Máster Universitario en
Nuevas Tecnologías en Informática

Asignatura “Visión Artificial”

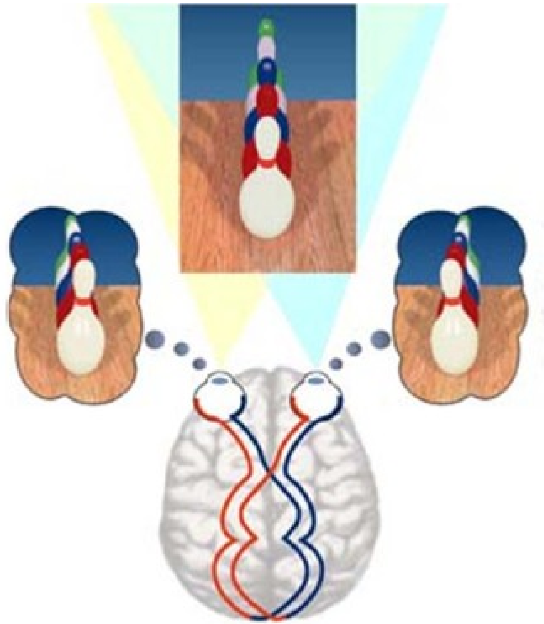
Geometría Proyectiva II

Structure from motion:
Geometría Epipolar
Autocalibración
Obtención de estructura 3D
Reconstrucción densa

Facultad de Informática
Universidad de Murcia
Curso 2018/19

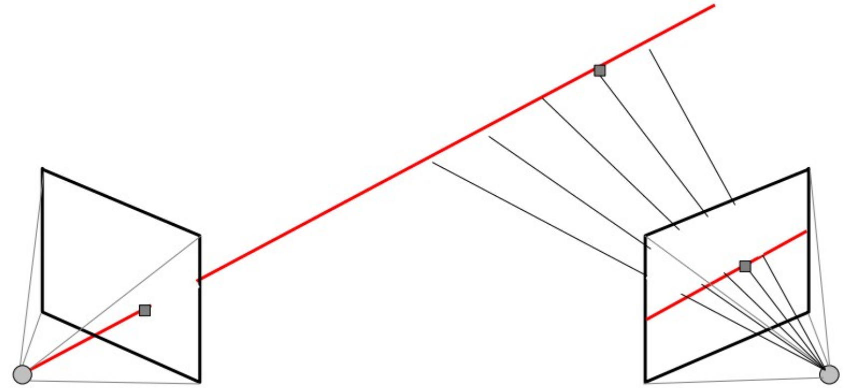
Introducción

- Objetivo:
  - Recuperar información de “profundidad” (estructura 3D de la escena) a partir de dos (o más) imágenes:
  - Fácil si se conoce totalmente la calibración (extrínseca e intrínseca) de las cámaras...
  - … pero también se puede hacer bajo ciertas circunstancias aunque no se conozca dicha información a priori.

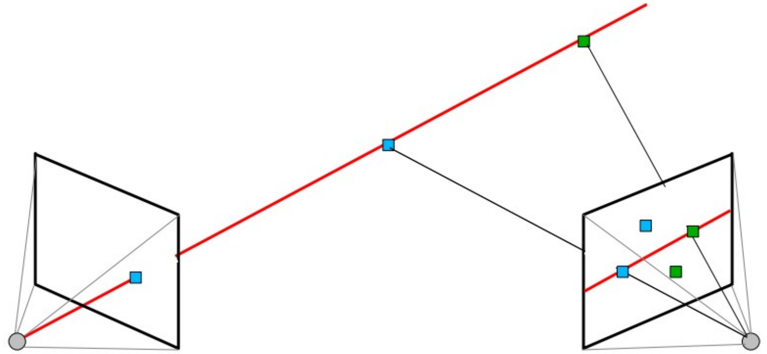


Restricción epipolar (I)

- Dado un punto 3D, cuya proyección 2D conocemos en la imagen izquierda → su proyección en la imagen derecha ha de estar forzosamente en una línea (**recta epipolar**):

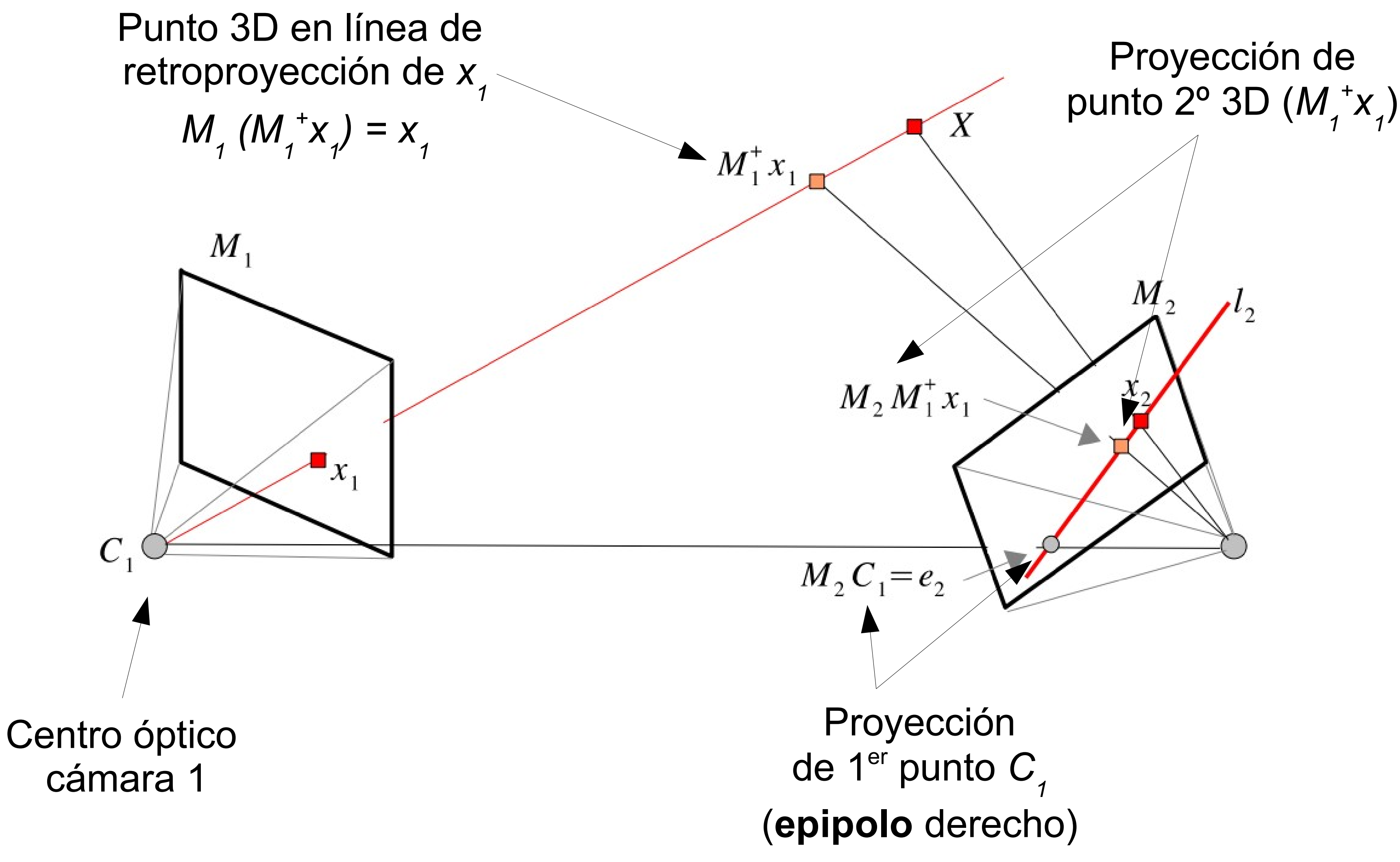


- → Esto restringe la zona en la que buscar *matchings*:



Restricción epipolar (II)

- ¿Cuál es la recta epipolar l_2 correspondiente al punto x_1?



Matriz fundamental (I)

- l_2 por tanto, es la recta que pasa por la proyección de dicho par de puntos 3D en la imagen 2 (*cross product*):

$$\boldsymbol{l}_2 = \boldsymbol{e}_2 \times \mathtt{M}_2\mathtt{M}_1^+ \boldsymbol{x}_1$$

- Y por tanto x_2 debe pasar por dicho l_2, es decir:

$$\boldsymbol{x}_2^\top \boldsymbol{l}_2 = \boldsymbol{x}_2^\top \underbrace{\boldsymbol{e}_2 \times \mathtt{M}_2\mathtt{M}_1^+}_{\mathtt{F}} \boldsymbol{x}_1 = 0$$

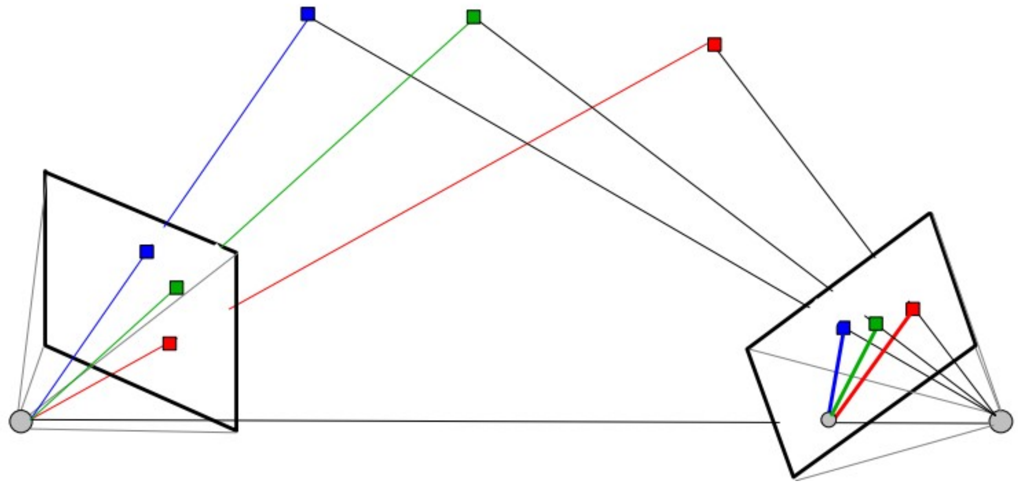
- Si recordamos que $[\boldsymbol{v}]_\times \boldsymbol{x} = \boldsymbol{v} \times \boldsymbol{x}$, entonces:

$$\boxed{\boldsymbol{x}_2^\top \, \mathtt{F} \, \boldsymbol{x}_1 = 0} \quad \text{con} \quad \boxed{\mathtt{F} = [\boldsymbol{e}_2]_\times \mathtt{M}_2\mathtt{M}_1^+} \quad \text{y} \quad \begin{bmatrix} a \\ b \\ c \end{bmatrix}_\times = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix}$$

Restricción epipolar **Matriz fundamental** *Cross product matrix*

Matriz fundamental (II)

- Todas las líneas epipolares de la imagen derecha pasan por el correspondiente epipolo e_2, independientemente del punto x_1 del que vengan:



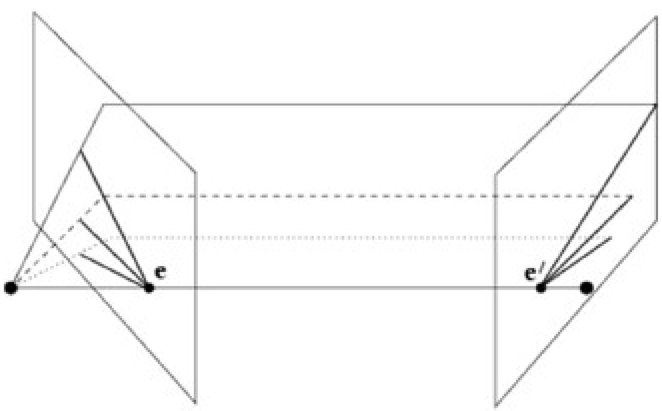
- Por ello, para todo x_1 se cumple que $\mathbf{e}_2^{\mathsf{T}} \mathrm{F} \boldsymbol{x}_1 = 0$; es decir que, forzosamente:

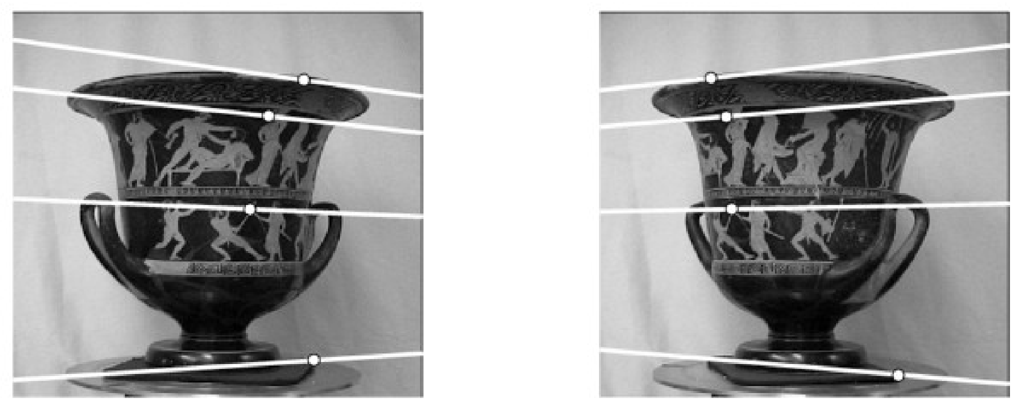
$$\mathbf{e}_2^{\mathsf{T}} \mathrm{F} = 0^{\mathsf{T}}$$

El epipolo derecho es justamente el espacio nulo de F^T ($F^T e_2 = \mathbf{0}$)

Líneas epipolares (I)

- Ejemplo real:
  - Comprobar visualmente cómo en cada par de líneas epipolares correspondientes han de estar todas las correspondencias de puntos:

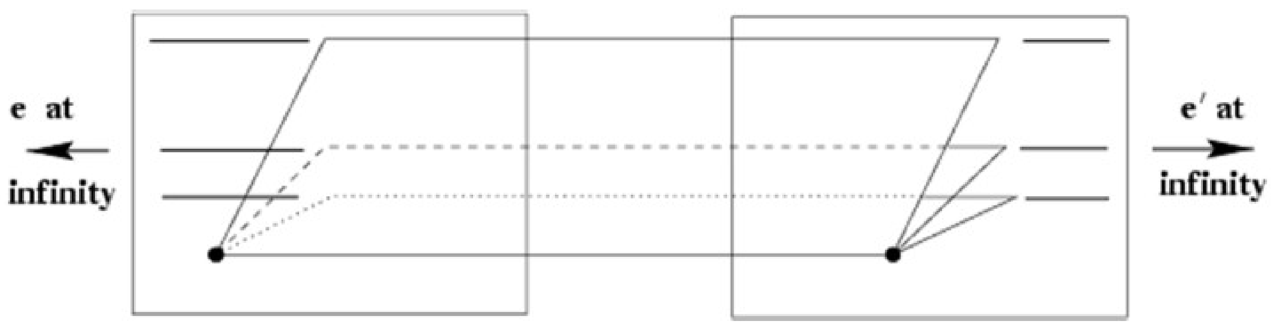




  - (en este caso ambos epipolos “caen fuera” de la imagen)

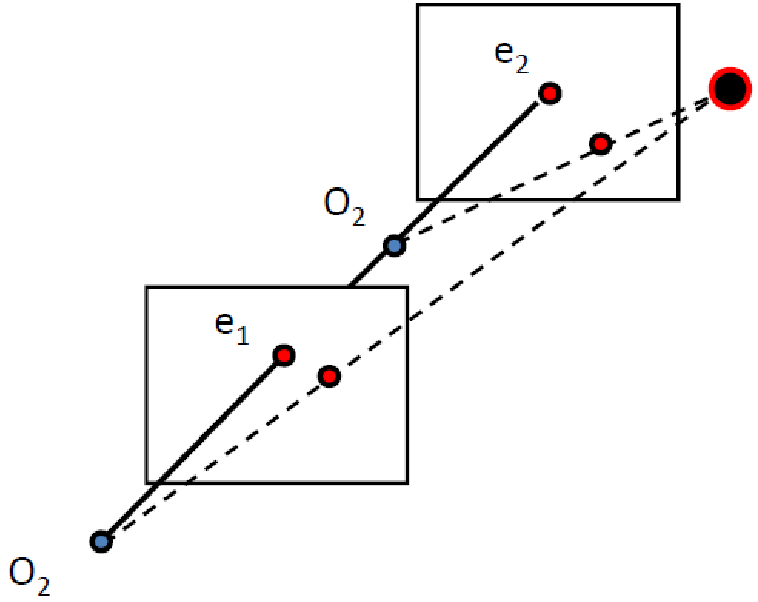
Líneas epipolares (II)

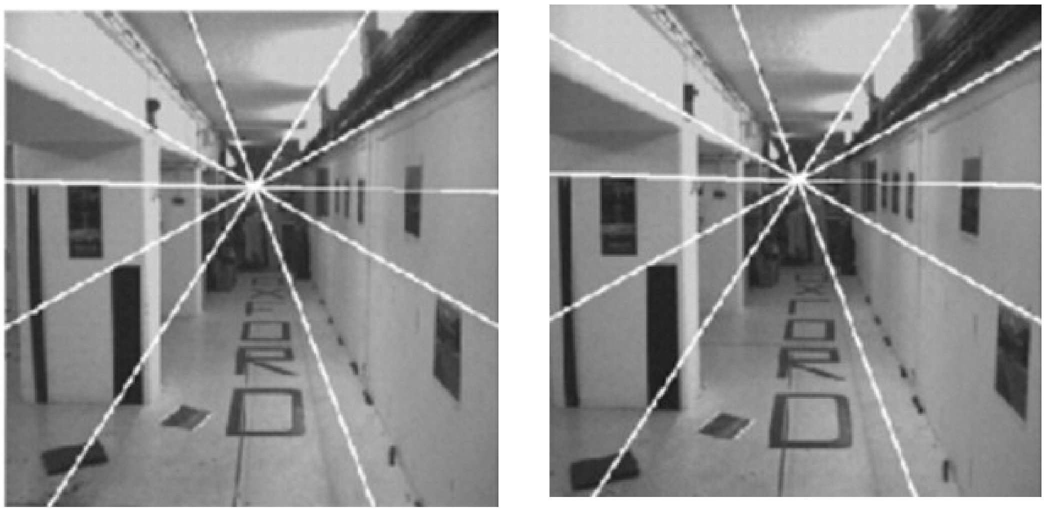
- Otro ejemplo: planos de imagen alineados:
  - Las líneas epipolares generadas son paralelas
  - Los epipolos caen en el infinito



Líneas epipolares (III)

- Un tercer ejemplo: traslación pura (sin rotación)
  - Las líneas epipolares convergen en un “foco de expansión”, correspondiente al epipolo (que cae en la misma posición en ambas imágenes):





Cómputo de F

- Para cada *matching* $\boldsymbol{p}=(u,v,1)^T \leftrightarrow \boldsymbol{p'}=(u',v',1)^T$, usando la restricción epipolar $p^TFp'=0$, podemos obtener una ecuación (homogénea) para los 9 coeficientes de F (F_{11}, … F_{33}):

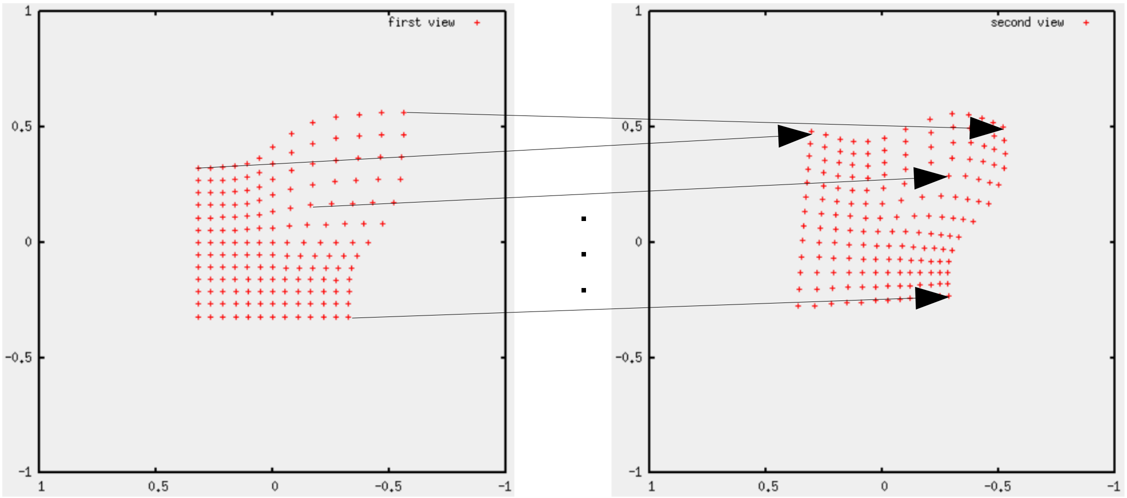
$$\mathrm{p}^{\mathrm{T}}\ \mathrm{F}\ \mathrm{p}' = 0 \quad \Rightarrow$$

$$(u, v, 1)\begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0$$

$$\Rightarrow \quad (uu', uv', u, vu', vv', v, u', v', 1)\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix} = 0$$

Algoritmo de los 8 puntos (I)

- Análogamente a lo que ocurría en la estimación de las homografías, hay 9 incógnitas, pero por el factor de homogeneidad sólo hay 9-1=8 gdl.
- Puesto que cada correspondencia induce ahora 1 ecuación (no 2 como en la estimación de H)...
- … hacen falta al menos 8 correspondencias para resolver y obtener una F completa (no 4 como para estimar una H):



Algoritmo de los 8 puntos (II)

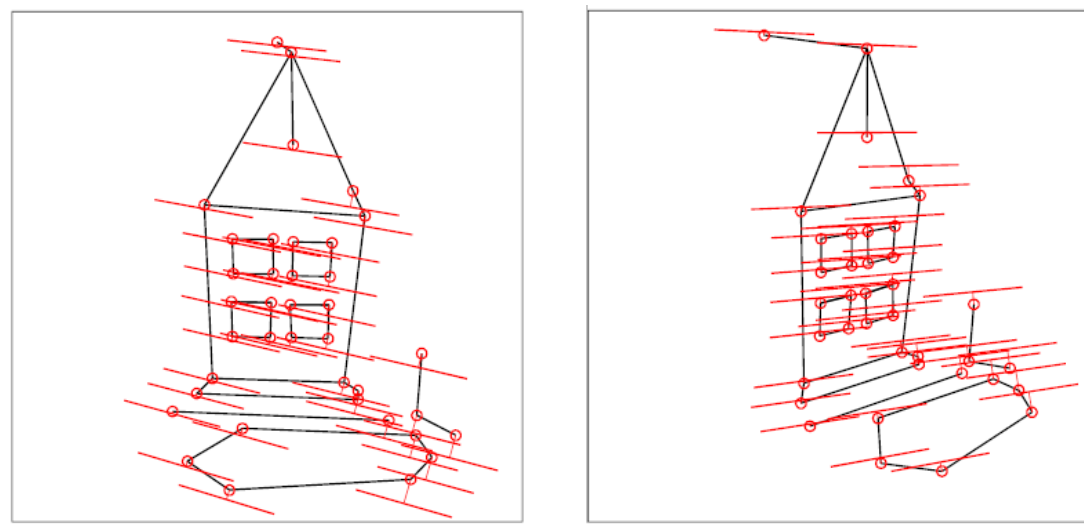
- Queda entonces un sistema homogéneo que se resuelve igual que hacíamos en el tema anterior para H (SVD, eigendecomposition, …):

$$\underbrace{\begin{pmatrix} u_1u_1' & u_1v_1' & u_1 & v_1u_1' & v_1v_1' & v_1 & u_1' & v_1' & 1 \\ u_2u_2' & u_2v_2' & u_2 & v_2u_2' & v_2v_2' & v_2 & u_2' & v_2' & 1 \\ u_3u_3' & u_3v_3' & u_3 & v_3u_3' & v_3v_3' & v_3 & u_3' & v_3' & 1 \\ u_4u_4' & u_4v_4' & u_4 & v_4u_4' & v_4v_4' & v_4 & u_4' & v_4' & 1 \\ u_5u_5' & u_5v_5' & u_5 & v_5u_5' & v_5v_5' & v_5 & u_5' & v_5' & 1 \\ u_6u_6' & u_6v_6' & u_6 & v_6u_6' & v_6v_6' & v_6 & u_6' & v_6' & 1 \\ u_7u_7' & u_7v_7' & u_7 & v_7u_7' & v_7v_7' & v_7 & u_7' & v_7' & 1 \\ u_8u_8' & u_8v_8' & u_8 & v_8u_8' & v_8v_8' & v_8 & u_8' & v_8' & 1 \end{pmatrix}}_{\mathbf{W}} \underbrace{\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix}}_{\mathbf{f}} = 0$$

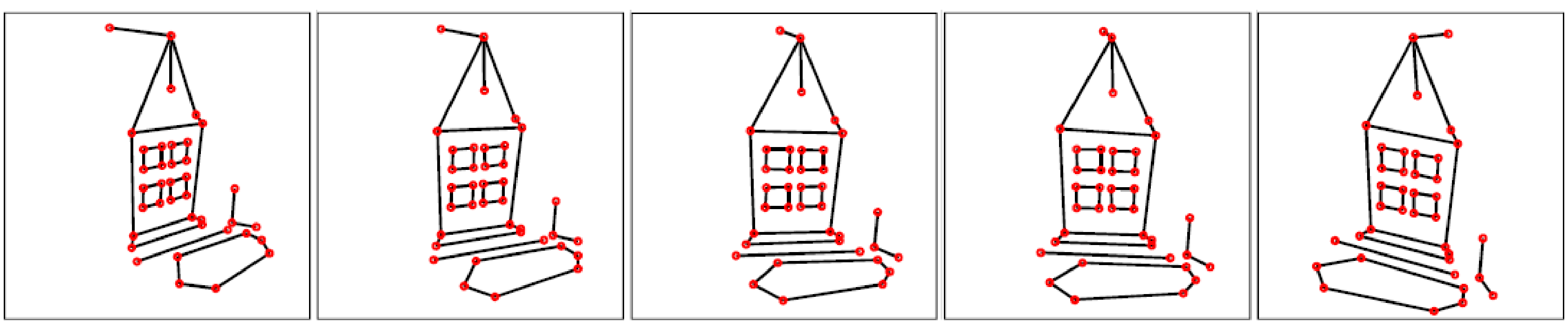
- Todo lo que allí aplicábamos vale también aquí:
  - Si hay más de 8 correspondencias → **mínimos cuadrados**.
  - Conveniente **prenormalizar** coordinadas de entrada (para evitar problemas numéricos).

Ejemplo de estimación de F (I)

Extracción de features
(sólo imagen izquierda)



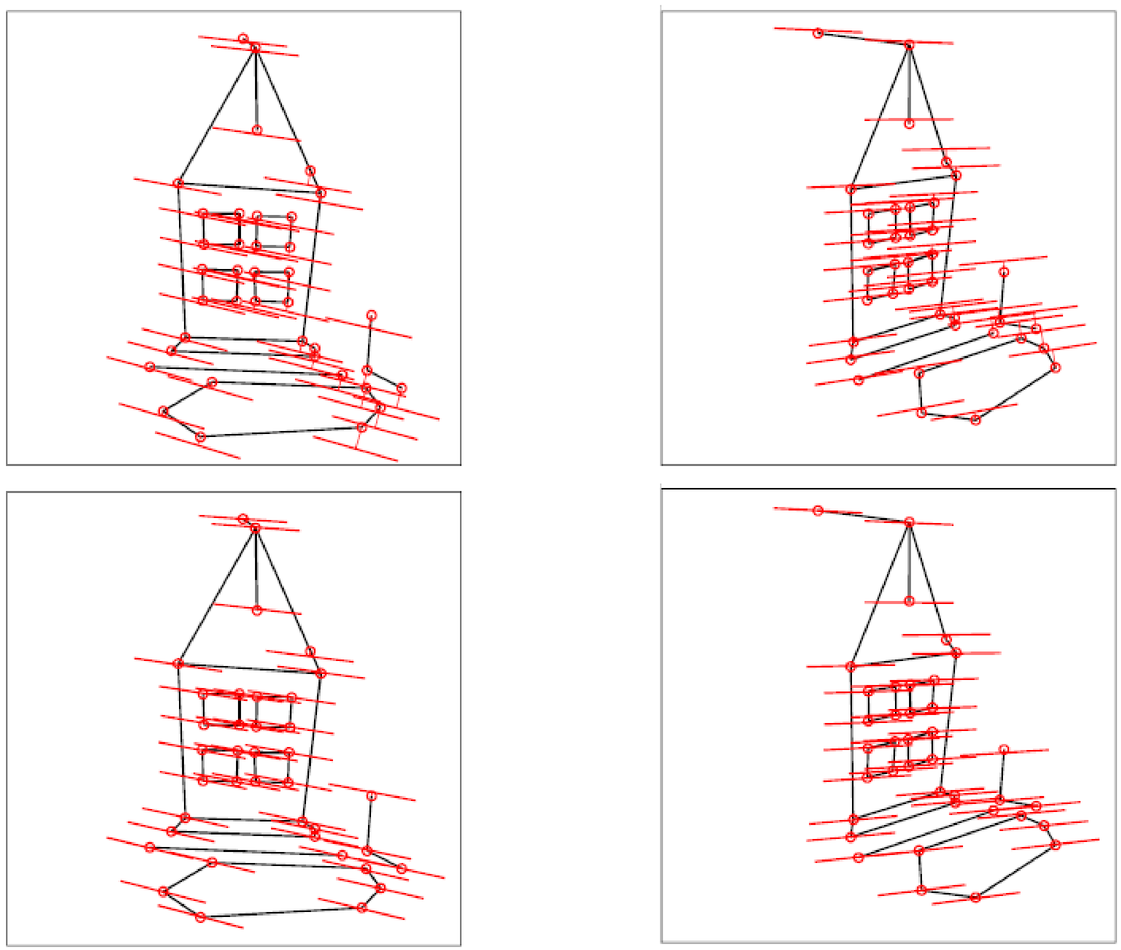
Líneas epipolares para la F
obtenida (cada $l'=Fx$, y $l=F^Tx'$)



Reconstrucción 3D obtenida, vista desde otros lugares
(veremos más adelante cómo se realiza)

Ejemplo de estimación de F (II)

- Influencia de la normalización:



Sin normalización (errores medios 9~10 píxeles)

Con normalización (errores medios ~1 píxel)

Propiedades de la matriz fundamental

- La restricción epipolar es <u>simétrica</u>:
  - Todos los resultados dados para la imagen derecha son válidos igualmente para la imagen izquierda (cambiando F por F^T)

- Por su propia forma ("contiene" en su factorización una matriz $[e_2]_x$ antisimétrica, y por tanto de rango 2), F es <u>deficitaria en rango</u>:
  - F tiene rango 2 (y por tanto, sólo dos valores singulares no nulos), a pesar de ser de 3x3.

Algoritmo de los 8 puntos (y III)

- Puesto que hemos visto que la F ha de cumplir la **restricción de tener rango 2**...
- … y dicha restricción no tiene por qué cumplirla la matriz obtenida por el algoritmo de los 8 puntos...
- … como último paso del mismo se suele **forzar** dicha condición realizando la SVD de $F=U \cdot diag(v_1,v_2,v_3) \cdot V^T$, (que debe tener un tercer valor singular v_3 cercano a cero)...
- … y reconstruyendo la F eliminando dicho v_3, es decir, $F_{rank2}=U \cdot diag(v_1,v_2,0) \cdot V^T$

Matriz esencial (I)

- Considérense dos cámaras en posición general, donde a la primera, por comodidad, se la considera alineada con el sistema de coordenadas del mundo ($R_1=I$, $C_1=(0,0,0)^T$), mientras que la otra está en una posición general ($R_2=R$, $C_2=C$):

$$\mathtt{M}_1 = \mathtt{K}_1[\mathtt{I}|\mathbf{0}] \qquad \boldsymbol{t} = -\mathtt{R}\boldsymbol{C}$$

$$\mathtt{M}_2 = \mathtt{K}_2[\mathtt{R}| - \mathtt{R}\boldsymbol{C}_2] = [\mathtt{K}_2\mathtt{R}|\mathtt{K}_2\boldsymbol{t}]$$

- Puesto que el centro (homogéneo) de M_1 es $(0,0,0,1)^T$, el epipolo derecho $e_2=M_2(0,0,0,1)^T$, *es decir:*

$$\boldsymbol{e}_2 = \mathtt{K}_2\boldsymbol{t}$$

Última columna de M_2

Matriz esencial (II)

- Por otro lado, por la sencilla forma de *M₁* se tiene que $M_1^+ = [I|\mathbf{0}]^\top K_1^{-1}$, con lo que:

$$M_2 M_1^+ = K_2 R K_1^{-1}$$

- Según la forma que sabíamos que tenía F:

$$F = [\boldsymbol{e}_2]_\times M_2 M_1^+$$

- Resulta ahora que F, en función ya de la calibración de M_1 y M_2, resulta ser:

$$\boldsymbol{t} = -R\boldsymbol{C}$$

$$F = [K_2 \boldsymbol{t}]_\times K_2 R K_1^{-1}$$

Matriz esencial (III)

- Aprovechando la siguiente propiedad de las matrices *cross product* (que no demostraremos aquí; HZ, apéndice):

Pasar una matriz *A* "a la izquierda" de una $[v]_x$

$$[\boldsymbol{v}]_\times \mathtt{A} = \mathtt{A}^{-\mathsf{T}} [\mathtt{A}^{-1}\boldsymbol{v}]_\times$$

- … podemos manipular la expresión anterior para "llevar" K_2 a la izquierda de la expresión

$$\mathtt{F} = \mathtt{K}_2^{-\mathsf{T}} [\mathtt{K}_2^{-1}\mathtt{K}_2\boldsymbol{t}]_\times \mathtt{R}\mathtt{K}_1^{-1} = \mathtt{K}_2^{-\mathsf{T}} \; [\boldsymbol{t}]_\times \mathtt{R} \; \mathtt{K}_1^{-1} = \mathtt{K}_2^{-\mathsf{T}} \mathtt{E} \, \mathtt{K}_1^{-1}$$

Matriz esencial:
- Depende sólo de extrínsecos (R y C)
- 5 gdl (3+3-1 por homogeneidad)

$$\boxed{\mathtt{E} = [\boldsymbol{t}]_\times \mathtt{R}}$$

Matriz esencial (IV)

- La matriz esencial trabaja con “rayos reales” (puntos de imagen a los que se “deshace” la transformación de imagen usando las inversas de las calibraciones intrínsecas K_1 y K_2):

$$\mathrm{F} = \mathrm{K}_2^{-\top} \mathrm{E}\, \mathrm{K}_1^{-1}$$

$$\boldsymbol{x}_1^n = \mathrm{K}_1^{-1} \boldsymbol{x}_1$$
$$\boldsymbol{x}_2^n = \mathrm{K}_2^{-1} \boldsymbol{x}_2$$

$$\boxed{\boldsymbol{x}_2^\top\, \mathrm{F}\, \boldsymbol{x}_1 = 0} \longleftrightarrow \boxed{\boldsymbol{x}_2^{n\top}\, \mathrm{E}\, \boldsymbol{x}_1^n = 0}$$

Propiedades de la matriz esencial

- Sólo depende de las posiciones relativas entre las cámaras, dadas por *R* y *C* (y no de K_1 y K_2):

Recordar que:

$$\boldsymbol{t} = -\mathrm{R}\boldsymbol{C}$$

$$\boxed{\mathrm{E} = [\boldsymbol{t}]_\times \mathrm{R}}$$

- Al igual que F, tiene tamaño 3x3 y rango 2 (por lo que su tercer valor singular es nulo)...
- ...pero además, en este caso por su forma particular (desaparecen K_1 y K_2), sus dos valores singulares no nulos <u>son iguales</u>.

Ambigüedad proyectiva (I)

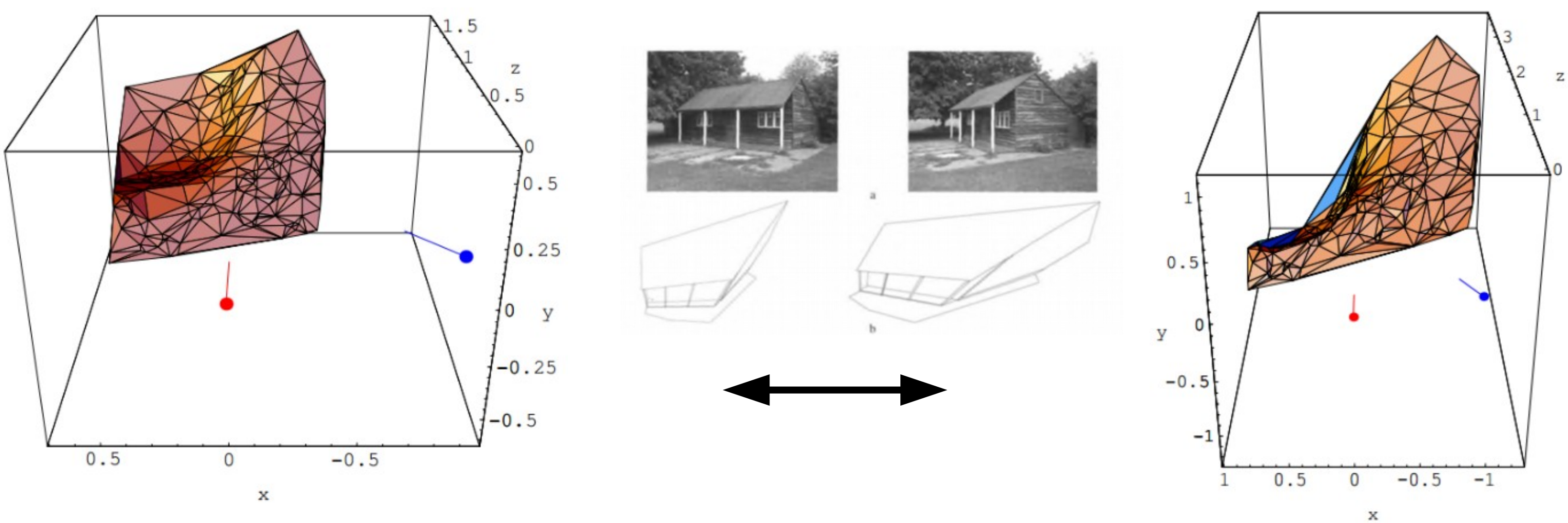
- Supongamos que, a partir de una serie de correspondencias entre puntos de dos imágenes ({*x1*↔*x2*}) llegamos a una determinada solución compuesta por dos cámaras M_1 y M_2, y el correspondiente conjunto de puntos 3D triangulados {X}.
- El problema es que, si dicha solución satisface las ecuaciones de proyección, también lo harán las siguientes matrices P_1 y P_2, y el conjunto de puntos 3D X' alternativos (obtenidos con cualquier matriz D 4x4, invertible):

$$\boldsymbol{x}_1 = \mathtt{M}_1 \boldsymbol{X} = \underbrace{\mathtt{M}_1 \mathtt{D}}_{\mathtt{P}_1} \underbrace{\mathtt{D}^{-1} \boldsymbol{X}}_{\boldsymbol{X}'} = \mathtt{P}_1 \boldsymbol{X}'$$

$$\boldsymbol{x}_2 = \mathtt{M}_2 \boldsymbol{X} = \underbrace{\mathtt{M}_2 \mathtt{D}}_{\mathtt{P}_2} \underbrace{\mathtt{D}^{-1} \boldsymbol{X}}_{\boldsymbol{X}'} = \mathtt{P}_2 \boldsymbol{X}'$$

Ambigüedad proyectiva (II)

- Hay **infinitas reconstrucciones** (par de cámaras y conjunto de puntos 3D) consistentes con un par de proyecciones.
  - Podríamos obtenerlas todas ellas simplemente usando las distintas matrices *D* 4x4 (con 16-1=15 *gdl*).
  - Las matrices *D* no son más que homografías 3D, que transforman el espacio de forma similar a como un plano es deformado por una homografía 2D.



Ambigüedad proyectiva (III)

- **Teorema fundamental de la geometría proyectiva**:
  - Si no se tiene información adicional de calibración...
  - ...sólo a partir de correspondencias entre puntos de dos imágenes se puede obtener una fundamental F...
  - ...y a partir de ella una reconstrucción 3D...
  - ...PERO siempre tendremos la mencionada <u>ambigüedad proyectiva</u>.
- Sin embargo, <u>conociendo la calibración</u> de las cámaras (K_1 y K_2), es posible obtener una **reconstrucción 3D similar**:
  - Donde la única ambigüedad sería la escala global del conjunto escena-cámaras:
    - "París *vs*. maqueta de París"

Autocalibración (I)

- Es importante, pues, disponer de la calibración:
  - Si la conocemos a priori, OK
  - ¿Y si no? → **procedimientos de autocalibración**.
- Múltiples posibilidades, pero casi todas basadas en la manipulación algebraica de las matrices de cámara / homografías para “cancelar” rotaciones aprovechando propiedad $R \cdot R^T = R^T \cdot R = I$
- Utilizaremos las siguientes matrices de conveniencia para, introduciéndolas en las expresiones, lograr dichas cancelaciones:

$$\omega = \mathrm{K}^{-\mathsf{T}}\mathrm{K}^{-1} \qquad \omega^* \equiv \omega^{-1} = \mathrm{K}\mathrm{K}^{\mathsf{T}}$$

- ... y obtener así ecuaciones para sus coeficientes y, a partir de ellos, extraer las matrices de calibración.

Autocalibración (II)

- Así pues, resolviendo ***w***, y sabiendo que:

$$\omega = K^{-T} K^{-1}$$

- *K* puede entonces obtenerse a partir de la **descomposición de Cholesky** de ***w*** (invirtiendo la K^{-1} así obtenida):
- Recordemos que:
  - La descomposición de Cholesky de una matriz real simétrica y definida positiva *A* se define como $A=L^{T}L$, con *L* una matriz triangular superior.
  - La inversa de una matriz triangular superior con diagonal no nula siempre existe, y es a su vez siempre otra matriz triangular superior.

Autocalibración (III)

- Diversas simplificaciones de cámara de las comúnmente empleadas (recordar tema anterior), surgen distintas matrices ***w*** (o bien ***w****) simplificadas, cada vez con menos *gdl*:

$$\mathtt{K} = \begin{bmatrix} f & 0 & o_x \\ 0 & fr & o_y \\ 0 & 0 & 1 \end{bmatrix} \qquad \rightarrow \qquad \boldsymbol{\omega} = \begin{bmatrix} a & 0 & c \\ 0 & b & d \\ c & d & 1 \end{bmatrix}$$

$$\mathtt{K} = \begin{bmatrix} f & 0 & 0 \\ 0 & fr & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \rightarrow \qquad \boldsymbol{\omega} = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathtt{K} = \mathtt{K}_f = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \rightarrow \qquad \boldsymbol{\omega} = \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Autocalibración mediante planos (I)

- <u>Primera posibilidad</u>: calibrar mediante **homografías plano-imagen**:
  - Sabemos (tema anterior) que cualquier *H* plano-imagen tenía la forma:

$$\mathrm{H} = \mathrm{K}[\hat{\mathrm{R}}|\boldsymbol{t}]$$

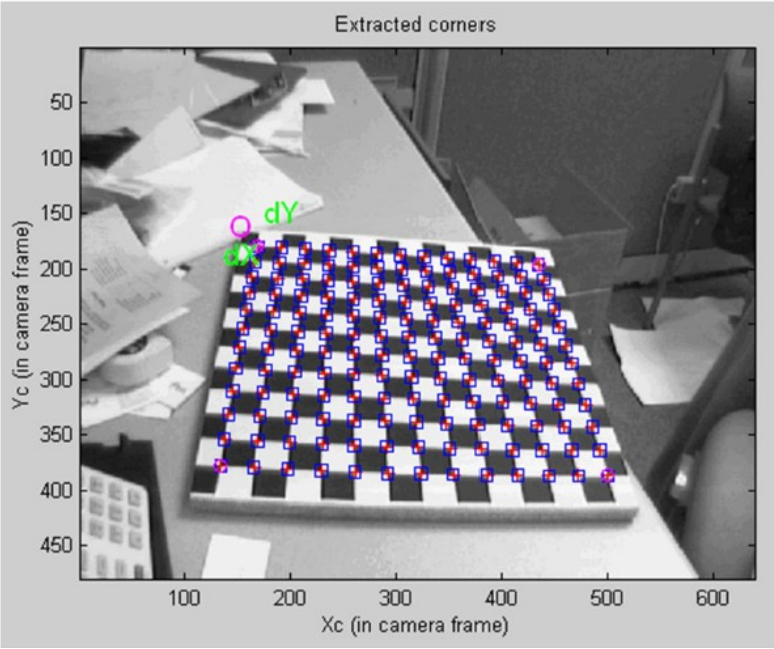
    (es decir, sus dos primeras columnas son ortogonales)
  - Entonces, dado que $\boldsymbol{\omega} = \mathrm{K}^{-\mathsf{T}}\mathrm{K}^{-1}$, sabemos que:

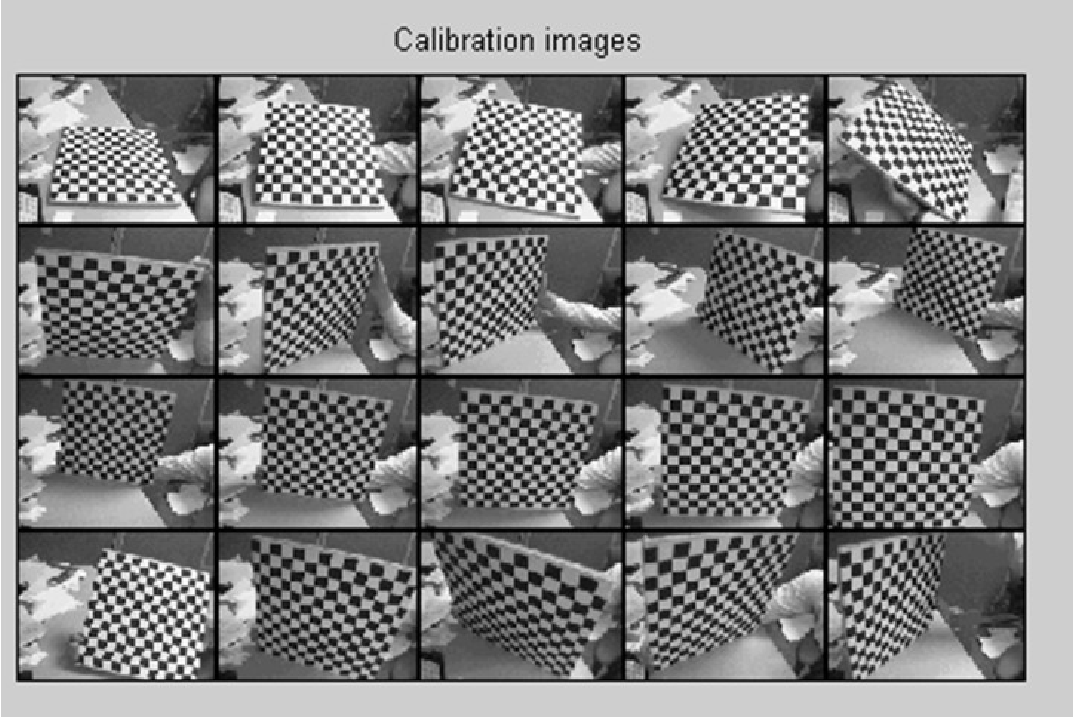
$$\mathrm{H}^{\mathsf{T}}\boldsymbol{\omega}\mathrm{H} = \begin{bmatrix} \lambda & 0 & ? \\ 0 & \lambda & ? \\ ? & ? & ? \end{bmatrix}$$

  - De cada plano, pues, podemos obtener 2 restricciones lineales sobre ***w:*** una igualando el elemento (1,2) de H^{T}wH a cero, y la otra igualando los elementos 1,1 y (2,2). Juntando las necesarias (según los gdl de ***w*** para nuestras suposiciones de cámara), podremos obtener nuestra *K*.

Autocalibración mediante planos (II)

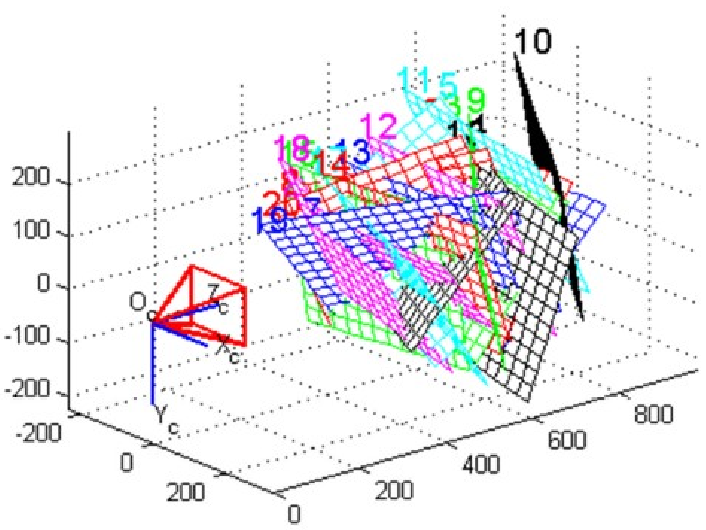
- Ejemplo típico de calibración de cámara (p.e., aplicación disponible en la OpenCV).
- Se mueve un plano con un patrón de calibración delante de la cámara, y se toman varias imágenes:

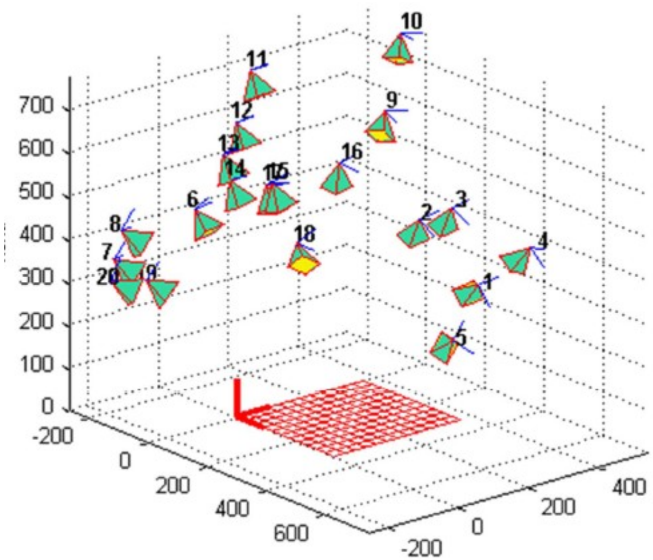


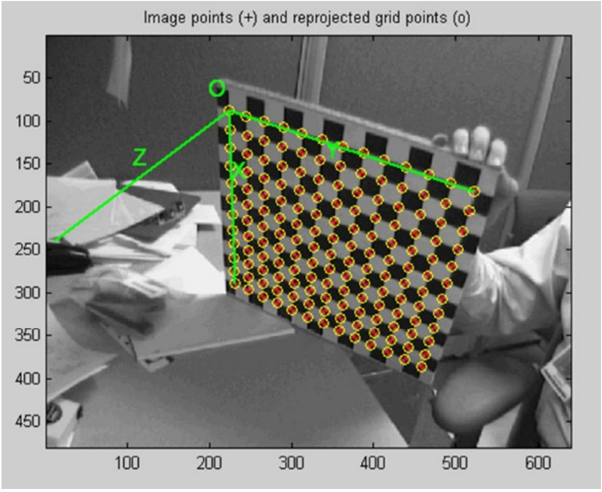


Autocalibración mediante planos (III)

- Con dichas homografías, se obtiene primero la *K* a través de las ecuaciones obtenidas para ***w***.
- Naturalmente, después se puede recuperar la R y la C de la cámara respecto al plano en cada imagen, y reproyectar el patrón y/o hacer realidad aumentada:







Autocalibración mediante rotaciones

- Segunda posibilidad: usar la **homografía inducida entre dos imágenes por una rotación de cámara**:
  - Si recordamos la forma de dicha *H* (tema anterior):

$$\mathrm{H} = \mathrm{K}\mathrm{R}\mathrm{K}^{-1}$$

  - Podemos volver a usar $\omega = \mathrm{K}^{-\mathsf{T}}\mathrm{K}^{-1}$ para obtener ahora ecuaciones de la forma $\mathrm{H}^{\mathsf{T}}\omega\mathrm{H} = \omega$:

$$\mathrm{H}^{\mathsf{T}}\omega\mathrm{H} = \mathrm{K}^{-\mathsf{T}}\mathrm{R}^{\mathsf{T}}\mathrm{K}^{\mathsf{T}}\ \mathrm{K}^{-\mathsf{T}}\mathrm{K}^{-1}\ \mathrm{K}\mathrm{R}\mathrm{K}^{-1} = \mathrm{K}^{-\mathsf{T}}\mathrm{K}^{-1} = \omega$$

  - De nuevo, con suficientes de ellas podremos despejar el número de incógnitas necesario para la forma de ***w*** de nuestra cámara, y a partir de ahí obtener la *K* por Cholesky.
  - Nota: podemos forzar *det(H)=1* para “deshomoge-neizar” la igualdad.

Autocalibración mediante fundamentales

- También está la <u>tercera posibilidad</u> de usar la forma estudiada para la matriz fundamental, $\mathtt{F} = [e_2]_\times \mathtt{KRK}^{-1}$:

$$\mathtt{F}\boldsymbol{\omega}^*\mathtt{F}^\mathsf{T} = [e_2]_\times \mathtt{KRK}^{-1}\mathtt{KK}^\mathsf{T}\mathtt{K}^{-\mathsf{T}}\mathtt{R}^\mathsf{T}\mathtt{K}^\mathsf{T}[e_2]_\times^\mathsf{T} = [e_2]_\times \boldsymbol{\omega}^*[e_2]_\times^\mathsf{T}$$

- Sin embargo, lamentablemente aquí no es posible eliminar la homogeneidad de forma sencilla:

$$\mathtt{F}\boldsymbol{\omega}^*\mathtt{F}^\mathsf{T} = \lambda[e_2]_\times \boldsymbol{\omega}^*[e_2]_\times$$

- … y eso nos lleva a ecuaciones NO lineales, no tan simples de resolver (**ecuaciones de Kruppa**).

Autocalibración mediante ecualización de SVD de E

- Una posibilidad alternativa, bastante sencilla, consiste en suponer una cámara "diagonal":

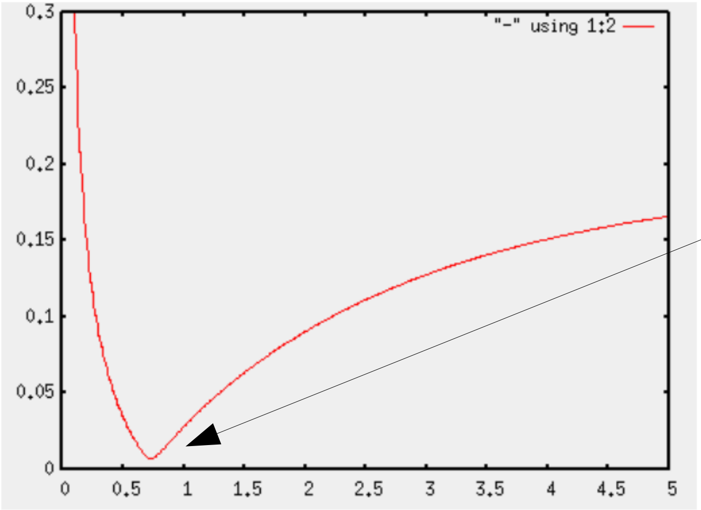
$$\mathrm{K}_f = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Y explotar el hecho de que la matriz esencial $\mathrm{E} = [\boldsymbol{t}]_\times \mathrm{R}$ tiene **IGUALES** sus dos valores singulares no nulos.
- Entonces, variando *f*, podemos calcular para cada matriz $E=K^T F K$ resultante, un "*factor de discrepancia*" entre sus dos primeros valores singulares, s_1 y s_2:

$$d = \frac{s_1 - s_2}{s_1 + s_2}$$

Calibración mediante ecualización de SVD de E

- Sólo queda buscar el mínimo de dicho factor de discrepancia al variar la *f*:



$$f = 0.73$$

- Estos son los valores de *F* y la *E* para este ejemplo:

$$\mathtt{F} = \begin{bmatrix} 0.208 & 0.113 & -0.337 \\ 0.144 & -0.128 & -0.574 \\ 0.366 & 0.573 & 0.043 \end{bmatrix} \qquad \mathtt{E} = \begin{bmatrix} 0.155 & 0.082 & -0.346 \\ 0.107 & -0.094 & -0.591 \\ 0.371 & 0.582 & 0.061 \end{bmatrix}$$

s1=0.727, s2=0.687 *s1=0.7073, s2=0.7069*

Extracción de cámaras

- Una vez que se conocen *K*, *K'* (autocalibración) y *F* (alg. 8 puntos), el procedimiento para extraer la *R* y la *C* entre ellas es cerrado (demostración en HZ):
  - Calculamos $E=K^{T}FK'$
  - Hacemos su SVD: $(\mathtt{U}, \_, \mathtt{V}) = svd(\mathtt{E})$
  - Y definiendo

$$\mathtt{W} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

  - Obtenemos ambas cámaras (incluso debidamente factorizadas, e.d. sabiendo sus extrínsecos e intrínsecos):

Rotación R

$$\mathtt{M} = \mathtt{K} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \mathtt{M}' = \mathtt{K}'[\mathtt{UAV}^{\mathsf{T}} \mid \pm \mathtt{U}_3]$$

C = espacio nulo de M'

A = W o W^T, y $\pm U_3$ → **<u>4 posibles M'</u>**

Extracción de cámaras

- 4 posibles M' (Hartley & Zisserman, 2003):

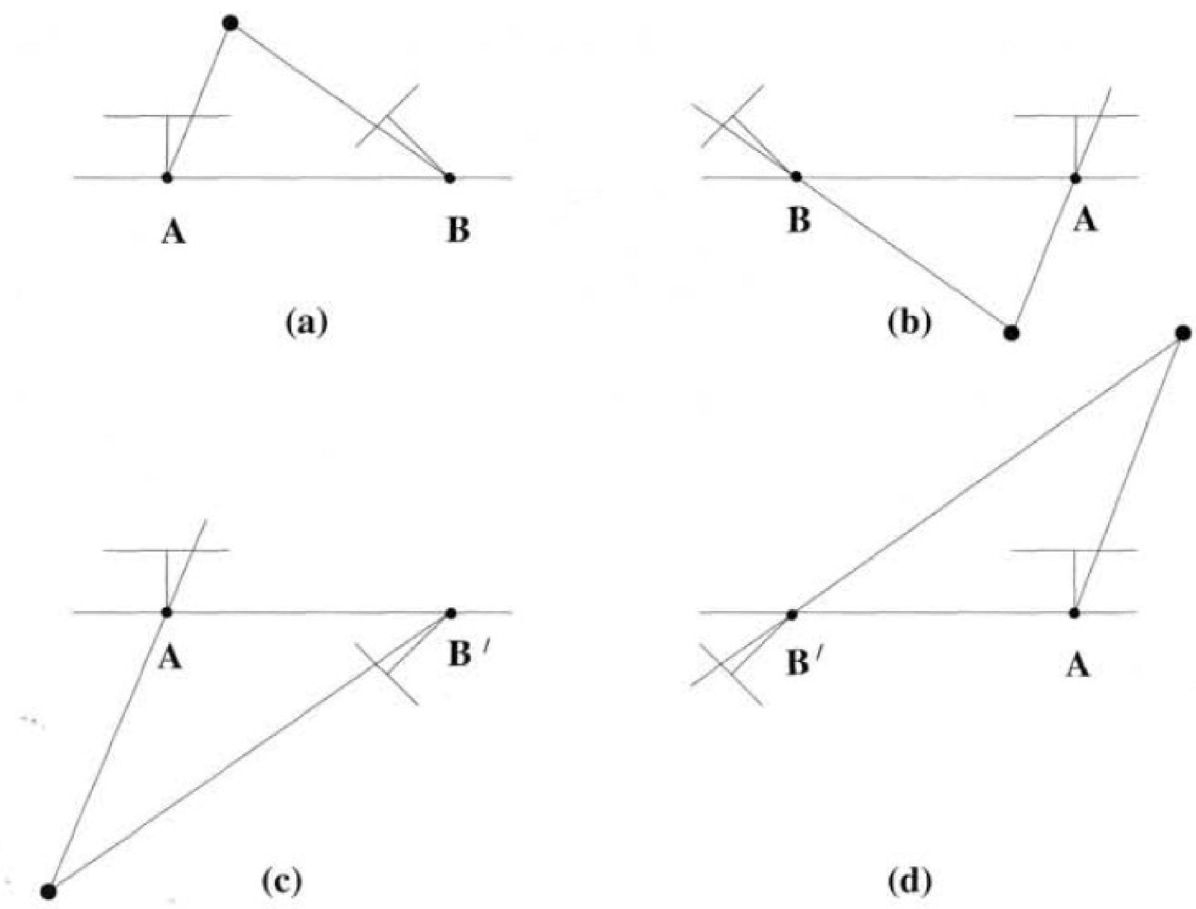
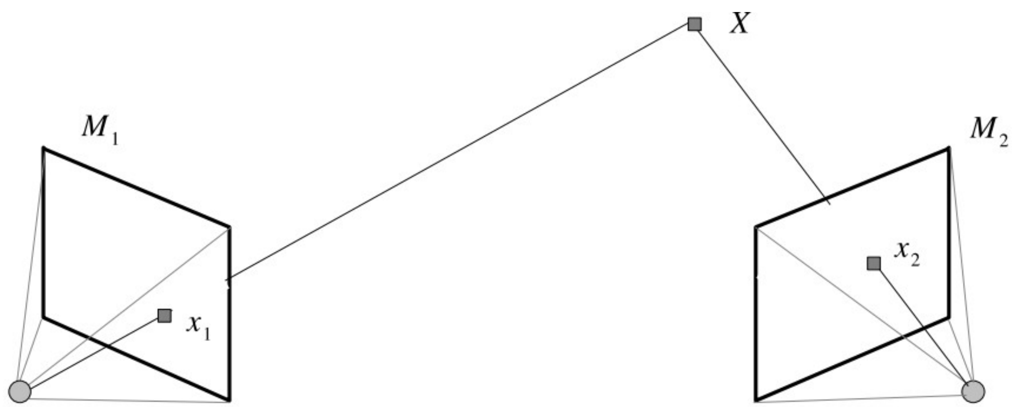


Fig. 9.12. **The four possible solutions for calibrated reconstruction from E.** *Between the left and right sides there is a baseline reversal. Between the top and bottom rows camera B rotates 180° about the baseline. Note, only in (a) is the reconstructed point in front of both cameras.*

Triangulación (I)

- Una vez conocida la posición relativa entre las cámaras, sólo queda recuperar la posición de un punto dadas sus proyecciones en las imágenes:



$$\boldsymbol{x}_1 = \mathtt{M}_1 \boldsymbol{X}$$

$$\boldsymbol{x}_2 = \mathtt{M}_2 \boldsymbol{X}$$

- Conocidos los coeficientes de M_1, M_2, x_1 y x_2, se puede plantear fácilmente un sistema lineal para determinar X (al estilo de como se hacía para la estimación de una homografía *H*). Lo veremos a continuación.

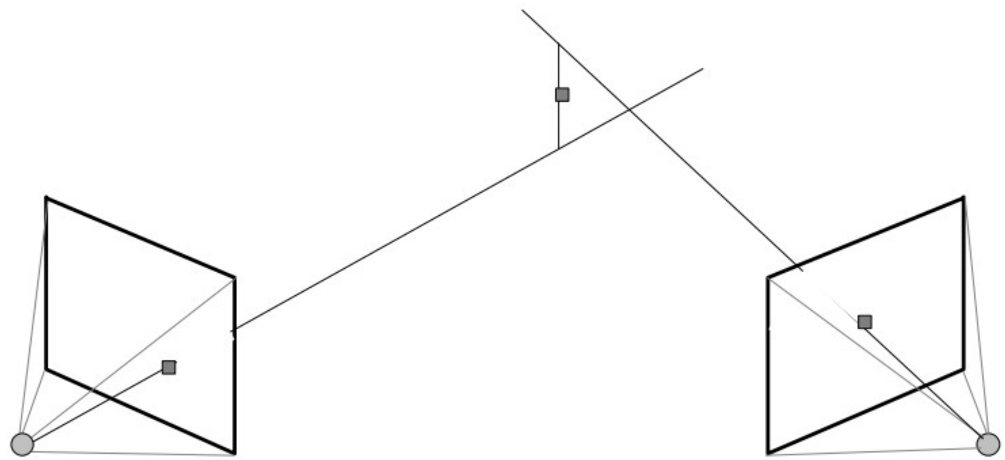
Triangulación (II)

- Por ambas igualdades homogéneas, tenemos que, para cada punto $x_1 \times M_1 X = 0$ y $x_2 \times M_2 X = 0$.
- Cada ecuación vectorial anterior se corresponde con 3 ecuaciones, sólo 2 de ellas independientes. Eso hace un total de 4 ecuaciones independientes por punto, suficiente para las 3 incógnitas (posición 3D del punto).
- La matriz de coeficientes que sale es la siguiente, donde m^i es la fila i de la matriz M_1, ídem para m'^i de M_2, y $x_1=(x,y,1)$ y $x_2=(x',y',1)$

$$A_{4\times 4} = \begin{pmatrix} x\mathbf{m}^3 - \mathbf{m}^1 \\ y\mathbf{m}^3 - \mathbf{m}^2 \\ x'\mathbf{m}'^3 - \mathbf{m}'^1 \\ y'\mathbf{m}'^3 - \mathbf{m}'^2 \end{pmatrix}$$

Triangulación (III)

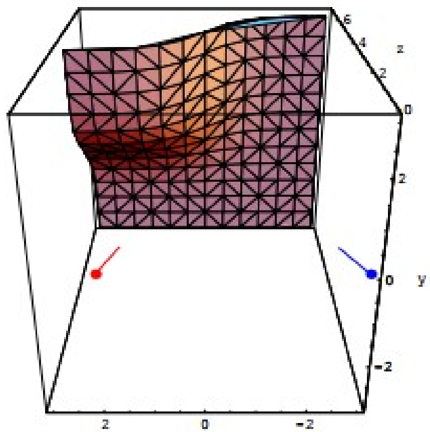
- Con ligero *ruido* de píxeles, en la triangulación no tienen por qué coincidir exactamente los rayos "retroproyectados":



- → Solución por **mínimos cuadrados** al sistema $A.(X,Y,Z,W)^T=0$. Es decir, de nuevo nos quedamos con el vector correspondiente al menor valor singular de A (o vector propio de A^TA). *De todos modos,* el error minimizado es "algebraico", no estrictamente "geométrico").
- (X/W,Y/W,Z/W) es nuestro punto 3D triangulado.

Triangulación (IV)

- Ejemplo:



Superficie original y posición de las cámaras:

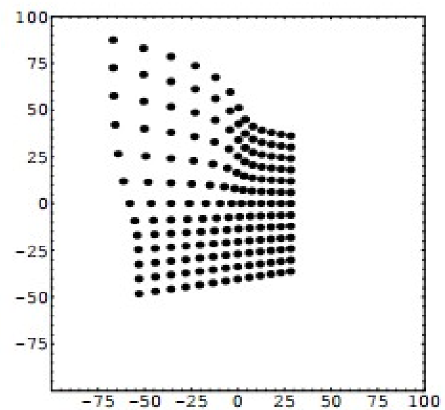


Imagen izquierda

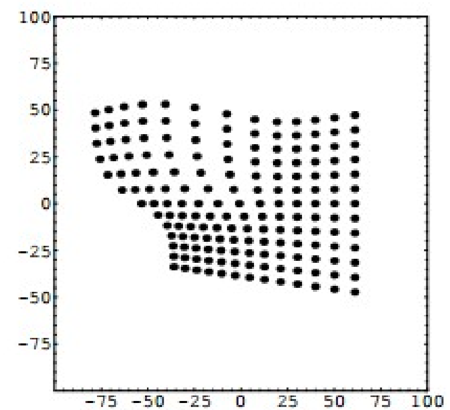
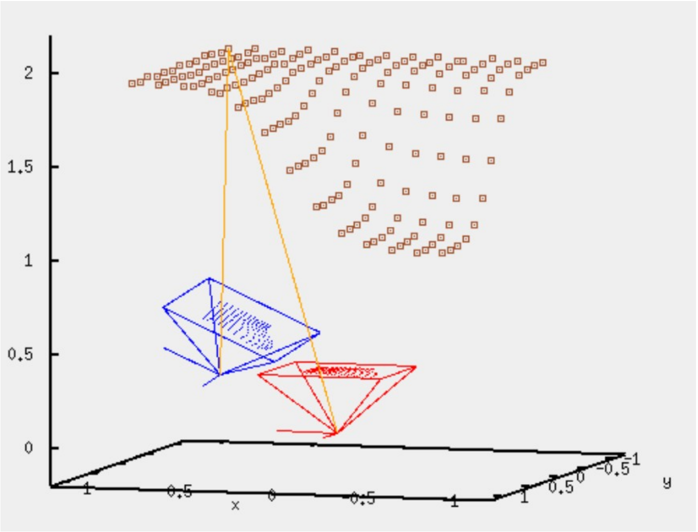


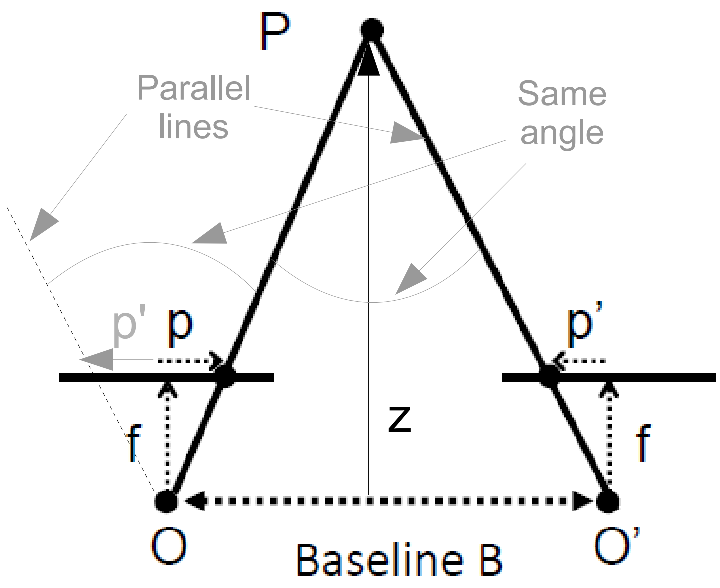
Imagen derecha

Reconstrucción obtenida tras la obtención de las cámaras y la posterior triangulación

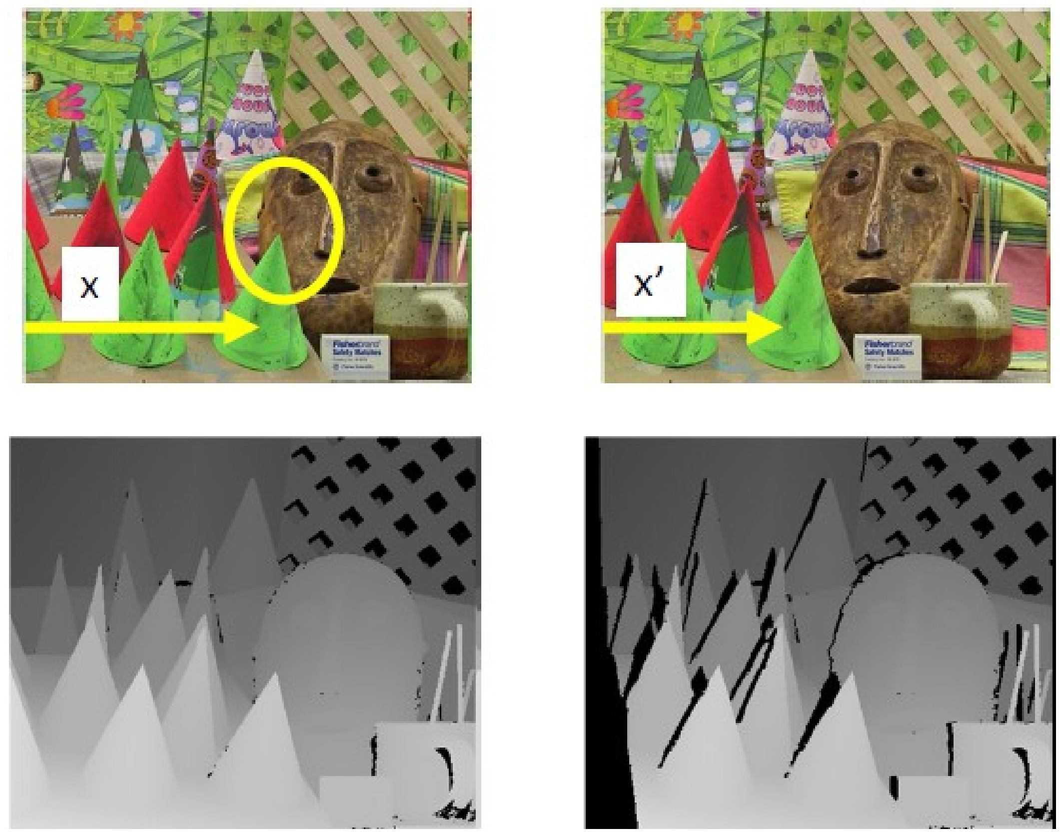


Reconstrucciones densas

- **Disparidad**:
  - Inversamente proporcional a *profundidad*.
  - Directamente proporcional a la focal y la distancia entre cámaras:

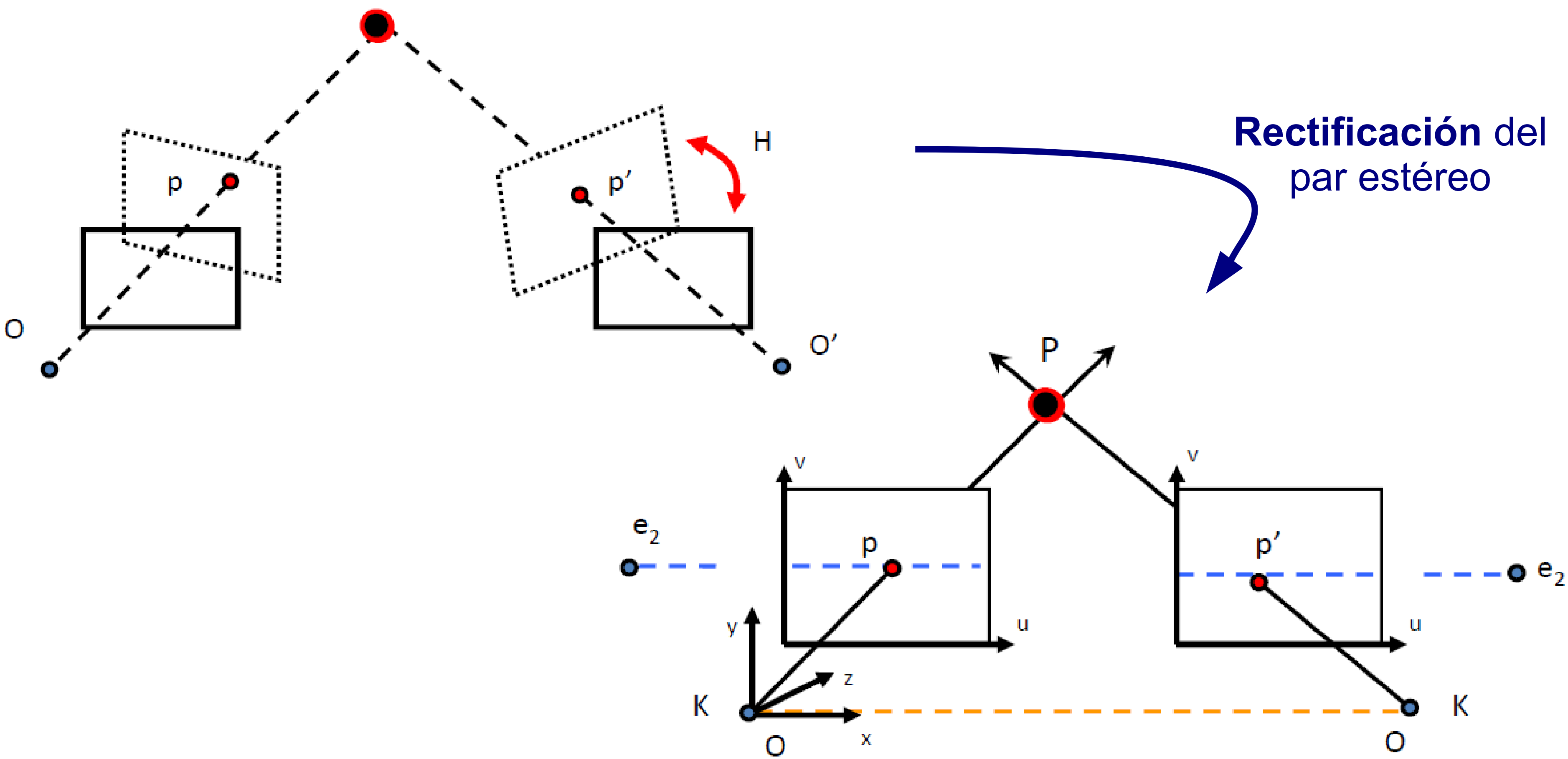


$$p - p' = \frac{B \cdot f}{z} \text{ = disparity}$$



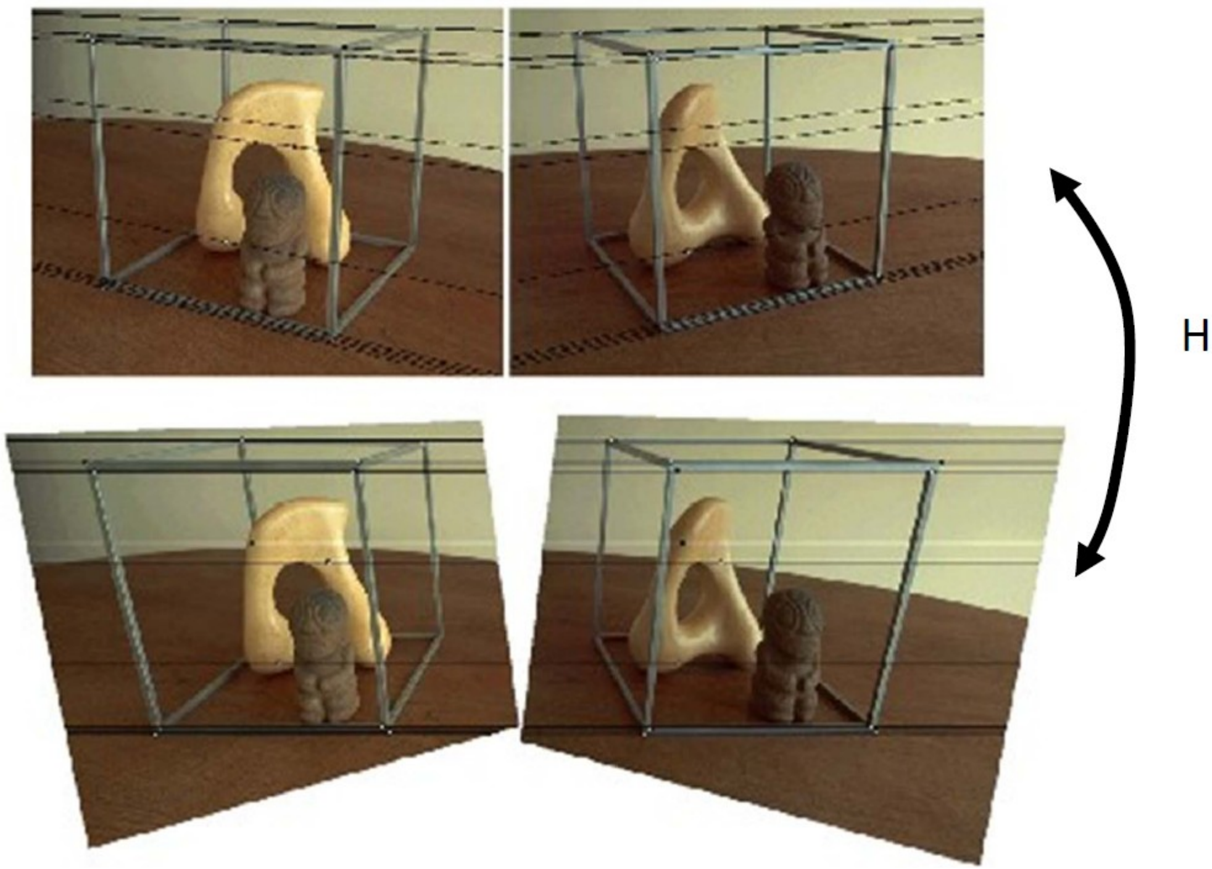
Rectificación

- Proceso previo para hacer las imágenes del par estéreo “paralelas”:
  - Líneas epipolares paralelas (epipolos en el infinito)



Ejemplo de rectificación de par estéreo

- Observar que, tras la rectificación, las correspondientes líneas epipolares son horizontales y están alineadas en ambas imágenes → más fácil hacer *matching* y calcular disparidad:

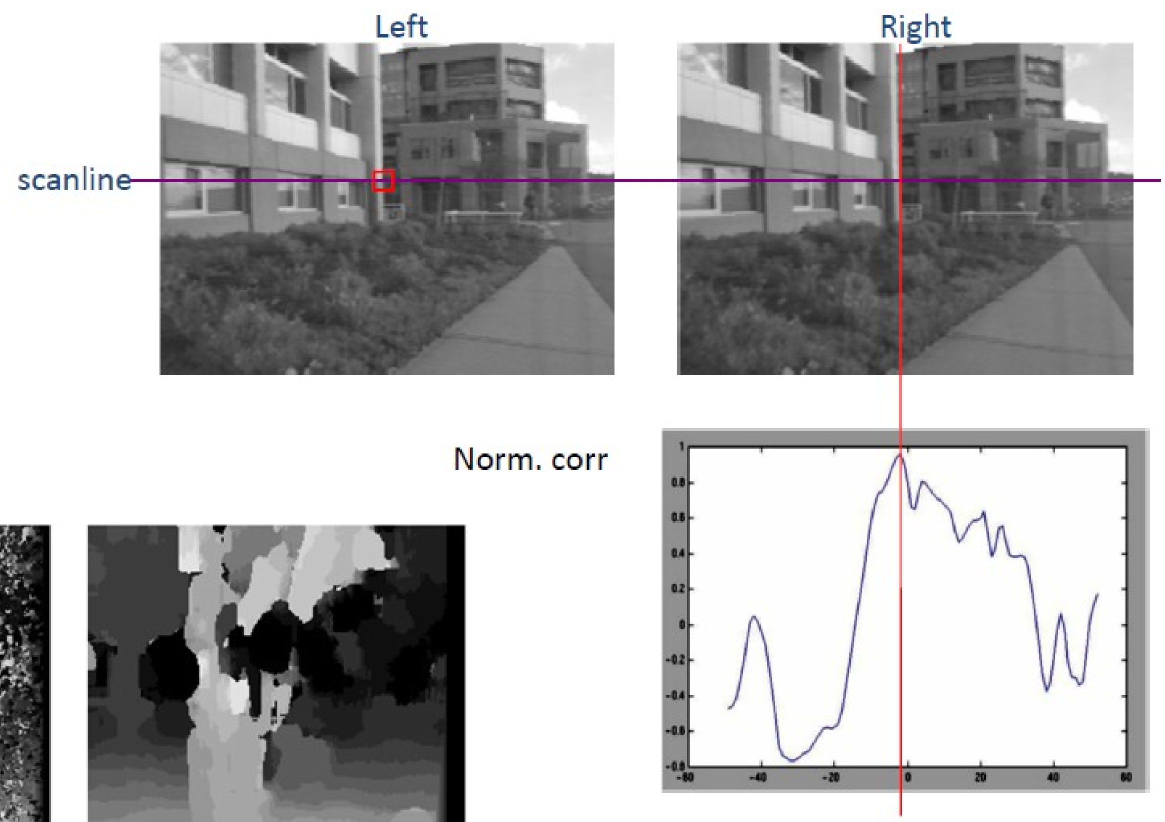


Rectificación par estéreo

- **Pasos a realizar** (a grandes rasgos):
  - Computar epipolos.
  - Girar ambas imágenes para “enviarlos”, primero, al eje X.
  - “Enviar” a continuación ambos epipolos al infinito.
  - Finalmente, encontrar también una transformación que ponga en la misma línea (=Y) líneas epipolares correspondientes, y que en el eje X minimice la disparidad.
  - Al final se obtienen sendas homografías H_1 y H_2 que hacen el trabajo.
  - Detalles de cómputo concreto en sección 11.12 de [HZ].

Método de correlación (I)

- Calcular correlación al desplazar una ventana de píxeles a lo largo de toda la correspondiente línea epipolar.
- Influye tamaño de ventana (más pequeña → más detalle, pero más ruido).



Método de correlación (II)

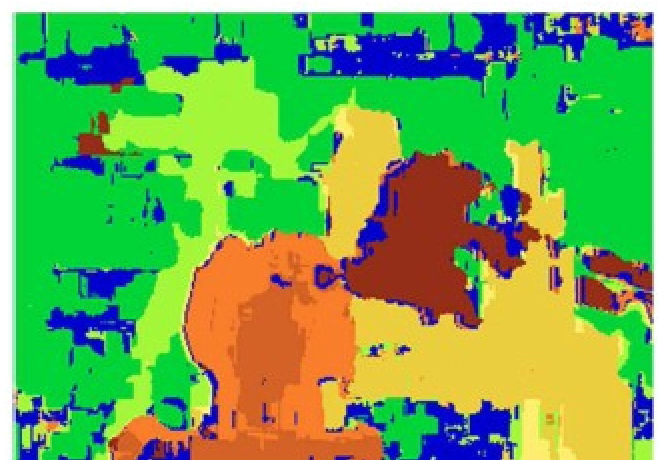
- Problemas con regiones homogéneas, o patrones repetitivos:

Método de correlación (III)

- Ejemplo de resultados obtenidos con correlación simple:

Imagen original (izquierda)

Mapa de profundidad "*ground truth*"



Mapa de profundidad obtenido

Múltiples vistas (I)

- El *paso final*: las múltiples vistas...
  - Proceso iterativo:
    - Con un par de vistas, reconstrucción inicial (dos cámaras + conjunto puntos 3D).
    - Con una tercera vista, reseccionar la cámara correspondiente, y seguramente añadir puntos 3D adicionales.
    - Iterar de forma similar con más vistas.
  - Finalmente, *bundle adjustment* (BA, o “ajuste de rayos”):
    - Minimización (no lineal) de una función que depende de todas las cámaras y los puntos 3D...

Múltiples vistas (II)

- *Bundle adjustment* (ajuste de rayos):
  - Minimización (no lineal) de la función “suma de todos los errores de reproyección”, que depende de todas las cámaras y los puntos 3D:

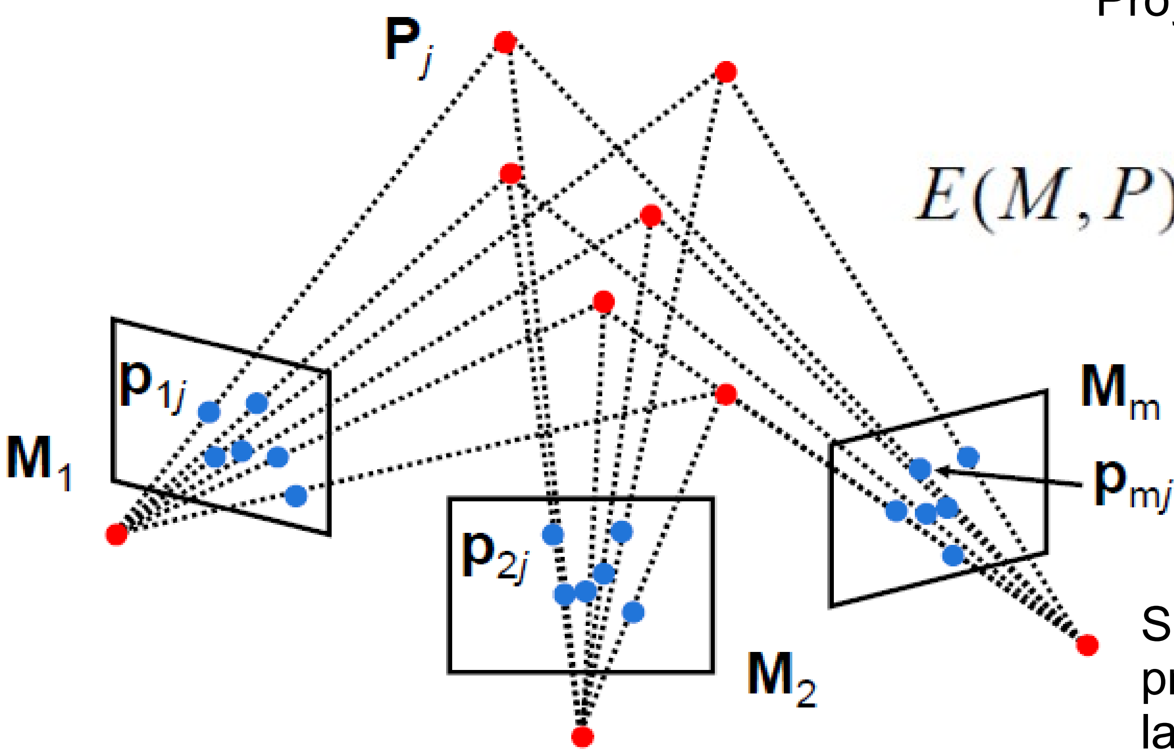
Cámaras

Puntos 3D

Proyecciones 2D

$$E(M,P)=\sum_{i=1}^{m}\sum_{j=1}^{n}\underbrace{D\left(p_{ij}, M_i P_j\right)^2}$$

Distancia proyecciones 3D→2D y puntos 2D detectados originales



Suma para todas las proyecciones en todas las cámaras.

Algunos vídeos ilustrativos

- Structure from motion pipeline:
  - https://www.youtube.com/watch?v=i7ierVkXYa8
- Proyecto LOCUM AR:
  - https://www.youtube.com/watch?v=qycdlROrtXE
- Structure from motion revisited:
  - https://www.youtube.com/watch?v=Gb086k7b0wg
- Software abierto: Visual SfM + CMVS:
  - https://www.youtube.com/watch?v=wBKidr0e-XA
- VisualSfM & Meshlab workflow (practical example):
  - https://www.youtube.com/watch?v=GEAbXYDzUjU

Bibliografía básica

- "*Multiple view geometry*" (2nd ed.), R. Hartley, A. Zisserman (2003), Cambridge University Press.

- "*Tutorial on 3D Modeling from Images*", M. Pollefeys (2000).

- "*Apuntes de Sistemas de Percepción y Visión por Computador*", A. Ruiz (2015).

  (http://dis.um.es/profesores/alberto/material/percep.pdf)