

# STRUCTURE FROM MOTION – VISIÓN ESTÉREO

---

ANDRÉS DANIEL GODOY ORTIZ

**VISIÓN COMPUTACIONAL**

IMÁGENES DE PROF. ANDREAS GEIGER, UNIVERSITY OF TÜBINGEN



# STRUCTURE FROM MOTION

---

Recuperar la estructura 3D de una escena y la posición de la cámara a partir de múltiples imágenes tomadas desde diferentes ángulos.

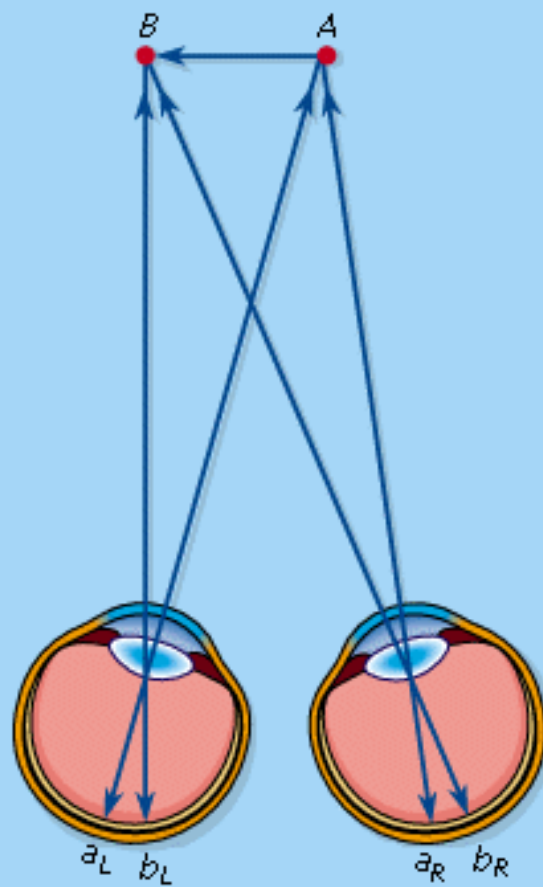


# ¿QUÉ PASOS DEBO REALIZAR PARA UNA RECONSTRUCCIÓN 3D?

---

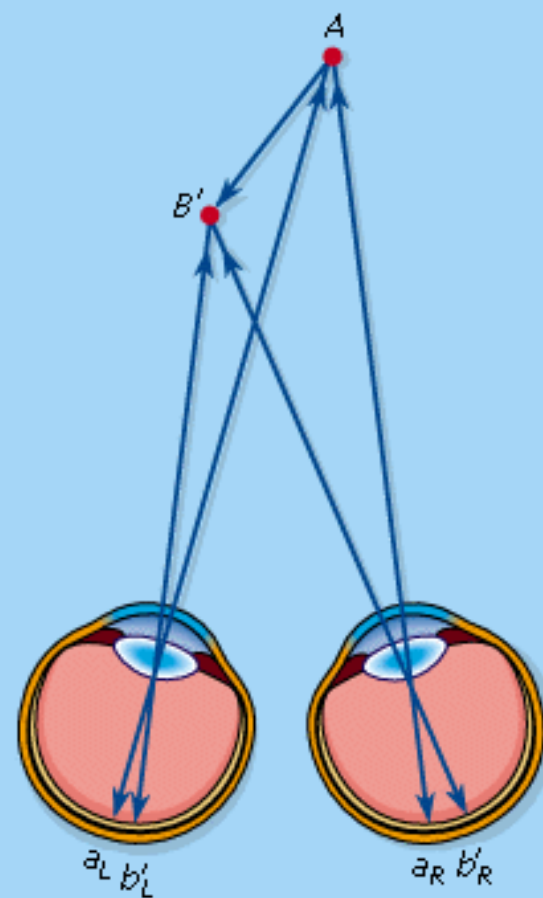
SFM: UNA SOLA CÁMARA EN MOVIMIENTO O VARIAS CÁMARAS SIN ALINEACIÓN  
FIJA (DRONES)





corresponding points

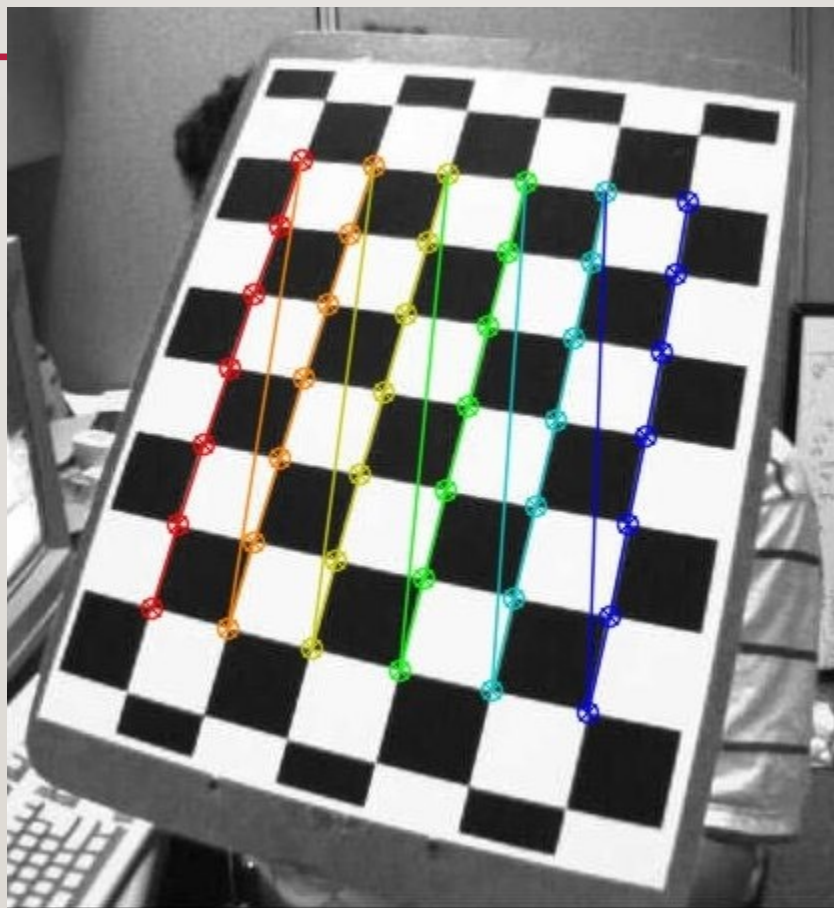
©1994 Encyclopaedia Britannica, Inc.



noncorresponding points

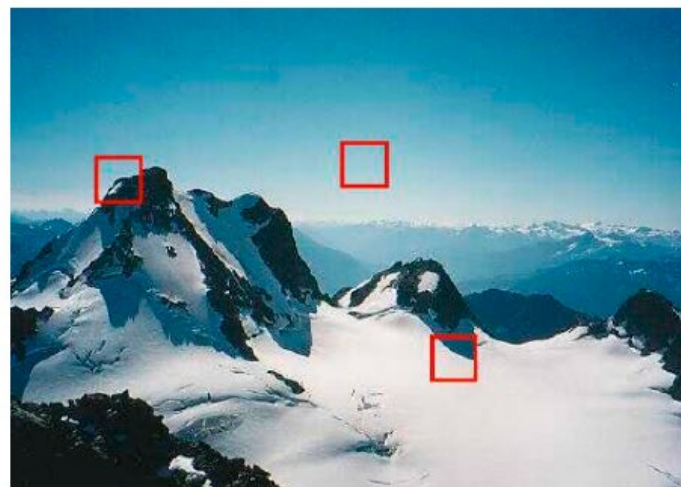
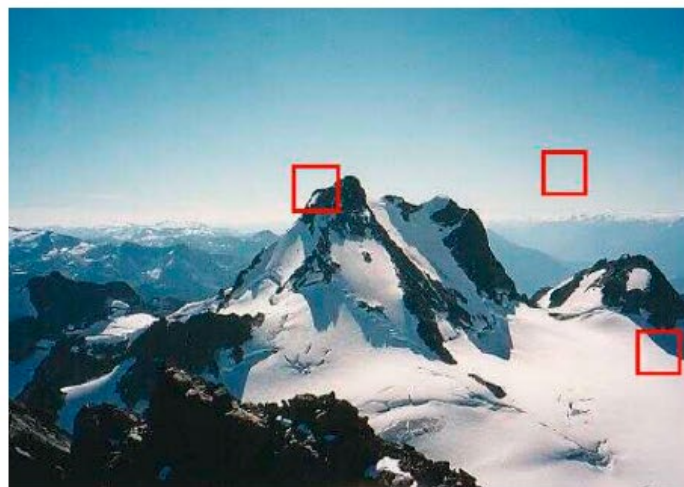


# PRIMERO HAY QUE CALIBRAR LA CÁMARA



# LUEGO HAY QUE EXTRAER CARACTERÍSTICAS

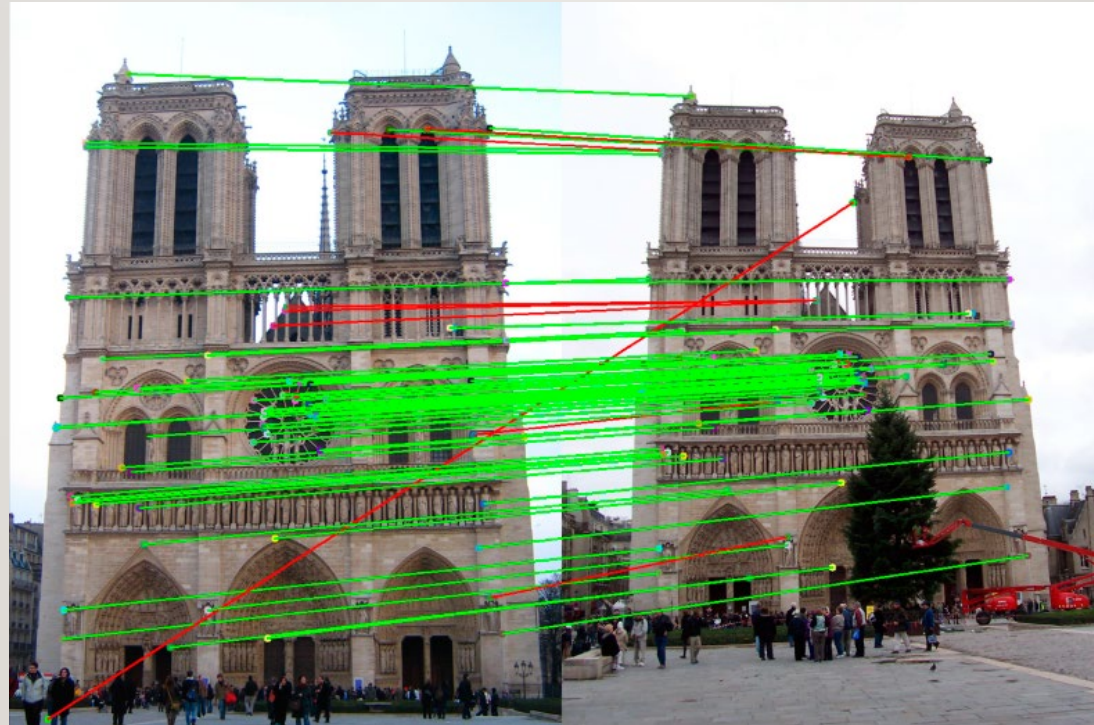
---



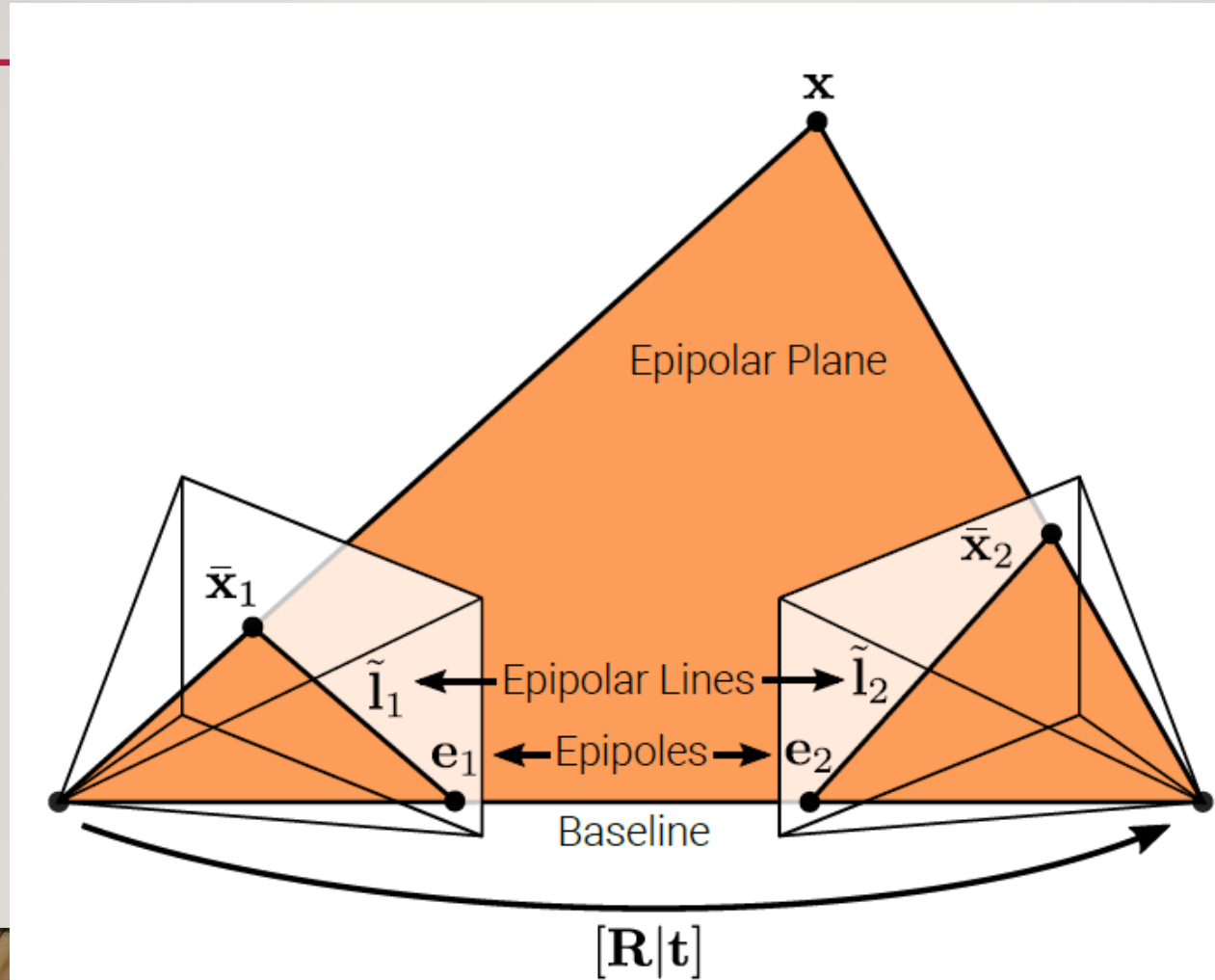


# SCALE INVARIANT FEATURE TRANSFORM (SIFT)

---



# GEOMETRÍA EPIPOLAR





Sabes que una cámara se modela con la ecuación de proyección:

$$\mathbf{x} = K[R|t]\mathbf{X}$$

donde:

- $\mathbf{X} = (X, Y, Z, 1)^T$  es un punto 3D en coordenadas homogéneas.
- $\mathbf{x} = (x, y, 1)^T$  es la proyección en la imagen (coordenadas homogéneas).
- $K$  es la matriz de calibración de la cámara.
- $R$  y  $t$  representan la orientación (rotación  $R$ ) y la posición (traslación  $t$ ) de la cámara en el espacio.

Expandiéndolo, tenemos:

$$\mathbf{x} \sim KR\mathbf{X} + Kt$$

y en coordenadas cartesianas:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} fX + cxZ \\ fY + cyZ \end{bmatrix}$$

donde  $f$  es la distancia focal y  $(cx, cy)$  son los parámetros de desplazamiento óptico.

Dado un punto  $X$  en el espacio tridimensional, su proyección en dos imágenes tomadas por dos cámaras es:

$$x_1 = K_1[R_1|t_1]X$$

$$x_2 = K_2[R_2|t_2]X$$

Para simplificar, trabajamos con **cámaras calibradas** y tomamos la primera cámara en el sistema de referencia mundial, es decir,  $R_1 = I$  y  $t_1 = 0$ . Esto nos da:

$$x_1 = KX$$

Para la segunda cámara, la matriz de rotación y traslación entre las cámaras está dada por  $R$  y  $t$ :

$$x_2 = K[R|t]X$$

Sea  $K$  la **matriz de calibración de la cámara**, y definimos la **dirección del rayo en coordenadas no calibradas** como:

$$\tilde{\mathbf{x}} = K^{-1} \mathbf{x}$$

Así mismo, se cumplirá:

$$\tilde{\mathbf{x}}_2 = R\tilde{\mathbf{x}}_1 + t$$

Tomamos el **producto cruzado** en ambos lados con  $t$

$$[t]_{\times} \tilde{\mathbf{x}}_2 = [t]_{\times} R\tilde{\mathbf{x}}_1$$



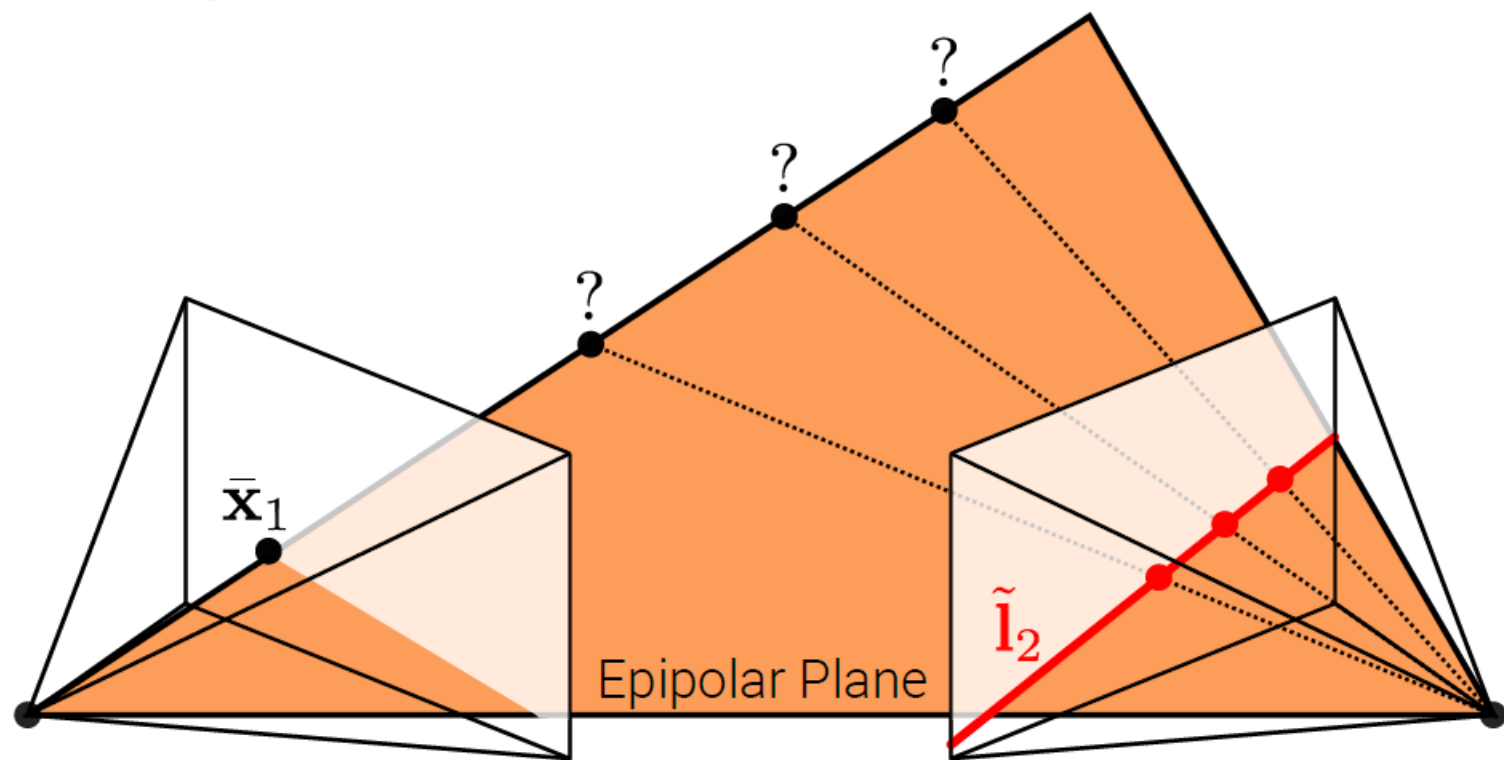
Finalmente, si multiplicamos por  $\tilde{\mathbf{x}}_2^T$ , obtenemos:

$$\tilde{\mathbf{x}}_2^T [t]_{\times} R \tilde{\mathbf{x}}_1 = 0$$

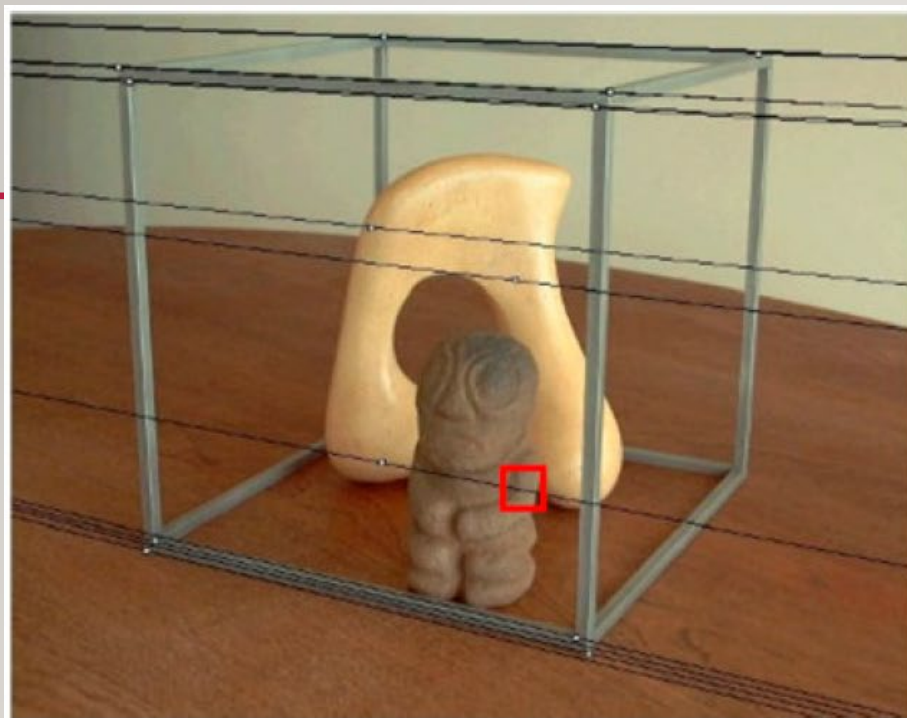
Restricción Epipolar, con matriz Esencial E:

$$\tilde{\mathbf{x}}_2^{\top} \tilde{\mathbf{E}} \tilde{\mathbf{x}}_1 = 0$$

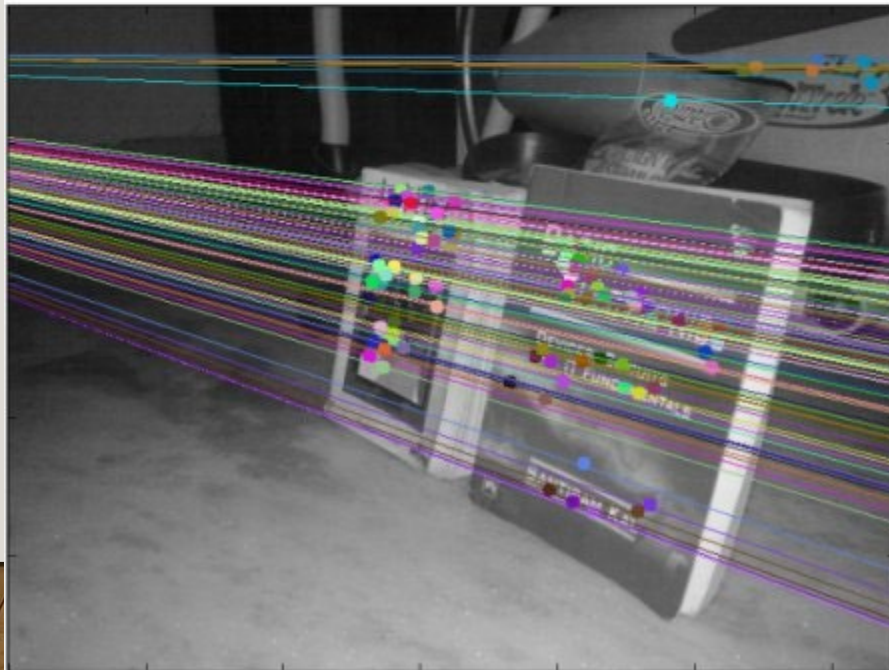
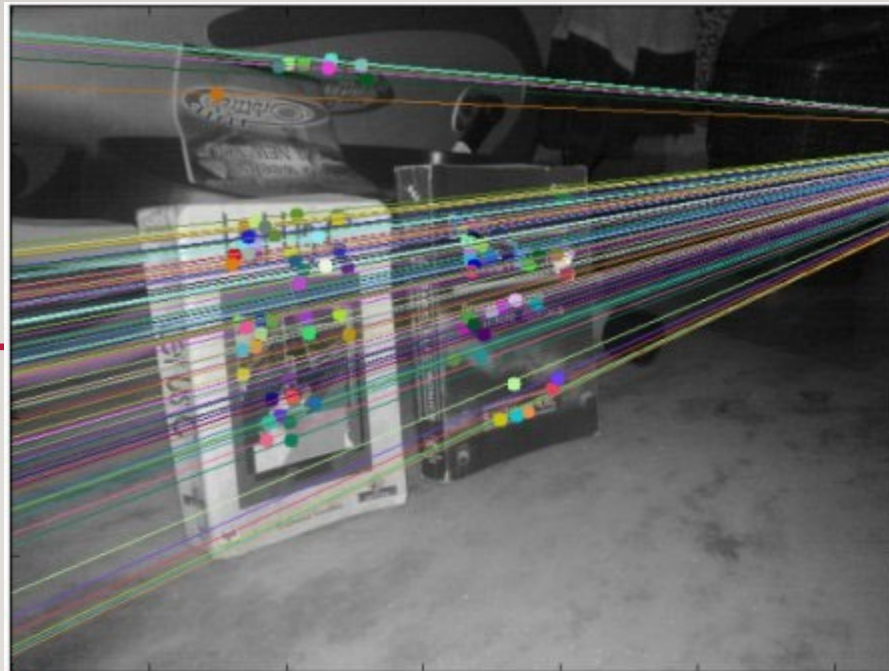
La matriz esencial se estima usando el **método de los ocho puntos** (Eight-Point Algorithm).



$$\tilde{\mathbf{l}}_2 = \tilde{\mathbf{E}} \bar{\mathbf{x}}_1$$







# Triangulación

Dado un punto observado en dos imágenes con matrices de proyección  $P_1$  y  $P_2$ :

$$P_1 = K[R_1|t_1], \quad P_2 = K[R_2|t_2]$$

y las coordenadas homogéneas  $x_1$  y  $x_2$ :

$$x_1 = P_1 X, \quad x_2 = P_2 X$$

Expandiendo:

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \propto P_1 X, \quad \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \propto P_2 X$$

Se resuelve como un **problema lineal** de la forma  $AX = 0$ , usando **SVD** para obtener  $X$ .

En la realidad, las imágenes contienen ruido, distorsión, errores de correspondencia. La triangulación sin optimización acumula errores y sesgos.

## Optimización (Bundle Adjustment)

$$\sum_{i,j} \|x_{ij} - \pi(P_j, X_i)\|^2$$

donde:

- $x_{ij}$  es la proyección observada.
- $\pi(P_j, X_i)$  es la proyección del punto  $X_i$  en la cámara  $j$ .

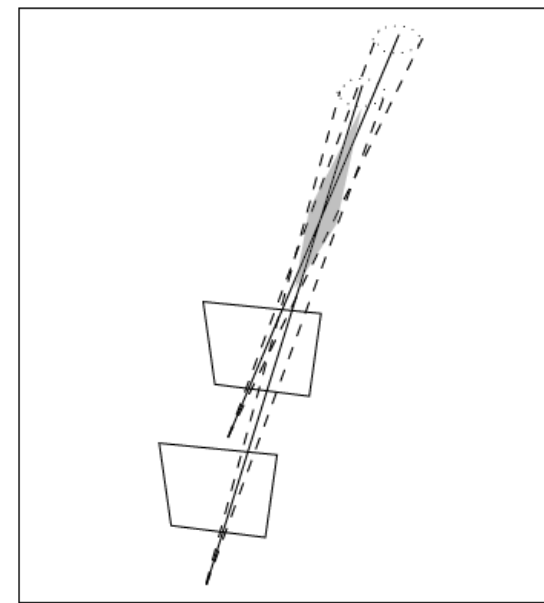
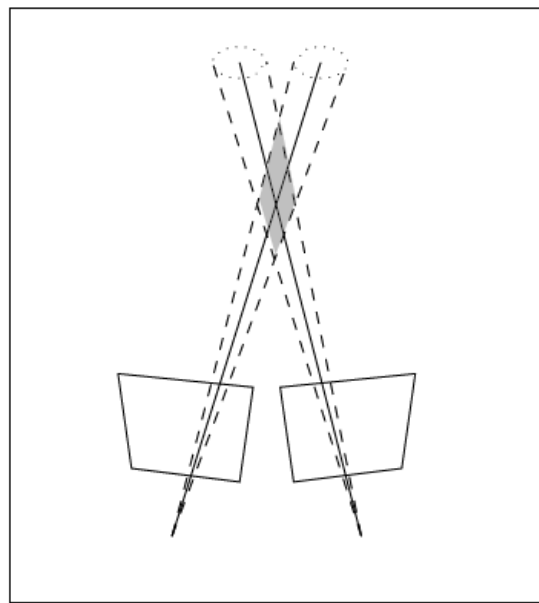
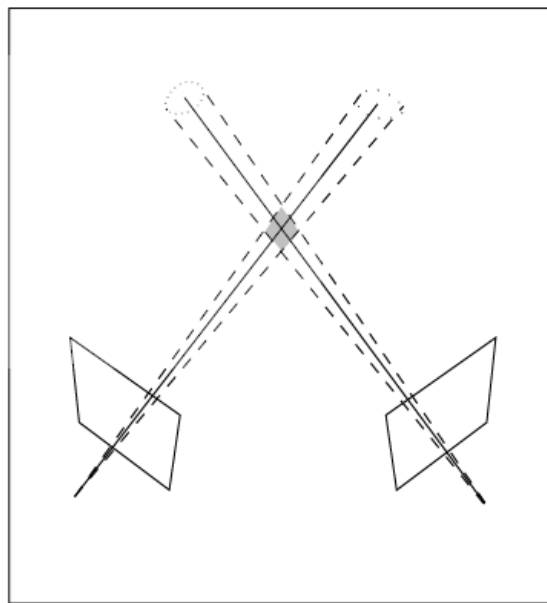
Se resuelve con **Levenberg-Marquardt** para refinar:

- Posiciones de cámara  $(R_j, t_j)$ .
- Puntos 3D  $X_i$ .



# LA INCERTIDUMBRE DE LA CÁMARA AUMENTA CUANDO....

---



# PASOS PARA RECONSTRUCCIÓN 3D

---

- **Detección de características** → Identificar puntos clave en múltiples imágenes.
- **Emparejamiento de características** → Asociar los mismos puntos en diferentes imágenes.
- **Estimación de movimiento** → Calcular la posición relativa de la cámara.
- **Triangulación** → Obtener la estructura 3D de la escena.
- **Optimización (BA - Bundle Adjustment)** → Refinar la estructura y la posición de la cámara.

# VISIÓN ESTÉREO

---

RECUPERAR PROFUNDIDAD A PARTIR DE DOS IMÁGENES OBTENIDAS  
SIMULTÁNEAMENTE POR CÁMARAS CALIBRADAS.





# RECUERDA ERES MEJOR QUE LEELA

---



PERO TAMPOCO EL/LA MÁS PRO...

---



# DISPARIDAD

---





# DISPARIDAD

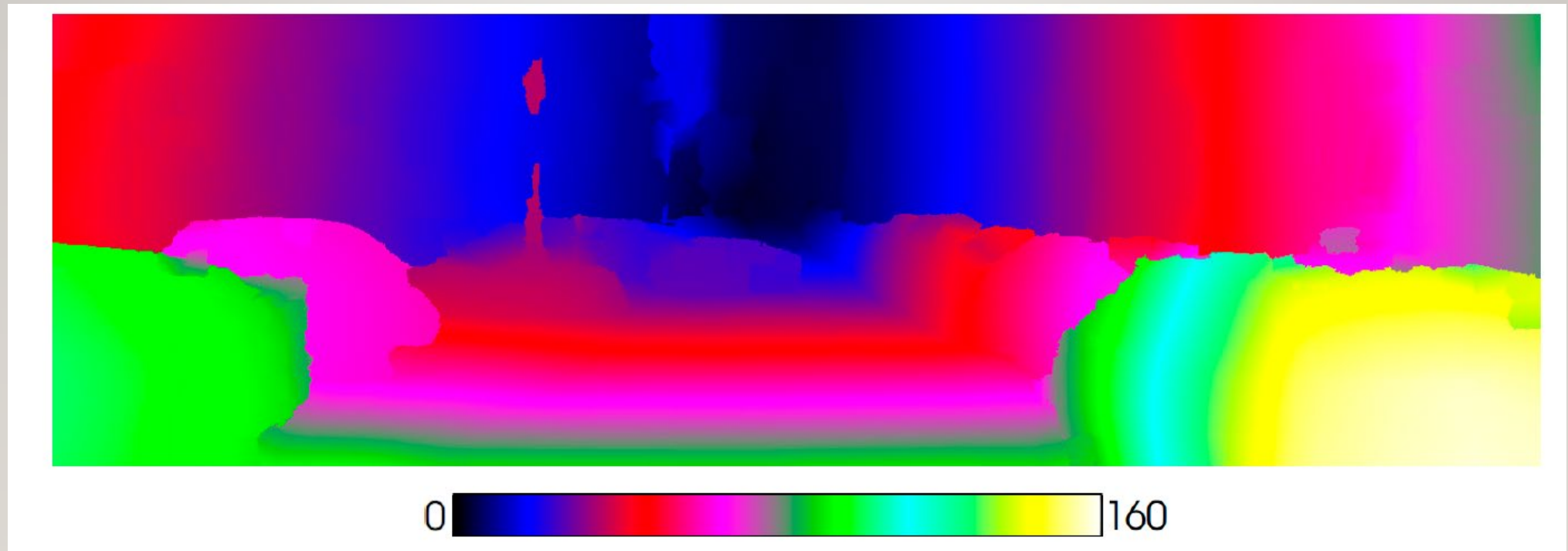
---





# DISPARIDAD

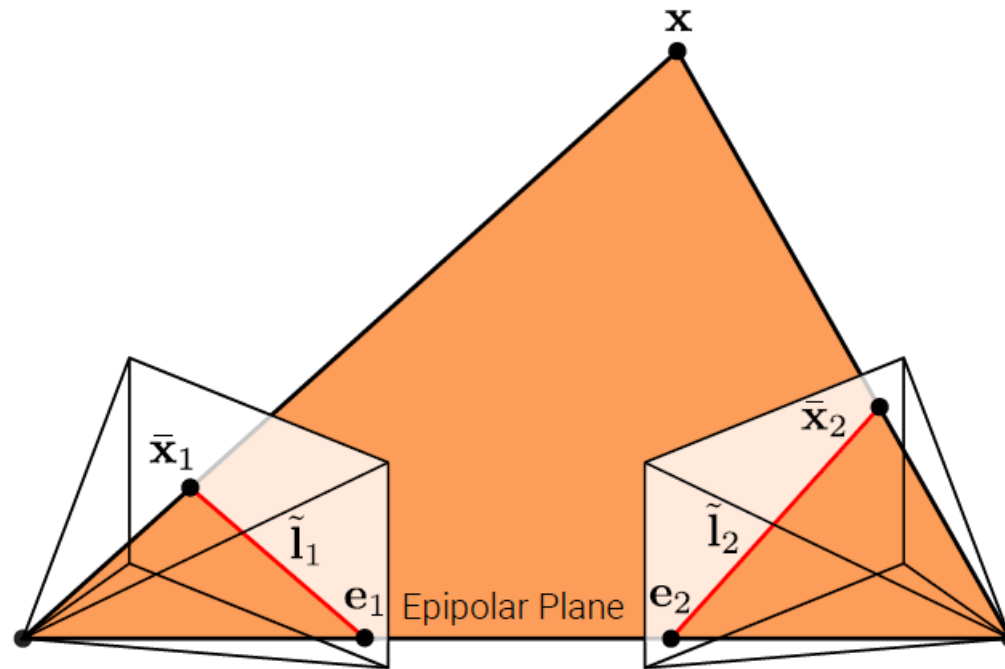
---



# LAS IMÁGENES DEBEN ESTAR SOBRE LA MISMA LÍNEA HORIZONTAL

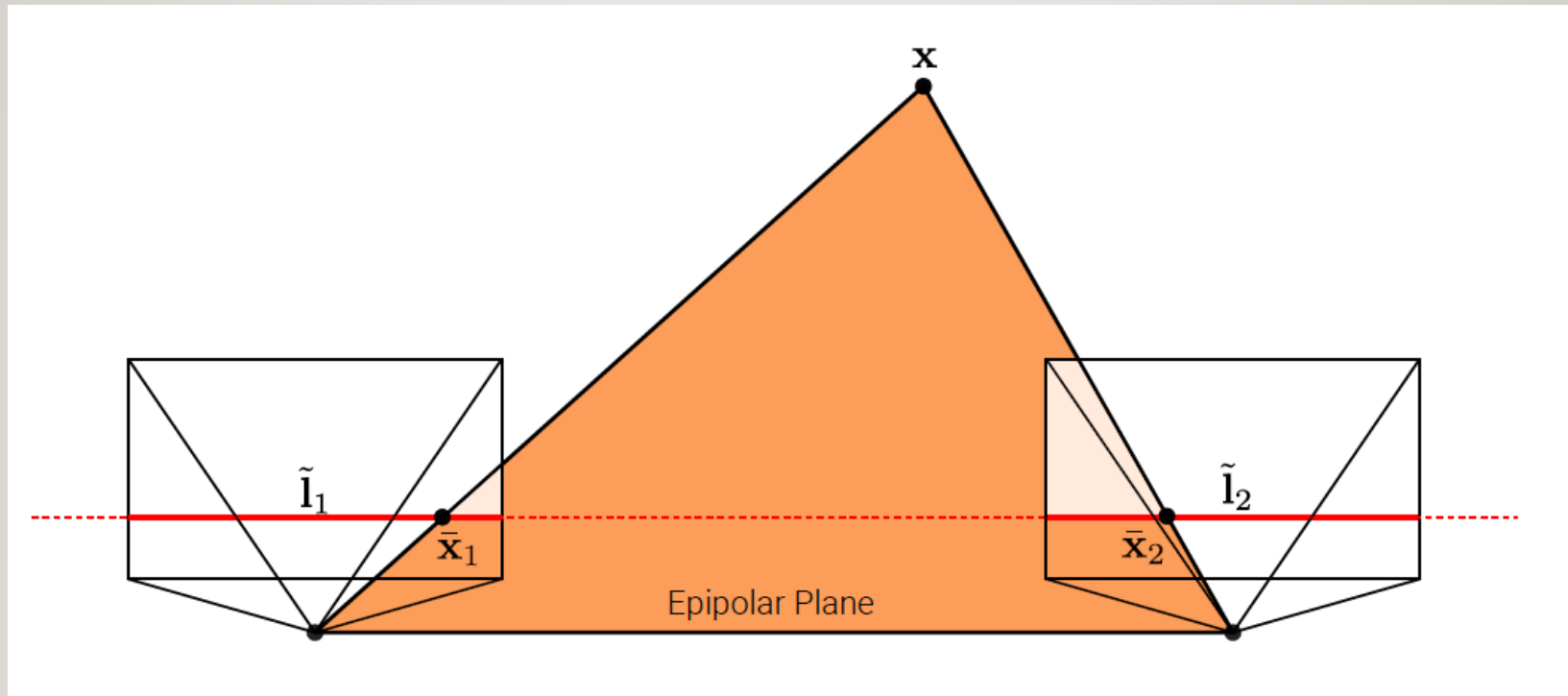
---

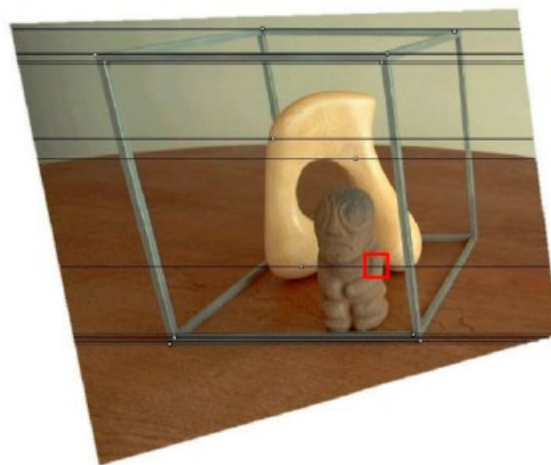
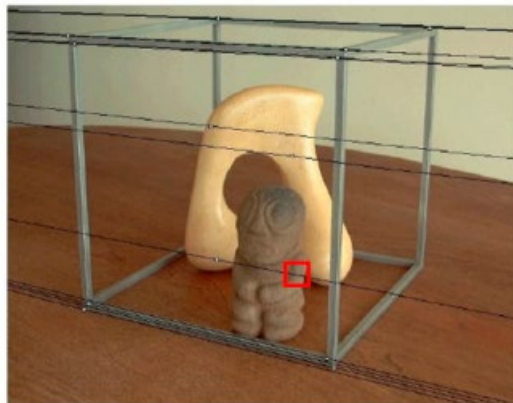
Esto no nos sirve...



# TOCA: IMAGE RECTIFICATION

---







# CORRESPONDENCIA DE CARACTERÍSTICAS

---

## Opciones:

- **Block Matching (BM)** → Compara ventanas de píxeles en ambas imágenes.
- **Semi-Global Matching (SGM)** → Más preciso pero más costoso computacionalmente.
- **Métodos basados en Deep Learning** → Como Redes Siameses, PSMNet, pero requieren entrenamiento.

- 
- **Calibración de cámaras** → Obtener matrices intrínsecas y distorsión.
  - **Rectificación estéreo** → Alinear imágenes para que los puntos correspondan en la misma línea horizontal.
  - **Correspondencia de puntos** → Comparar píxeles en la imagen izquierda y derecha.
  - **Calcular disparidad**
  - **Recuperar profundidad**