}_A tiene la forma

$$\mathbf{F}_A = \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{pmatrix}.$$

Las correspondencias entre las imágenes viene dado por una transformación afín de forma que $\mathbf{x}' = \mathbf{H}_A \mathbf{x}$.

La línea epipolar se obtiene a través de \mathbf{x}' y el epipolo \mathbf{e}' , $\mathbf{l}' = \mathbf{e}' \times \mathbf{H}_A \mathbf{x} = \mathbf{F}_A \mathbf{x}$, así que

$$\begin{aligned} \mathbf{F}_A &= [\mathbf{e}']_{\times} \mathbf{H}_A = \begin{pmatrix} 0 & 0 & * \\ 0 & 0 & * \\ * & * & 0 \end{pmatrix} \begin{pmatrix} * & * & * \\ * & * & * \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & * \\ 0 & 0 & * \\ * & * & * \end{pmatrix}. \end{aligned}$$

También podemos deducir la matriz algebraicamente sabiendo que $\mathbf{l}' = [\mathbf{e}']_{\times} \mathbf{P}' \mathbf{P}^+$.

La matriz fundamental afín tiene, por tanto, 4 grados de libertad. El epipolo de la primera imagen se puede obtener como $\mathbf{F}_A \mathbf{e} = 0$, por lo que $\mathbf{e} = (-d, c, 0)$ y $\mathbf{e}' = (-b, a, 0)$.

Las líneas epipolares son paralelas. Esto se puede ver como $\mathbf{l}' = \mathbf{F}_A \mathbf{x} = (a, b, cx + dy + e)^T$. Aquí todas las líneas epipolares tienen la misma dirección dada por (a, b) .

Las correspondencias entre puntos se establecen como $\mathbf{x}'^T \mathbf{F}_A \mathbf{x} = 0$, $ax' + by' + cx + dy + e = 0$.

Las matrices de las cámaras canónicas también tienen una forma bastante simple:

$$\mathbf{P}_A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \mathbf{P}'_A = \left(\begin{array}{ccc|c} \mathbf{M}_{2 \times 3} & & & \mathbf{t} \\ 0 & 0 & 0 & 1 \end{array} \right)$$

con $a = m_{23}$, $b = -m_{13}$, $c = m_{13}m_{21} - m_{11}m_{23}$, $d = m_{13}m_{22} - m_{12}m_{23}$ y $e = m_{13}t_2 - m_{23}t_1$.

11.2 Cálculo de \mathbf{F}_A a partir de puntos en correspondencia

Algorithm 18 *Cálculo del MLE de \mathbf{F}_A a partir de un conjunto de $n \geq 4$ correspondencias. Representamos cada correspondencia como $\mathbf{X}_i = (x'_i, y'_i, x_i, y_i)$*

1. Calcular el centroide $\bar{\mathbf{X}} = \frac{1}{N} \sum_i \mathbf{X}_i$ y centrar los puntos $\Delta \mathbf{X}_i = \mathbf{X}_i - \bar{\mathbf{X}}$
2. Calcular la matriz \mathbf{A} de $n \times 4$ con $\Delta \mathbf{X}_i$ como filas
3. $\mathbf{N} = (a, b, c, d)^T$ es el vector singular correspondiente al valor singular más pequeño de \mathbf{A} y $e = -\mathbf{N}^T \bar{\mathbf{X}}$.

References

- [1] Richard Hartley y Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000
- [2] Olivier Faugeras, *Three-Dimensional Computer Vision, a Geometric Viewpoint*, MIT Press, 1993
- [3] Olivier Faugeras, Quang-Tuan Luong and Theo Papadopoulos, *The Geometry of Multiple Images*, MIT Press, 2001