La matriz fundamental y la matriz esencial. Concepto y aplicaciones

David Santo Orcero

irbis@orcero.org

http://www.orcero.org/irbis

Mayo de 2002

Resumen

En este documento vamos a analizar qué es una matriz fundamental, qué es una matriz esencial, sus formulaciones y sus deducciones. También estudiaremos cómo se calculan en la práctica, y algunos casos de ejemplo de aplicaciones. Incorporamos en este documento los conceptos básicos necesarios para entender las demostraciones, y hacemos énfasis en la comprensión intuitiva y en el rigor de los temas comentados.

Índice

1. Base matemática

Antes de comenzar a estudiar las matrices fundamentales y esenciales, vamos a repasar algunos de los conceptos matemáticos que necesitaremos para seguir las demostraciones y los algoritmos.

1.1. La matriz pseudoinversa

Definimos como pseudoinversa por la derecha de una matriz A, que denotamos como A^+, a una matriz que multiplicada por la derecha por A da la matriz diagonal con todos los elementos diagonales iguales a 1 -matriz I-; es decir:

$$I = AA^+ \tag{1}$$

esta definición de matriz pseudoinversa permite que la matriz A no sea cuadrada. Esto es fundamental en nuestro problema, ya que las matrices de proyección no son cuadradas. Si tenemos las matrices de proyección P_1 y P_2, estas tendrán como dimensión 4x3. Esto significa que nuestra matriz pseudoinversa por la derecha debe ser una matriz con dimensiones 3x4. Calculamos la pseudoinversa de A con la ecuación:

$$A^+ = A^T[A^TA]^{-1} \tag{2}$$

como la matriz $[A^TA]$ es cuadrada, suponiendo que sea una matriz regular, es una matriz que puede ser invertida, y su inversa es única; por lo que el cálculo de esta expresión no debe tener ningún problema adicional.

1.2. La matriz antisimétrica

Definimos matriz antisimétrica por la izquierda $[A]^\times$ de una matriz A a aquella matriz tal que:

$$A \times B = [A]^\times B \tag{3}$$

En las demostraciones que necesitaremos para llegar a una formulación simple de la matriz epipolar, necesitaremos conocer la matriz antisimétrica por la izquierda de una matriz de 3x1 posiciones. Si tenemos la matriz:

$$A = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \tag{4}$$

Su antisimétrica $[A]^\times$ será:

$$[A]^{\times} = \begin{pmatrix} 0 & -a_2 & a_1 \\ a_2 & 0 & -a_0 \\ -a_1 & a_0 & 0 \end{pmatrix} \tag{5}$$

1.3. La transformada SVD

La transformada SVD -Singular Value Decomposition- es una herramienta común para resolver sistemas de ecuaciones, y para diagonalizar matrices cuando el sistema es singular, o numéricamente muy próximo del singular -el error acumulado del método numérico por la precisión finita de los cálculos de una máquina es suficiente muchas veces para convertir sistemas no singulares en sistemas singulares-. En estos casos, podemos hacer una resolución por mínimos cuadrados del sistema, o plantearnos resolución del sistema por SVD.

Cualquier matriz A, sea singular o no, puede ser descompuesta en un producto de tres matrices, de la forma:

$$A = UDV^T \tag{6}$$

donde U y V son matrices ortogonales, por lo que sus inversas son iguales a sus transpuestas; y D es una matriz diagonal.

La aplicación más común de la SVD es el cálculo rápido de la inversa de una matriz para resolver un sistema de ecuaciones. La inversa de la matriz A, conocida su descomposición SVD, es:

$$A^{-1} = VD^{-1}U^T \tag{7}$$

El cómputo de la inversa de la matriz D es fácil de calcular; aplicando la propiedad de las matrices diagonales por la que la inversa de una matriz diagonal es la inversa de los elementos de dicha matriz. Esto funciona si la matriz no es singular, lo que significa en la práctica que todos los elementos de la matriz D son distintos de 0; lo que no es el caso ni en la matriz fundamental ni en la matriz esencial, que son matrices singulares -y que en particular tendrán un elemento de D igual a 0.

SVD es interesante para nosotros porque define un espacio ortogonal en U cuyo número de vectores directores coincide con el rango de la matriz A, y este espacio define la parte no-singular de la matriz, y un espacio ortogonal en V cuyo número de vectores directores coincide con la dimensión de la matriz A menos su rango -es decir, el número de ceros en la diagonal de la matriz D-, y este espacio define la parte singular de la matriz.

1.4. Propiedades de la SVD

Las propiedades más importantes de la SVD son:

- Sea $A = UDV^T$ una descomposición SVD, σ_1 el mayor elemento diagonal de D y σ_n el menor elemento diagonal de D. Llamamos *grado de singularidad* al cociente $\frac{\sigma_1}{\sigma_n}$; y si la inversa del grado de singularidad es del orden de magnitud que la precisión de la máquina, denominamos a la matriz *mal condicionada*, y la podemos considerar singular.

- Sea $A = UDV^T$ una descomposición SVD, y A una matriz cuadrada no singular. La inversa de A será:

$$A^{-1} = VD^{-1}U^T \tag{8}$$

- Sea $A = UDV^T$ una descomposición SVD, y A una matriz cuadrada singular. La pseudoinversa de A será:

$$A^+ = VD_0^{-1}U^T \tag{9}$$

  donde D_0 es una matriz en la que sustituimos los elementos de la diagonal que valgan cero por ∞; lo que es lo mismo que decir que los elementos de D_0^{-1} que valgan ∞ los sustituiremos por cero.

- Los autovalores no nulos de A^TA y de AA^T son los cuadrados de los valores singulares no nulos de A.

- Sea $A = UDV^T$ una descomposición SVD, las columnas de U son los autovectores de AA^T y las columnas de V son los autovectores de A^TA.

- Sea $A = UDV^T$ una descomposición SVD, todo elemento de la diagonal σ_k cumple que:

$$Au_k = \sigma_k v_k \tag{10}$$

  y que:

$$A^T v_k = \sigma_k u_k \tag{11}$$

  siendo:

  - u_k la columna de U correspondiente a σ_k
  - v_k la columna de V correspondiente a σ_k

2. Análisis geométrico de una imagen

2.1. El modelo pin-hole

El modelo de cámara pin-hole es uno de los más sencillos que podemos plantearnos, pero que es de mucho interés, ya que modela razonablemente bien una cámara común con una formulación matemática muy simple.

El modelo de cámara pin-hole supone que para todo punto que impacta en la película, el rayo de luz que sale rebotado de un cuerpo y llega a la cámara, atraviesa único punto, independientemente del punto de origen y del punto de impacto en la película. Esto hace el modelo matemático de la cámara muy simple, ya que de una aplicación directa del teorema de Thales de Mileto, podemos sacar las ecuaciones:

$$\begin{cases} Zx & = -fX \\ Zy & = -fY \end{cases} \tag{12}$$

suponiendo que:

- Las cordenadas del punto en el espacio que se proyecta serán $(X, Y, Z)^T$, donde el eje de la coordenada Z es la línea que pasa por el centro óptico y es perpendicular al plano de proyección, y las coordenadas $(0, 0, 0)$ están situadas en el centro óptico.
- Las coordenadas del punto proyectado serán $(x, y)^T$.
- f es la *distancia focal* -es decir, la distancia que separa el centro óptico del plano de proyección-.

Si queremos operar con álgebra matricial -lo que es más cómodo para trabajar con ordenadores- podemos reformular la ecuación anterior en coordenadas homogeneas como:

$$Z\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -fX \\ -fY \\ Z \end{bmatrix} = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{13}$$

sabiendo que hemos definido el punto proyectado m como:

$$\begin{bmatrix} -fX \\ -fY \\ Z \end{bmatrix} \tag{14}$$

el punto que se proyecta M como:

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{15}$$

y la matriz de proyección P como:

$$\begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{16}$$

la ecuación 13 queda como:

$$\lambda m = PM \tag{17}$$

Donde λ es una argucia matemática, denominada factor de escalado, y que coincide con la coordenada Z del punto M.

Aunque a simple vista no lo parezca, hemos ganado bastante con esta reformulación: a pesar de la nomenclatura matricial, el hecho de operar con coordenadas homogeneas supone que tratamos siempre con un sistema lineal de ecuaciones, por lo que dispondremos de gran cantidad de técnicas numéricas sencillas, rápidas y precisas para resolver el problema en un ordenador.

La formula matemática de la matriz de proyección P puede ser más complicada, en caso que no empleemos el modelo pinhole al modelar la imagen. También puede involucrar más parámetros que apenas la distancia focal f. Al proceso de cálculo de estos parámetros, que nos permiten deducir la matriz de proyección, lo denominamos *calibración de la cámara*.

El estudio detallado del cálculo de P en casos más complicados es algo que queda fuera del objetivo de nuestro trabajo, y recomendamos la lectura de la bibliografía para tener más información sobre cálculo de matrices de proyección P para modelos más complejos de cámara.

2.2. Cálculo del centro óptico

A veces no tendremos las coordenadas espaciales del centro óptico, pero más adelante las necesitaremos. Por ello, necesitamos saber como calcular el centro óptico C a partir de la matriz de proyección P.

Para calcularlo, emplearemos una propidedad curiosa: el centro óptico es el único punto del espacio que no puede ser proyectado en el plano. La forma de hacer esto matemáticamente es comprobando que la última coordenada de su proyección sea 0, por lo que nos saldrá el punto en la recta del infinito.

Podemos poner las otras dos coordenadas a 0 para facilitar la formulación matemática, quedando esta como:

$$PC = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{18}$$

2.3. Modelo de cámara CCD

La cámara CCD –charge-coupled-device– es un tipo de cámara muy común en aplicaciones informáticas. Internamente está formada por una matriz de seminconductores fotosensibles.

Entre sus características, destacamos que la imagen ya está espacialmente discretizada, y que son cámaras que tienen una resolución muy buena, operan en condiciones de poca luz, y presentan pocas aberraciones geométricas, en relación con las otras tecnologías existentes. Para nuestro caso actual nos interesan porque los casos de estudio los hemos realizado con una Canon Ixus digital, con un CCD.

Estas cámaras tienen en su matriz de perspectiva cuatro aspectos importantes que debemos modelar:

- *Translación*: El punto $(0, 0)$ de nuestra imagen capturada puede no coincidir con el centro de coordenadas de nuestro sistema. Llamaremos a los parámetros de translación x_0 y y_0, y corresponde con la posición del punto $(0, 0)$ de la imagen de la cámara respecto al sistema global.

- *Escalado*: Un punto de una imagen no tiene por que coincidir necesáriamente con la unidad en la que estamos midiendo. El factor de escalado determina la proporción entre la unidad del mundo real que empleamos y el punto en la cámara, y lo denotaremos por s.

- *Rotación*: Una imagen puede estar rotada respecto al sistema de coordenadas. Esto supone tres parámetros, α, β, γ que corresponden, respectivamente, al ángulo del eje de la cámara con los ejes X, Y y Z del sistema. Es común poner arbitrariamente el eje Z perpendicular a la imagen de la cámara, ya que simplifica los cálculos -el angulo de la cámara respecto al sistema de coordenadas es igual al ángulo de giro en la imagen-.

- *Deformación*: los CCDs pueden provocar una deformación en la imagen. Esta deformación suele ser lineal, y corresponde a que los pixeles del CCD que captura la imagen pueden ser rectangulares -en lugar de cuadrados-, por lo que tendremos una deformación homogénea y lineal de ensanchamiento -o estiramiento- de la imagen. El factor de deformación, llamado d, es la razón entre la altura y la anchura de un pixel del CCD de la cámara, y puede ser calibrado con facilidad haciendo la operación a mano sobre una foto de un cuadrado. Observamos que podemos tener una variable de escalado y otra de deformación, o una variable de escalado horizontal y otra variable de escalado vertical, y no sería necesario el factor de demostración.

La ecuación matricial del modelo de cámara de una cámara CCD sin deformación es:

$$\begin{pmatrix} m_{k_x} \\ m_{k_y} \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha & s & x_0 \\ 0 & \beta & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} m_x \\ m_y \\ 1 \end{pmatrix} \tag{19}$$

es decir, las ecuaciones en formato matricial son:

$$m_k = Km \tag{20}$$

Donde K es la denominada *matriz de calibración*.

$$m = KPM \tag{21}$$

desarrollando las matrices, obtenemos:

$$\begin{pmatrix} m_x \\ m_y \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha & s & x_0 \\ 0 & \beta & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} M_x \\ M_y \\ M_z \\ 1 \end{pmatrix} \tag{22}$$

si hacemos las sustituciones:

$$\begin{cases} \alpha_f & = \alpha f \\ \beta_f & = \beta f \\ s_f & = s f \end{cases} \tag{23}$$

tenemos:

$$\begin{pmatrix} m_x \\ m_y \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_f & s_f & x_0 \\ 0 & \beta_f & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} M_x \\ M_y \\ M_z \\ 1 \end{pmatrix} \tag{24}$$

es decir, hemos prescindido de la longitud focal.

En la práctica, significa que podemos trabajar con una longitud focal 1, y compensarlo al realizar la calibración de la cámara.

3. Análisis geométrico de dos imagenes

El hecho de disponer de dos imágenes del mismo entorno nos permite descubrir nuevas informaciones sobre dicho entorno.

Supongamos que tenemos un punto M del espacio, que se proyecta en el punto m_1 de la primera imagen, y el punto m_2 de la segunda imagen. Por aplicación directa de la ecuación 17 a este escenario, tenemos el sistema de ecuaciones matriciales:

$$\begin{cases} \lambda_1 m_1 = & P_1 M \\ \lambda_2 m_2 = & P_2 M \end{cases} \tag{25}$$

Donde:

- P_1 es la matriz de proyección de la primera imagen.
- P_2 es la matriz de proyección de la segunda imagen.
- M son las coordenadas homogéneas del punto en el espacio.

- λ_1 es el factor de escalado para la primera imagen.
- λ_2 es el factor de escalado para la segunda imagen.
- m_1 son las coordenadas homogéneas en el plano de proyección de la primera imagen de la primera proyección del punto.
- m_2 son las coordenadas homogéneas en el plano de proyección de la segunda imagen de la segunda proyección del punto.

3.1. Visión estéreo y geometría epipolar

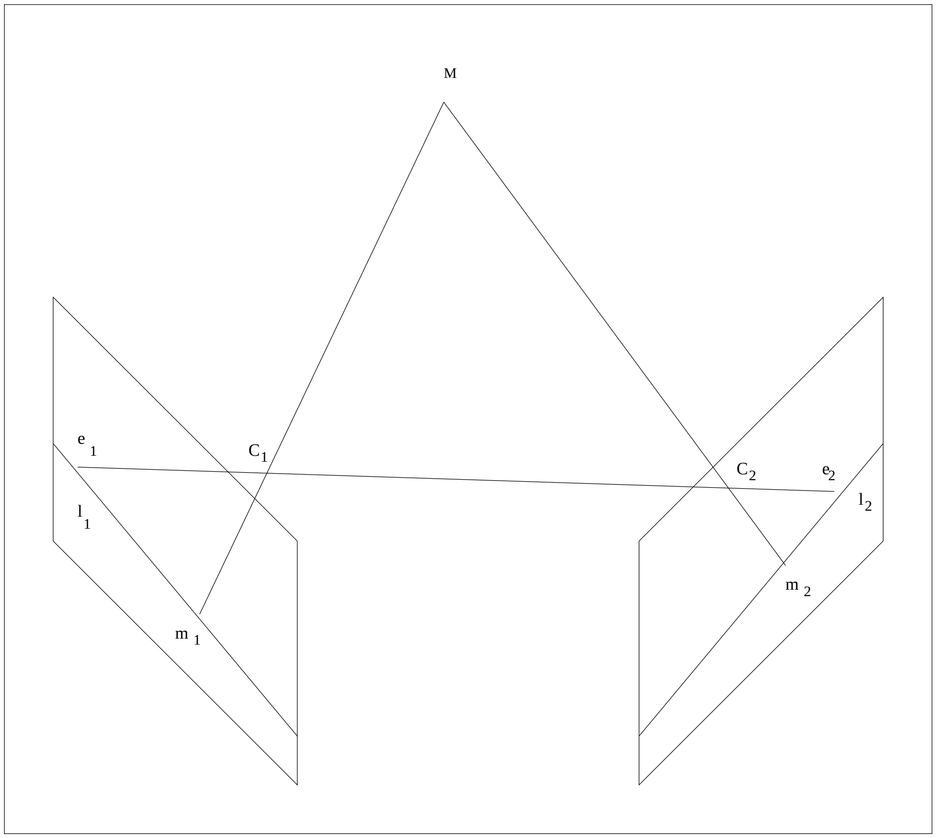
Definimos visión estéreo como aquella en la que empleamos más de una imagen para obtener una idea de tridimensionalidad. Según el número de imágenes que empleemos, hablaremos de visión *bifocal* -dos imágenes-, *trifocal* -tres imágenes-, *quadrifocal* -cuatro imágenes- o *N-focal* -N imágenes-.

Para nuestro caso, estudiaremos los conceptos de un sistema bifocal; aunque lo estudiado es extrapolable a cualquier sistema N-focal.

El primer concepto que necesitaremos es el de *geometría epipolar*. Denominamos *geometría epipolar* como a la geometría generada por dos vistas; y se basa en dos conceptos fundamentales, quue son la línea epipolar y el epipolo.

Dado un sistema bifocal, definimos como epipolo e_i de una de las dos imágenes del sistema a la proyección del centro óptico de la otra imagen. Un sistema bifocal tiene exactamente dos epipolos, uno por cada imágen. Los dos epipolos pueden tener las mismas coordenadas, o coordenadas distintas.

Definimos como línea epipolar l_i para uno de los dos planos de proyección a la linea que pasa por el epipolo e_i contenido en el plano de proyección y la projección m_i. Observamos que la linea epipolar $e_i m_i$ es también la proyección de la línea que pasa por el centro óptico C_i y la proyección m_i sobre el plano de proyección. Tenemos exactamente dos lineas epipolares, una por cada imagen. Las dos lineas epipolares pueden tener las mismas coordenadas, o coordenadas distintas.



Definimos como plano epipolar al plano definido por los centros ópticos de las dos imágenes, y el punto M del espacio tridimensional. Tanto las dos proyecciones m_1 y m_2 del punto M como sus epipolos e_1 y e_2 están contenidos en el plano epipolar. si M está fuera de los dos planos de proyección y los planos de proyección no son paralelos, existe un único plano epipolar.

4. La matriz fundamental

4.1. Deducción de la matriz fundamental

Para deducir la formulación de la matriz fundamental tomaremos las ecuaciones 25, que son:

$$\begin{cases} \lambda_1 m_1 = & P_1 M \\ \lambda_2 m_2 = & P_2 M \end{cases} \tag{26}$$

El primer paso que tomaremos será buscar la formulación matemática de la linea epipolar para la segunda proyección P_2, que será l_2; nos interesa que sea función de m_1, P_1 y P_2.Podríamos usar como punto M si lo conocieramos, pero partimos de que M es desconocido.

Para encontrar el primer punto usaremos un punto arbitrario M_a, cuya formulación es:

$$M_a = P_1^+ m_1 \tag{27}$$

Es decir, nuesto punto será el producto de la pseudoinversa de la primera matriz de proyección por la primera proyección del punto. Siendo P_1 de 4x3 elementos, P_1^+ será forzosamente de 3x4 elementos según la definición de pseudoinversa por la derecha estudiada en la primera sección de este informe técnico. Dado que la matriz m_1 es de 3x1 elementos, la matriz resultante M_a es de 4x1 elementos, lo que significa que M_a formula matemáticamente un punto.

Este punto cuya ecuación hemos definido tiene una propiedad importante. Si lo aplicamos a la primera de las ecuaciones 25, tendremos:

$$\lambda_1 m_{1_{(M_a)}} = P_1 M_a \tag{28}$$

Sustituyendo el valor de M_a, obtenemos:

$$\lambda_1 m_{1_{(M_a)}} = P_1 P_1^+ m_1 \tag{29}$$

Aplicando la definición de pseudoinversa, tenemos que $P_1 P_1^+ = I$, por lo que tenemos que:

$$\lambda_1 m_{1_{(M_a)}} = I m_1 \tag{30}$$

Esto quiere decir que la proyección m_a de este punto M_a es igual a m_1, por lo que el punto está en la línea entre el centro óptico C_1 y la proyección m_1 en P_1 del punto M. Ya tenemos uno de los dos puntos que estamos buscando: M_a. Por lo tanto, hacemos:

$$m_{2_{(M_a)}} = P_2 M_a \tag{31}$$

Aplicando la definición de M_a de 27:

$$m_{2_{(M_a)}} = P_2 P_1^+ m_1 \tag{32}$$

Para el segundo punto, vamos a aplicar a la segunda ecuacion 25, el centro óptico C_1 de la primera imagen, con objeto de obtener el segundo epipolo. tendremos:

$$e_2 = m_{2_{(C_1)}} = P_2 C_1 \tag{33}$$

Conocidos estos dos puntos, el cálculo de la línea epipolar es trivial:

$$l_2 = e_2 \times m_{2_{(M_a)}} \tag{34}$$

Sustituyendo el cálculo de e_2 de 33 y de $m_{2_{(M_a)}}$ de 32, tenemos:

$$l_2 = P_2 C_1 \times P_2 P_1^{+} m_1 \tag{35}$$

Podemos aún simplificar más la formulación de la línea epipolar l_2 de la segunda proyección P_2, sustituyendo la matriz resultado del producto de $P_2 C_1$ por su antisimétrica; y así tenemos que la nueva formulación de la linea epipolar l_2 es:

$$l_2 = [P_2 C_1]^{\times} P_2 P_1^{+} m_1 \tag{36}$$

Definamos la matriz $\mathfrak{F}$ como:

$$\mathfrak{F} = [P_2 C_1]^{\times} P_2 P_1^{+} \tag{37}$$

La línea epipolar l_2 pasa a ser:

$$l_2 = \mathfrak{F} m_1 \tag{38}$$

Y la llamaremos *matriz fundamental*. La matriz fundamental depende exclusívamente de la geometría bifocal, y es independiente de los puntos del sistema y de sus proyecciones -es decir, no depende de M, ni de m_1, ni de m_2-.

La forma de usar esta matriz es para relacionar las dos proyecciones m_1 y m_2 de un punto en el espacio. Para deducir la expresión que relaciona las dos proyecciones,, analizaremos la linea epipolar l_2 de nuevo. Como m_2 pertenece a la línea epipolar l_2, tendremos que:

$$m_2^T l_2 = 0 \tag{39}$$

pero, aplicando la nueva definición de línea epipolar de 38, tenemos que:

$$m_2^T \mathfrak{F} m_1 = 0 \tag{40}$$

Por lo que podremos usar la matriz fundamental para relacionar las dos proyecciones de un punto en el espacio.

4.2. Propiedades de la matriz fundamental

La matriz fundamental $\mathfrak{F}$ tiene algunas propiedades importantes, que necesitaremos para calcularla y usarla. No vamos a demostrar las propiedades matemáticamente -se puede encontrar en cualquier libro donde describa la matriz fundamental-, pero si vamos a especificar las propiedades más importantes, que son:

- La matriz fundamental $\mathfrak{F}$ es homogénea. En la práctica, significa que $\forall k$ si $k \neq 0$ la matriz $k\mathfrak{F}$ sigue siendo matriz fundamental.
- El determinante de la matriz fundamental $\mathfrak{F}$ es cero.
- La matriz fundamental $\mathfrak{F}$ sólo tiene siete grados de libertad, es decir, de los nueve componentes de la matriz, dos pueden ser calculados a partir de los otros siete.

- El rango de la matriz fundamental $\mathfrak{F}$ es 2.

Además de estas propiedades algebraicas, que son útiles para calcular la matriz fundamental, tenemos un conjunto de propiedades geométricas, que usaremos para operar con la matriz fundamental. Estas propiedades son:

- Dada una geometria bifocal, la matriz fundamental $\mathfrak{F}$ es función exclusiva de dicha geometría. Esto significa que es independiente de todos los puntos M de la escena y de las proyecciones m_1 y m_2 de los puntos de la escena.

- La matriz fundamental $\mathfrak{F}$ define las lineas epipolares. Podemos calcular las lineas epipolares a partir de la matriz fundamental con las ecuaciones:

$$l_2 = \mathfrak{F} m_1 \tag{41}$$

$$l_1 = \mathfrak{F}^T m_2 \tag{42}$$

- La matriz fundamental $\mathfrak{F}$ define los epipolos. Podemos calcular los epipolos a partir de la matiz fundamental con las ecuaciones:

$$\mathfrak{F} e_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{43}$$

$$\mathfrak{F}^T e_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{44}$$

- La matriz fundamental $\mathfrak{F}$ redefine la restricción epipolar -es decir, la relación entre las dos proyecciones- como:

$$m_2^T \mathfrak{F} m_1 = 0 \tag{45}$$

5. Cálculo de la matriz fundamental

5.1. Introducción conceptual al cálculo de la matriz fundamental

Existen dos enfoques posibles al cálculo de la matriz fundamental. El primer enfoque es emplear la fórmula 37:

$$\mathfrak{F} = [P_2 C_1]^{\times} P_2 P_1^{+} \tag{46}$$

Este planteamiento es rápido y muy eficiente, involucrando apenas unas pocas operaciones algebraicas -principalmente, sumas y productos-. Sin embargo, necesita el conocimiento de las matrices de proyección P_1 y P_2, y del centro óptico C_1. El centro óptico podemos calcularlo directamente a partir de P_1 como ya hemos visto; sin embargo, P_1 y P_2 son matrices complicadas de descubrir, y esto será posible apenas en entornos fuertemente controlados.

El caso más común es que no conozcamos las matrices P_1 y P_2, por lo que este planteamiento no nos vale. Sin embargo, conocidas las correspondencias entre n puntos de la primera proyección m_{1_i} con n puntos de la segunda proyección m_{2_i}, podemos usar la ecuación:

$$m_{2_i}^T \mathfrak{F} m_{1_i} = 0 \tag{47}$$

para todos los pares de puntos de proyeccion m_{1_i}, m_{2_i}.

Si desarrollamos las ecuaciones resultantes para n puntos, tendremos un sistema de n ecuaciones lineales con nueve incognitas. Tomemos, por ejemplo, un par de puntos arbitrarios m_{1_i}, m_{2_i}; el desarrollo completo de la ecuación 25 para estos dos puntos será:

$$\begin{aligned} 0 =& x_{m_{2_i}} x_{m_{1_i}} \mathfrak{F}_{00} + x_{m_{2_i}} y_{m_{1_i}} \mathfrak{F}_{01} + x_{m_{2_i}} \mathfrak{F}_{02} + \\ &+ y_{m_{2_i}} x_{m_{1_i}} \mathfrak{F}_{10} + y_{m_{2_i}} y_{m_{1_i}} \mathfrak{F}_{11} + y_{m_{2_i}} \mathfrak{F}_{12} + \\ &+ x_{m_{1_i}} \mathfrak{F}_{20} + y_{m_{1_i}} \mathfrak{F}_{21} + \mathfrak{F}_{22} \end{aligned} \tag{48}$$

donde son constantes $x_{m_{1_i}}$, $y_{m_{1_i}}$, $x_{m_{2_i}}$ e $y_{m_{2_i}}$; y son las variables las distintas $\mathfrak{F}_{ij}$.

Estos sistemas, en principio, tienen una base matemática muy simple, y pueden ser resueltos con las técnicas de resolución de cualquier sistema de ecuaciones, tanto algebráicas -que cualquier alumno de bachillerato domina- como numéricas -que se aprenden en cualquier curso de iniciación a los métodos numéricos de una carrera-.

Sin embargo, las propiedades de la matriz fundamental anteriormente comentadas permiten que no sean necesarios tantos pares de puntos, y que podramos incluir restricciones adicionales basadas en estas propiedades. Por ello, vamos a ver los distintos casos de resolución que tenemos.

5.2. Cálculo de la matriz fundamental con ocho pares de puntos

Podemos calcular la matriz fundamental con ocho pares de puntos. Para ello, desarrollamos la ecuación 48. Con esto tendremos ocho ecuaciones con nueve incógnitas. La novena ecuación la obtenemos sabiendo que el rango de la matriz fundamental es dos. Por ello, una de las filas debe ser linealmente dependiente de otra. Escogemos una fila al azar, y decimos que es igual a otra fila al azar multiplicada por la constante que queramos -salvo 0-. Esto nos dará una ecuación más.

Esto significa que con ocho puntos tendremos exactamente nueve ecuaciones con nueve incognitas.

5.3. Cálculo de la matriz fundamental con más de ocho pares de puntos

Si tenemos ocho puntos o más, siempre podemos emplear las restricciones del algoritmo de ocho puntos, que tiene como característica que nos asegura que la matriz fundamental va a cumplir las propiedades matemáticas de tal matriz; y aplicamos un procedimiento de optimización donde las ecuaciones son las del algoritmo de los ocho puntos; solo que ahora serán las restricciones del algoritmo de optimización, donde optimizamos el error cuadrático medio de ocho variables respecto a las restricciones derivadas de las ecuaciones anteriormente citadas.

Dentro de estos métodos de minimización del error cuadrático medio, podemos usar una función de pseudogradiente -como hace el algoritmo de

RANSAC-, hacer un descenso de gradiente, usar una red neuronal, agregación simulada, un algoritmo genético, un algoritmo mimético o cualquier otro método que queramos para minimizar dicho error cuadrático medio.

5.4. Optimización del algoritmo de cálculo de la matriz fundamental

Aún se puede optimizar más el cálculo de la matriz fundamental. Para ello, suponemos que el resultado de los algoritmos anteriores dará la matriz fundamental estimada, en lugar de la matriz fundamental.

Hacemos:

- Determinamos la SVD de la matriz fundamental estimada F.
- Sea $F = UDV^T$ la descomposición SVD de la matriz fundamental estimada F, la matriz D debe tener la forma:

$$D = \begin{pmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{49}$$

  si la matriz fundamental estimada F fuese exactamente la matriz fundamental $\mathfrak{F}$; pero probablemente, debido a errores numéricos en el calculo y a errores en la exactitud de los puntos, tendrá la forma:

$$D = \begin{pmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & k_3 \end{pmatrix} \tag{50}$$

Por ello, definimos D' como:

$$D' = \begin{pmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{51}$$

- Estimamos la matriz fundamental definitiva $\mathfrak{F}$ como:

$$\mathfrak{F} = UD'V^T \tag{52}$$

Este algoritmo no es necesario si tenemos exactamente ocho puntos; pero si hemos resuelto el problema con más de ocho puntos y en nuestro algoritmo hemos tenido en cuenta nueve incógnitas -es decir, se ha relajado la condición de dependencia lineal de una de las filas-, es importante que apliquemos este algoritmo para mejorar el resultado, y que la matriz fundamental sea realmente la matriz fundamental.

6. La matriz esencial

6.1. Qué es la matriz esencial

En 20 hemos visto las ecuaciones de los puntos de la imagen después de haber calibrado la cámara. Conocida la matriz de calibración de la primera cámara K_1 y de la segunda cámara K_2, podemos aplicarla sobre los puntos m_1 y m_2, obteniendo:

$$\begin{cases} m_{k_1} = & K_1 m_1 \\ m_{k_2} = & K_2 m_2 \end{cases} \tag{53}$$

Si aplicamos estos puntos sobre la formulación de la matriz fundamental, según la vimos en 40, y operamos, obtenemos la expesión:

$$m_{k_1}^T T^{\times} R^{+} m_{k_2} = 0 \tag{54}$$

donde:

- T es el vector de translación entre las dos imágenes.
- R es la matriz de rotación entre las dos imágenes.

Podemos definir la matriz esencial $\mathfrak{E}$ como:

$$\mathfrak{E} = T^{\times} R^{+} \tag{55}$$

por lo que tendremos la expresión:

$$m_{k_1}^T \mathfrak{E} m_{k_2} = 0 \tag{56}$$

Dicho de una forma verbal, al realizar la transformación de calibración para cada uno de los dos puntos, lo que relaciona el punto calibrado de una imagen con el punto calibrado de la otra imagen es la composición de la translación y la rotación de las vistas. Por ello, podemos considerar a la matriz fundamental $\mathfrak{F}$ como la matriz que relaciona dos proyecciones, es decir, que relaciona las dos imágenes de una misma escena; mientras que la matriz esencial $\mathfrak{E}$ relaciona el movimiento relativo para llegar de una vista a la otra vista.

6.2. Propiedades de la matriz esencial

Como ocurría en la matriz fundamental, la matriz esencial $\mathfrak{E}$ tienen un conjunto de propiedades, que son:

- El rango de la matriz esencial $\mathfrak{E}$ es 2.
- Siendo T el vector de translación entre las dos vistas, tenemos que:

$$T^T \mathfrak{E} = 0 \tag{57}$$

- Siendo T el vector de translación entre las dos vistas, y R la matriz de rotación, tenemos que:

$$\mathfrak{E}(RT) = 0 \tag{58}$$

- Siendo la descomposición SVD de la matriz esencial:

$$\mathfrak{E} = UDV^T \tag{59}$$

la forma general de la matriz D será:

$$D = \begin{pmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{60}$$

- Siendo la descomposición SVD de la matriz esencial:

$$\mathfrak{E} = UDV^T \tag{61}$$

la matriz de rotación R será:

$$R = U \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} V^T \tag{62}$$

- Siendo la descomposición SVD de la matriz esencial:

$$\mathfrak{E} = UDV^T \tag{63}$$

el vector de translación T será:

$$T = V \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} V^T \tag{64}$$

7. Cálculo de la matriz esencial

El cálculo de la matriz esencial sigue los siguientes pasos:

- Calculamos la matriz fundamental $\mathfrak{F}$.
- Calculamos la matriz de calibración de la primera cámara K_1.
- Calculamos la matriz de calibración de la segunda cámara K_2.
- La matriz esencial $\mathfrak{E}$ la obtendremos con la expresión:

$$\mathfrak{E} = K_1^T \mathfrak{F} K_2 \tag{65}$$

8. Aplicaciones

Existen muchas aplicaciones de las matrices fundamentales y esenciales. Vamos a hacer una revisión de las aplicaciones más comunes de estos dos tipos de matrices.

8.1. Calibración de una cámara

Es posible emplear la matriz fundamental para calibrar una cámara. Para ello empleamos las denominadas *ecuaciones de Kruppa*, que toman como parámetro la matriz fundamental, los epipolos y dichas matrices de calibración.

Las ecuaciones de Kruppa son la formulación matemática de un concepto matemático denominado *cónica absoluta*. La cónica absoluta es una conica que tiene como característica que está contenida en el plano del infinito; y nos interesa porque su proyección depende exclusívamente de los parámetros intrínsecos de la cámara.

Las ecuaciones de Kruppa se desarrollan a partir de la ecuación:

$$\mathfrak{F}K\mathfrak{F}^T = \beta e \times K(e^{\times})^T \tag{66}$$

donde β es un escalar desconocido, distinto de 0. Esta ecuación matricial se puede desarrollar por cada elemento de la matriz de cualquiera de los dos lados, dando lugar a la formulación matemática de las ecuaciones de Kruppa:

$$\begin{aligned}
\alpha &= \frac{\mathfrak{F}K(\mathfrak{F}^T)_{00}}{(e \times K(e^{\times})^T)_{00}} \\
\alpha &= \frac{\mathfrak{F}K(\mathfrak{F}^T)_{01}}{(e \times K(e^{\times})^T)_{01}} \\
\alpha &= \frac{\mathfrak{F}K(\mathfrak{F}^T)_{02}}{(e \times K(e^{\times})^T)_{02}} \\
\alpha &= \frac{\mathfrak{F}K(\mathfrak{F}^T)_{11}}{(e \times K(e^{\times})^T)_{11}} \\
\alpha &= \frac{\mathfrak{F}K(\mathfrak{F}^T)_{12}}{(e \times K(e^{\times})^T)_{12}} \\
\alpha &= \frac{\mathfrak{F}K(\mathfrak{F}^T)_{22}}{(e \times K(e^{\times})^T)_{22}}
\end{aligned} \tag{67}$$

Donde:

- α es la variable usada para simplificar la formulación de las ecuaciones, pero de ningún valor intrínseco para nuestro problema.
- $\mathfrak{F}$ es la matriz fundamental de las dos imágenes.
- $(\mathfrak{F}^T)_{ij}$ es el elemento i, j de la transpuesta de la matriz fundamental $\mathfrak{F}$.
- K es la matriz de calibración.
- e es la proyección en la segunda imagen del epipolo de la primera imagen.
- $(e \times K(e^{\times})^T)_{ij}$ es el elemento i, j de la matriz resultado de realizar la operación $(e \times K(e^{\times})^T)$.

Estas ecuaciones hacen que K esté sobredeterminado en unos polinomios de segundo orden, y que, por lo tanto, no pueda ser resuelto de forma simple mediante las reglas de sistemas lineales. Sin embargo, puede ser resuelto numéricamente como cualquier otro sistema no lineal sobredimensionado; o puede ser resuelto por algunas de las técnicas desarrolladas específicamente para las ecuaciones de Kruppa.

8.2. Egomotion

Denominamos un algoritmo de *egomotion* a aquel algoritmo que nos permite describir el movimiento de un sensor -en este caso, una cámara- a partir de la imagen generada por dicho sensor.

La matriz esencial puede ser empleada directamente para el cálculo de egomotion; para hacer esto, recordamos que entre las propiedades de la SVD estaba en que a partir de las matrices de la descomposición podíamos obtener la translación y la rotación entre una imagen y la otra. Definida una transformación SVD con la ecuación $\mathfrak{E} = UDV^T$, sabemos que la rotación R viene formulada por la ecuación:

$$R = U \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} V^T \tag{68}$$

y el vector de translación T viene definido por la ecuación:

$$T = V \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} V^T \tag{69}$$

Sin embargo, si hacemos esto tal y como hemos comentado, no funcionará, ya que el error acumulado al escoger los pares de puntos, el error por la correspondencia de puntos, el error al calcular la matriz fundamental, el error en la matriz de calibración, el error al calcular la matriz esencial, el error del método numérico de cálculo de la SVD y los errores por usar aritmética finita en todos los calculos involucrados, harán que se parezca muy poco el resultado al desplazamiento y la rotación real de la segunda cámara respecto a la primera.

Pero hay una solución: a la hora de calcular la SVD, sabemos que existen un dos propiedades importantes que tiene la matriz esencial, y que se deben de cumplir: el rango de la matriz esencial debe ser dos, y sus dos autovalores distintos de cero deben ser iguales. En la práctica, significa que en la descomposición SVD debemos tener dos elementos en la matriz diagonal distintos de cero, y los dos en la matriz diagonal. Por ello, igualando a cero el tercer elemento de la diagonal y calculando la media de los elementos que encontramos en la matriz diagonal nos encontramos con una nueva matriz diagonal, a la que si empleamos para montar de nuevo la matriz -realizando el producto con U por la izquierda y el producto con la transpuesta de V por la derecha-, obtendremos una mejor aproximación de la matriz esencial. La transformada SVD de la nueva matriz esencial, si bien sigue sin ser exacta, tiene una precisión razonable, por lo que podemos aproximarla a la matriz esencial $\mathfrak{E}$. El proceso lo podemos repetir iterattivamente, pero no lo haremos porque la mayor parte de las veces esto no será necesario, y bastará con hacer la corrección una única vez.

Por tanto, el algoritmo de egomotion con la matriz esencial tiene como pasos:

- Encuentra la correspondencia entre n puntos de la primera imagen y de la segunda imagen. Deben ser, al menos, 7 puntos.
- Normaliza los datos, para reducir el error numérico en los cálculos. El objetivo de la normalización es que la media de los datos sea 0. Un ejemplo de matriz de normalización es:

$$H = \begin{pmatrix} a & 0 & c \\ 0 & b & d \\ 0 & 0 & 1 \end{pmatrix} \tag{70}$$

- Calculamos la matriz fundamental $\mathfrak{F}$, según vimos en la sección el cálculo de la matriz fundamental.
- Calculamos la matriz esencial estimada E, según vimos en la sección sobre el cálculo de la matriz esencial.
- Determinamos la SVD de la matriz esencial estimada E.
- Sea $E = UDV^T$ la descomposición SVD de la matriz esencial estimada E, la matriz D debe tener la forma:

$$D = \begin{pmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{71}$$

  si la matriz esencial estimada E fuese exactamente la matriz esencial $\mathfrak{E}$; pero probablemente, debido a errores numéricos en el calculo y a errores en la exactitud de los puntos, tendrá la forma:

$$D = \begin{pmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & k_3 \end{pmatrix} \tag{72}$$

  Por ello, definimos D' como:

$$D' = \begin{pmatrix} \frac{k_1+k_2}{2} & 0 & 0 \\ 0 & \frac{k_1+k_2}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{73}$$

- Estimamos la matriz esencial definitiva $\mathfrak{E}$ como:

$$\mathfrak{E} = UD'V^T \tag{74}$$

- Siendo la descomposición SVD de la matriz esencial:

$$\mathfrak{E} = UD'V^T \tag{75}$$

la matriz de rotación R será:

$$R = U \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} V^T \tag{76}$$

- Siendo la descomposición SVD de la matriz esencial:

$$\mathfrak{E} = UDV^T \tag{77}$$

el vector de translación T será:

$$T = V \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} V^T \tag{78}$$

8.3. Estereo

Conocidas las matrices de calibración y la matriz esencial, es facil deducir para un punto M cuales son sus proyecciones m_1 y m_2.

Aqui hablaremos de una reconstrucción tridimensional rápida y eficiente, basada en la similaridad y partiendo del supuesto de que la matriz de calibración ha sio calculada por otro método, por lo que es conocida. Existen otros métodos capaces de resolver el problema de una reconstrucción tridimensional euclidea plena, pero no he sido capaz de entender el método a tiempo.

El algoritmo de la reconstrucción de estéreo por similaridad es:

- Encuentra la correspondencia entre n puntos de la primera imagen y de la segunda imagen. Deben ser, al menos, 7 puntos.
- Normaliza los datos, para reducir el error numérico en los cálculos. El objetivo de la normalización es que la media de los datos sea 0. Un ejemplo de matriz de normalización es:

$$H = \begin{pmatrix} a & 0 & c \\ 0 & b & d \\ 0 & 0 & 1 \end{pmatrix} \tag{79}$$

- Calculamos la matriz fundamental $\mathfrak{F}$, según ya hemos visto en la sección sobre el cálculo de la matriz fundamental.
- Calculamos la matriz esencial estimada E, según vimos hemos visto en la sección sobre el cálculo de la matriz esencial.
- Determinamos la SVD de la matriz esencial estimada E.
- Sea $E = UDV^T$ la descomposición SVD de la matriz esencial estimada E, la matriz D debe tener la forma:

$$D = \begin{pmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{80}$$

si la matriz esencial estimada E fuese exactamente la matriz esencial $\mathfrak{E}$; pero probablemente, debido a errores numéricos en el calculo y a errores en la exactitud de los puntos, tendrá la forma:

$$D = \begin{pmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & k_3 \end{pmatrix} \tag{81}$$

Por ello, definimos D' como:

$$D' = \begin{pmatrix} \frac{k_1+k_2}{2} & 0 & 0 \\ 0 & \frac{k_1+k_2}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{82}$$

- Estimamos la matriz esencial definitiva $\mathfrak{E}$ como:

$$\mathfrak{E} = UD'V^T \tag{83}$$

- Siendo la descomposición SVD de la matriz esencial:

$$\mathfrak{E} = UD'V^T \tag{84}$$

la matriz de rotación R será:

$$R = U \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} V^T \tag{85}$$

- Siendo la descomposición SVD de la matriz esencial:

$$\mathfrak{E} = UDV^T \tag{86}$$

el vector de translación T será:

$$T = V \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} V^T \tag{87}$$

- Concidas las matrices de calibración K_1 y K_2, la matriz de rotacion entre las imágenes R, y el vector de translación T, obtenemos el punto real M del desarrollo de las ecuaciones:

$$\begin{cases} m_1 & \approx [K_1|0]M \\ m_2 & \approx [(K_2R)|(-K_2RT)]M \end{cases} \tag{88}$$

  Si incluimos todos los parámetros conocidos en estas ecuaciones, podremos despejar algebraicamente los cuatro parámetros de M.

9. Conclusión

En este informe técnico hemos revisado los conceptos necesarios para entender y usar las matrices fundamentales y las matrices esenciales, y hemos estudiado dichas matrices, incluyendo deducción, formulación, cálculo de estas matrices y algunos de sus usos.

10. Bibliografía comentada

- *Base de álgebra necesaria*: José Martinez Salas, Elementos de Matemáticas, ed. Lex Nova

- *Base de geometría proyectiva*: Departamento de Ingeniería de Sistemas y Automática, Universidad Politécnica de Madrid. Apuntes de geometría proyectiva y visión artificial.

- *Calculo de la matriz fundamental*: Nicolás Pérez de la Blanca Capilla, Apuntes de visión artificial, tema 8.

- *Estéreo y Egomotion*: ?
- *Calibración de la cámara*: Manolis I.A. Lourakis, Rachid Deriche; Camera Self-Calibration Using Singular Value Decomposition of the Fundamental Matrix. INRIA research report.
- *Error asociado al cálculo de la matriz fundamental*: Csurka, G; Zeller, C. et. al.; Characterizing the Uncertainy of the Fundamental Matrix. INRIA research report.