Visión Artificial

Tema 5: Visión estéreo

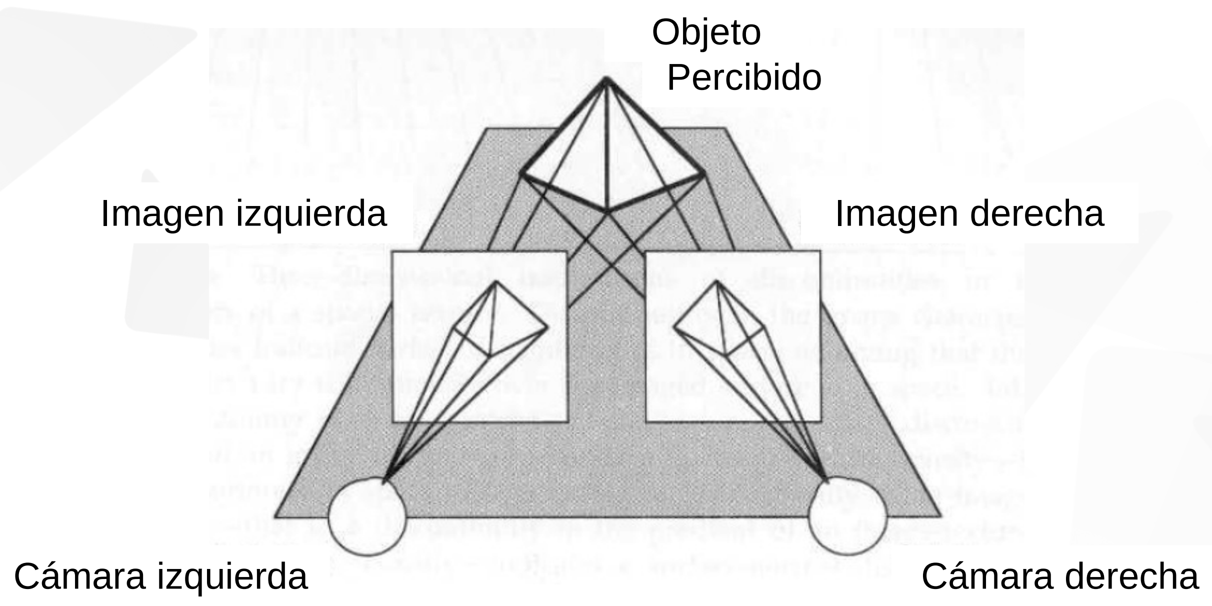
- Introducción a la visión estéreo
- El problema de la correspondencia
- Geometría epipolar
- Reconstrucción 3D

Introducción a la visión estéreo

- Los dos problemas de la visión estéreo
- Un sistema estéreo simple
- Parámetros de un sistema estéreo

Introducción a la visión estéreo

- El principal objetivo de la visión estéreo es recuperar la estructura 3D de una escena utilizando dos o más imágenes de dicha escena, cada una tomada desde un punto de vista diferente.
- El término "visión binocular" se utiliza cuando se usa la información de dos cámaras (par estéreo)

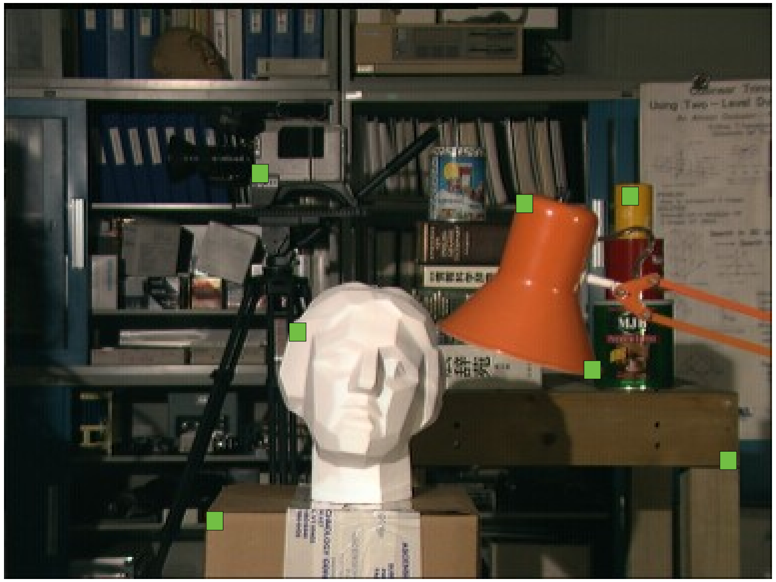


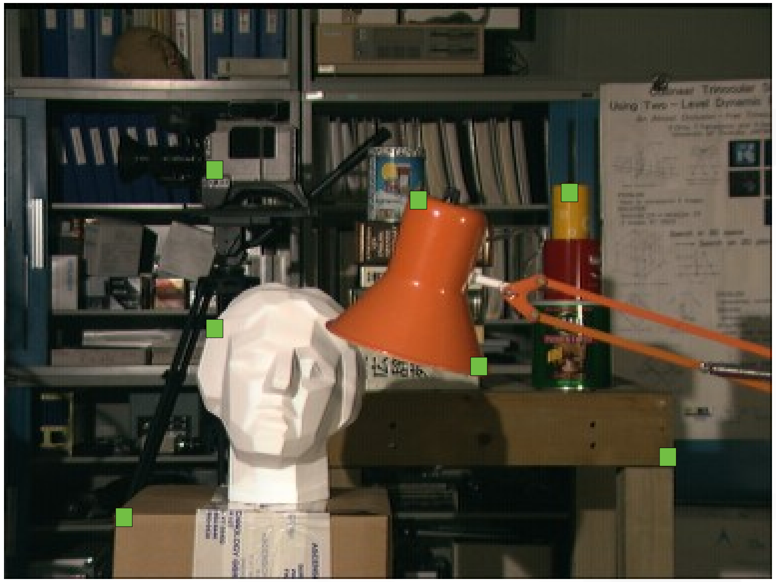
Introducción a la visión estéreo

- Alternativa a la visión estéreo: RGB-D
- Visión estéreo frente a RGB-D:
  - **RGB-D**:
    - Ventajas: proporcionan una reconstrucción densa; funcionan en superficies poco texturadas
    - Inconvenientes: rango de profundidades muy limitado; no funcionan bien en exteriores.
  - **Visión estéreo**:
    - Ventajas: mayor rango de profundidades; buenos resultados en exteriores; mayor precisión
    - Inconvenientes: resultados incompletos en zonas de baja textura; alto coste computacional.

Los dos problemas de la visión estéreo

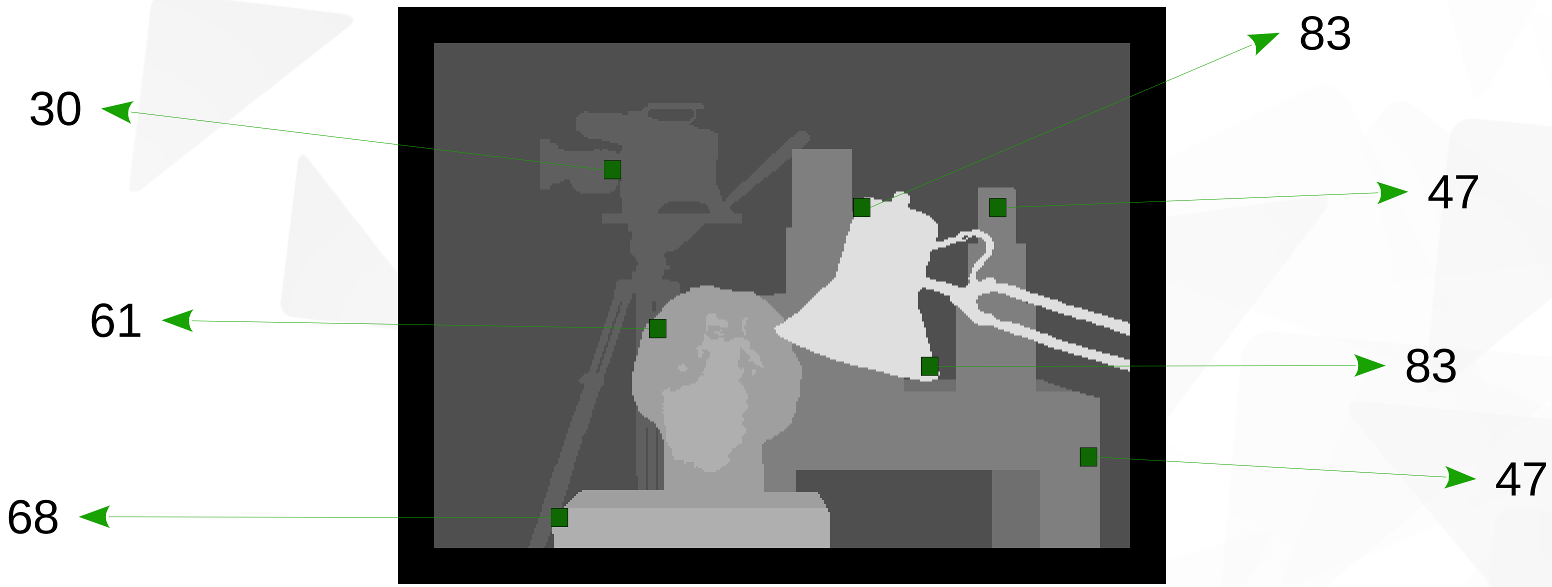
- El problema de la correspondencia:
  - Determinar qué píxel de una imagen se corresponde con un determinado píxel de la otra imagen.
  - Algunos objetos sólo son visibles desde una de las cámaras: el sistema debería poder determinar qué parte de la imagen no puede ponerse en correspondencia





Los dos problemas de la visión estéreo

- **Mapa de disparidad**: representa la diferencia de posición entre una pareja de puntos homólogos en las dos imágenes
  - Por cada punto de la imagen izquierda (x_l, y_l) emparejado con un punto de la imagen derecha (x_r, y_r):
    - DisparidadX[x_l, y_l] = $x_l - x_r$
    - DisparidadY[x_l, y_l] = $y_l - y_r$



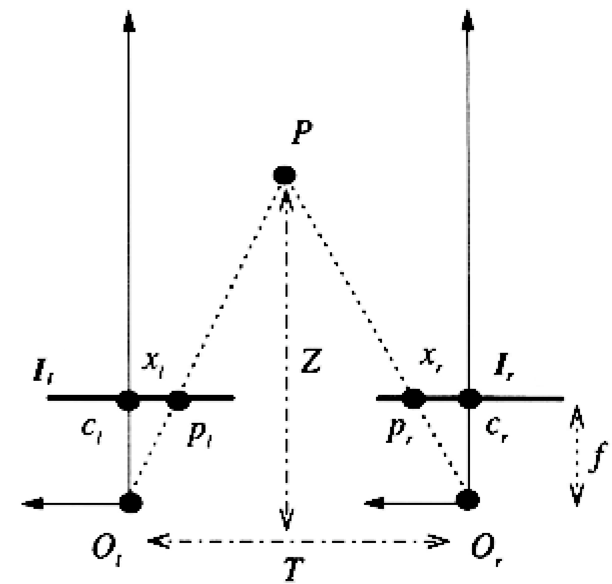
Los dos problemas de la visión estéreo

- Reconstrucción 3D:
  - Interpretación de la diferencia de posición en dos imágenes de un mismo punto de la escena.
  - Si la geometría del sistema es conocida, el mapa de disparidad puede transformarse en un mapa 3D.

Un sistema estéreo simple

- Dos cámaras *pinhole*.
- O_l y O_r son los centros de proyección.
- *f* es la distancia focal (igual en ambas cámaras)
- c_l y c_r son los puntos principales.
- La distancia *T* entre los dos centros de proyección se denomina *línea base*.
- Planos de imagen coplanares.
- Los ejes ópticos son paralelos: la intersección de ambos ejes está situada en el infinito.

Un sistema estéreo simple

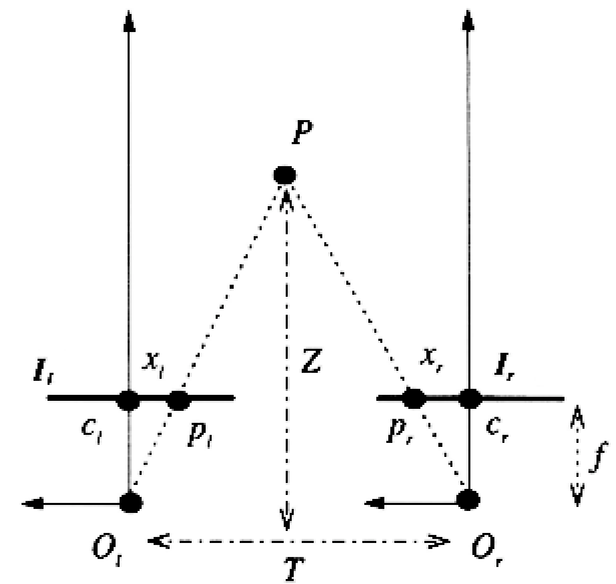


Un sistema estéreo simple

- Un punto *P* en el espacio se proyecta en los puntos de imagen p_l y p_r.
- x_l y x_r son las coordendas de p_l y p_r con respecto a los puntos principales.
- $d = x_l - x_r$ es la disparidad (diferencia de posición entre dos puntos en correspondencia).
- Considerando los triángulos equivalentes (p_l, P, p_r) y (O_l, P, O_r), las ratios *base/altura* en ambos triángulos tienen que ser iguales:

$$\frac{T-(x_l-x_r)}{Z-f}=\frac{T}{Z} \longrightarrow Z=f\frac{T}{d}$$

Un sistema estéreo simple

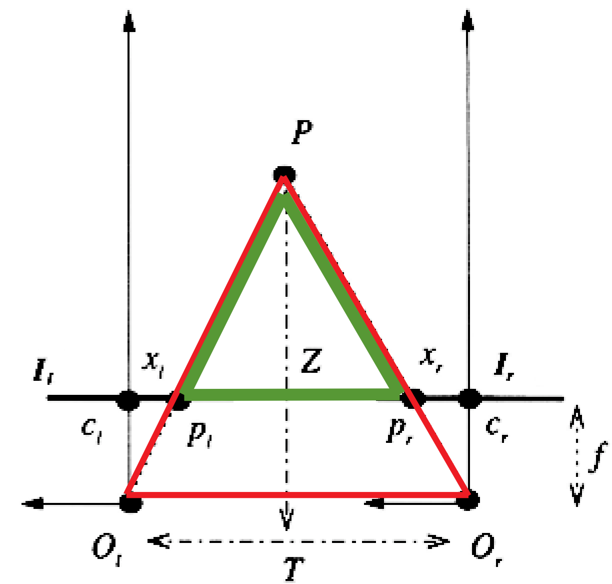


Un sistema estéreo simple

- Un punto P en el espacio se proyecta en los puntos de imagen p_l y p_r.
- x_l y x_r son las coordendas de p_l y p_r con respecto a los puntos principales.
- $d = x_l - x_r$ es la disparidad (diferencia de posición entre dos puntos en correspondencia).
- Considerando los triángulos equivalentes (p_l, P, p_r) y (O_l, P, O_r), las ratios *base/altura* en ambos triángulos tienen que ser iguales:

$$\boxed{\frac{T-(x_l-x_r)}{Z-f}}=\boxed{\frac{T}{Z}} \longrightarrow Z=f\frac{T}{d}$$

Un sistema estéreo simple



Parámetros de un sistema estéreo

- El sistema estéreo anterior está parametrizado por: f,T, c_l y c_r.
- En general, un sistema estéreo tiene dos tipos de parámetros:
  - Parámetros intrínsecos: relacionan coordenadas de puntos en el sistema de referencia de cada cámara con coordenadas de píxel (K_l, K_r).
  - Parámetros extrínsecos: describen la transformación que permite expresar puntos visto desde el sistema de referencia de una cámara en puntos en el sistema de referencia de la otra cámara (R, T).
- La calibración de un sistema estéreo es similar a la calibración de una sola cámara: el sistema de referencia del mundo se fija en una de las cámaras.

El problema de la correspondencia

- Métodos basados en correlación.
- Métodos basados en características.
- Criterios adicionales en la búsqueda de correspondencias.

El problema de la correspondencia

- Antes de reconstruir una escena, es necesario encontrar correspondencias entre las dos imágenes.
- Inicialmente puede asumirse que:
  - La mayoría de los puntos de la escena son visibles por ambas cámaras.
  - Las imágenes de dos regiones en correspondencia son similares.
- Esto se cumple si el punto de fijación está situado en una posición lejana en relación a la línea base.
- De no ser así, la búsqueda de correspondencias en el par de imágenes se complica debido principalmente al cambio de aspecto de una misma zona de la escena en las dos vistas.

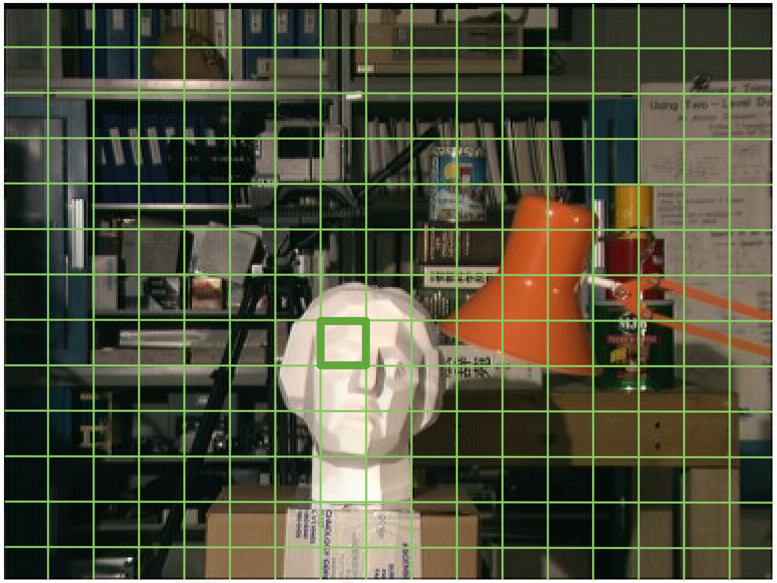
El problema de la correspondencia

- Cuando se realiza un proceso de búsqueda de correspondencias hay que decidir:
  - Qué elementos de imagen se van a poner en correspondencia (puntos, líneas, esquinas, bordes, …).
  - Qué medida de similitud se utilizará para compararlos.
- Hay dos estrategias principales:
  - Métodos basados en correlación.
  - Métodos basados en características.

Métodos basados en correlación

- En los **métodos basados en correlación**, los elementos de imagen que se ponen en correspondencia son ventanas de imagen de un tamaño fijo.
- El **criterio de correspondencia** es una **medida de similitud entre ventanas** de las dos imágenes: correlación cruzada, suma de las diferencias al cuadrado, …
- Una ventana en una imagen se corresponde con la ventana de la otra imagen que maximiza el criterio de similitud.
- Este método funciona bien si la línea base no es muy extensa y las imágenes están texturadas.
- **<u>Ventajas</u>**: dan como resultado un mapa de disparidad denso.
- **<u>Desventajas</u>**: puede ser muy lento y no funcionar bien en zonas uniformes.

Métodos basados en correlación

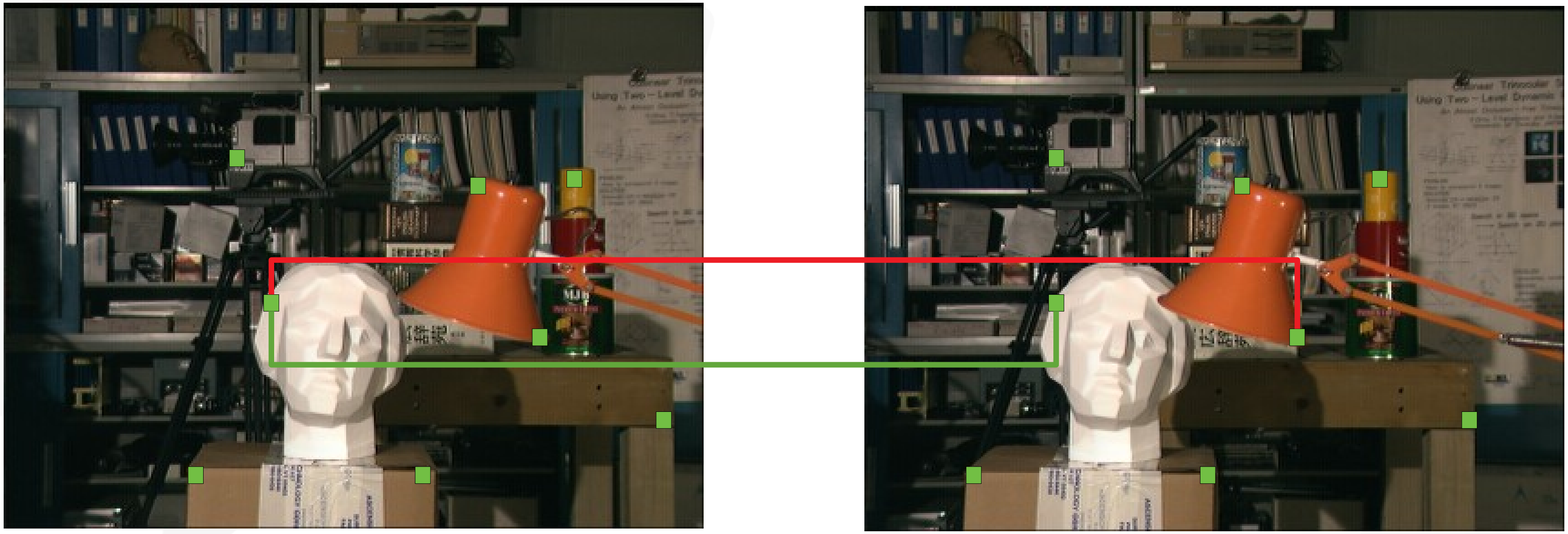


- **Similitud entre ventanas** (w_I, w_D):
  - Suma de las diferencias al cuadrado: $\sum \left(w_I(x,y) - w_D(x,y)\right)^2$
  - Coeficiente normalizado de correlación: $\dfrac{w'_I \otimes w'_D}{|w'_I||w'_D|}$

Métodos basados en características

- En los **métodos basados en características**, los elementos de imagen que se ponen en correspondencia son un conjunto de características extraídas previamente de cada imagen (por ejemplo, puntos característicos).
- Para comparar características pueden utilizarse descriptores invariantes a diferentes transformaciones.
- En función de las características utilizadas, el método proporciona buenos resultados cuando la separación entre vistas es significativa o las imágenes no tienen suficiente textura.
- **Ventajas**: rapidez (depende de la velocidad en la extracción de características).
- **Desventajas**: proporciona un conjunto de correspondencias dispersas.

Métodos basados en correlación



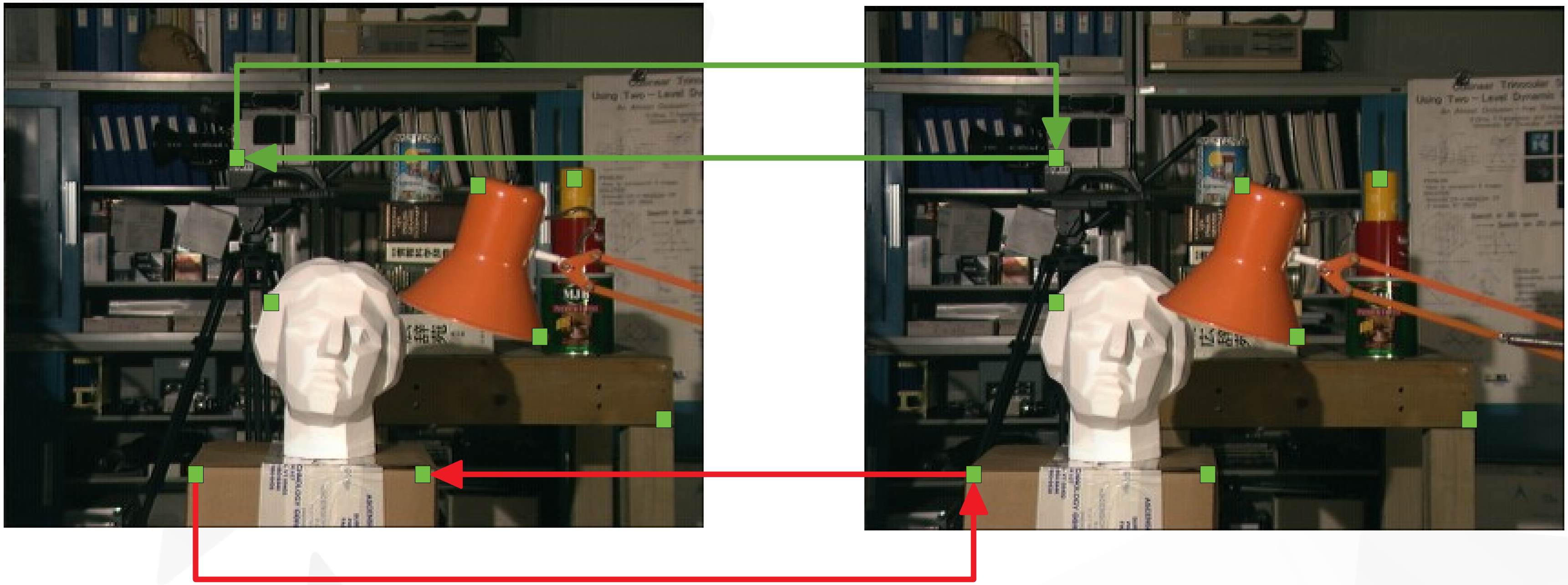
- Comparación entre puntos dispersos:
  - Se reduce el conjunto de posibles homólogos de cada punto.
  - Si se utilizan descriptores de región, los emparejamientos son más fiables.
  - Para obtener un 3D denso es necesario extrapolar los resultados al resto de puntos de la imagen.

Criterios adicionales

- El rendimiento de un método de búsqueda de correspondencias se ve afectado por las oclusiones o las correspondencias espurias.
- Imponiendo restricciones adicionales se reducen los efectos de ambos problemas.
- Dos restricciones importantes son:
  - **Consistencia entre pares**: dos elementos de imagen en correspondencia maximizan el criterio de similitud en los dos sentidos (izquierda-derecha, derecha-izquierda).
  - **Restricción epipolar**: los emparejamientos deben ser coherentes con la geometría del sistema estéreo.

Métodos basados en correlación

Consistencia entre pares

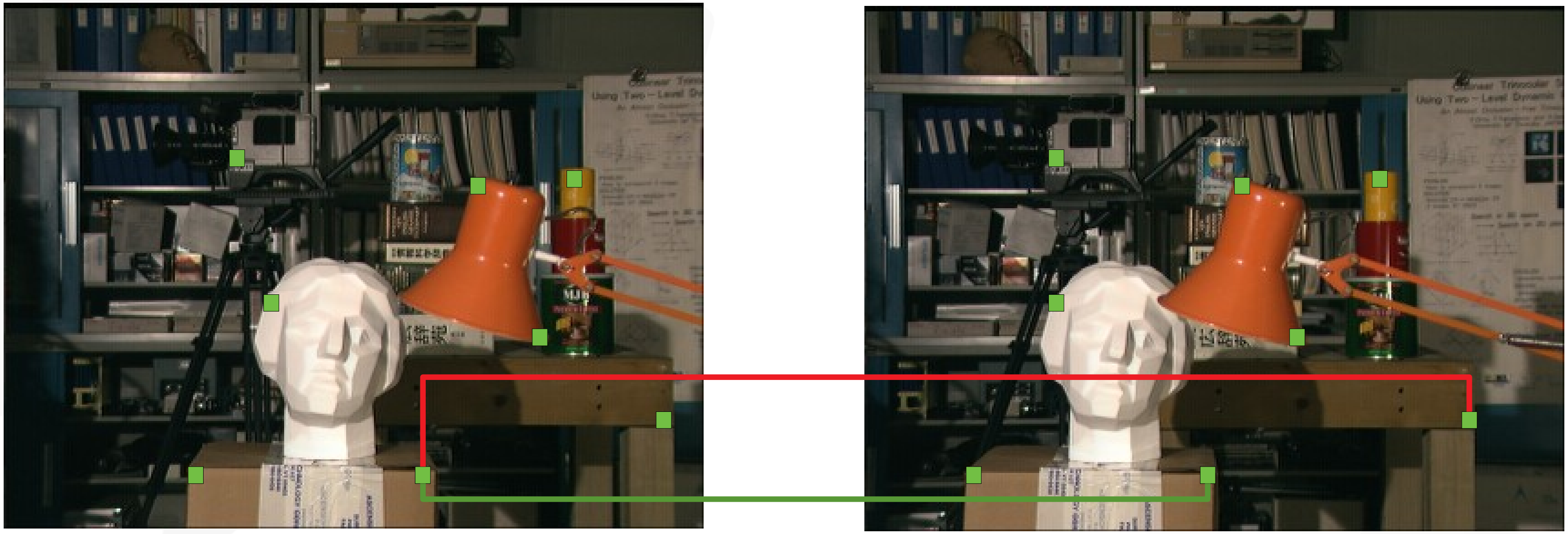


Los puntos son consistentes. El mejor candidato en un sentido de la comparación genera la misma pareja en el otro sentido.

Los puntos no son consistentes. El mejor candidato en un sentido de la comparación genera otra pareja diferente.

Métodos basados en correlación

Restricción epipolar



La geometría del sistema estéreo permite el emparejamiento. Los dos puntos pueden provenir del mismo punto 3D.

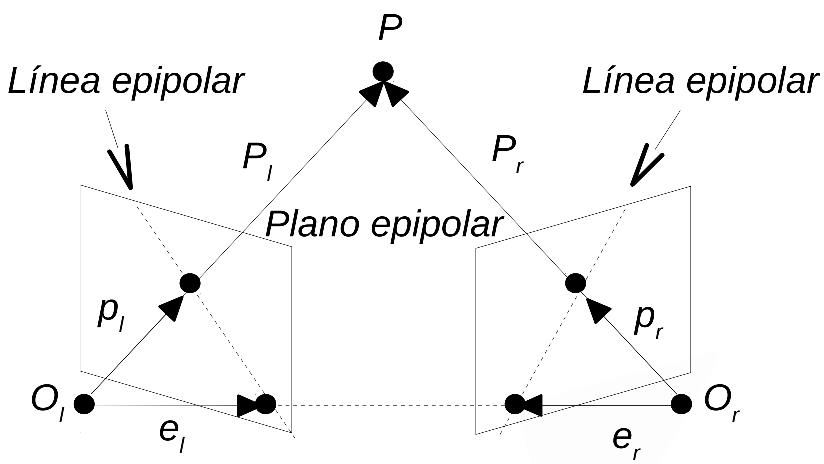
La geometría no permite el emparejamiento. Los puntos emparejados no pueden ser proyecciones del mismo punto 3D.

Geometría epipolar

- La matriz esencial
- La matriz fundamental
- Estimación de las matrices esencial y fundamental
- Rectificación de imágenes

Geometría epipolar

- La geometría de un sistema estéreo se conoce como geometría epipolar.
- Plano epipolar: el plano que pasa por los dos centros de proyección y un punto 3D.
- Línea epipolar: intersección entre el plano epipolar y el plano de imagen.
- Epipolos: puntos de corte en los planos de imagen de la línea que pasa por los dos centros de proyección.

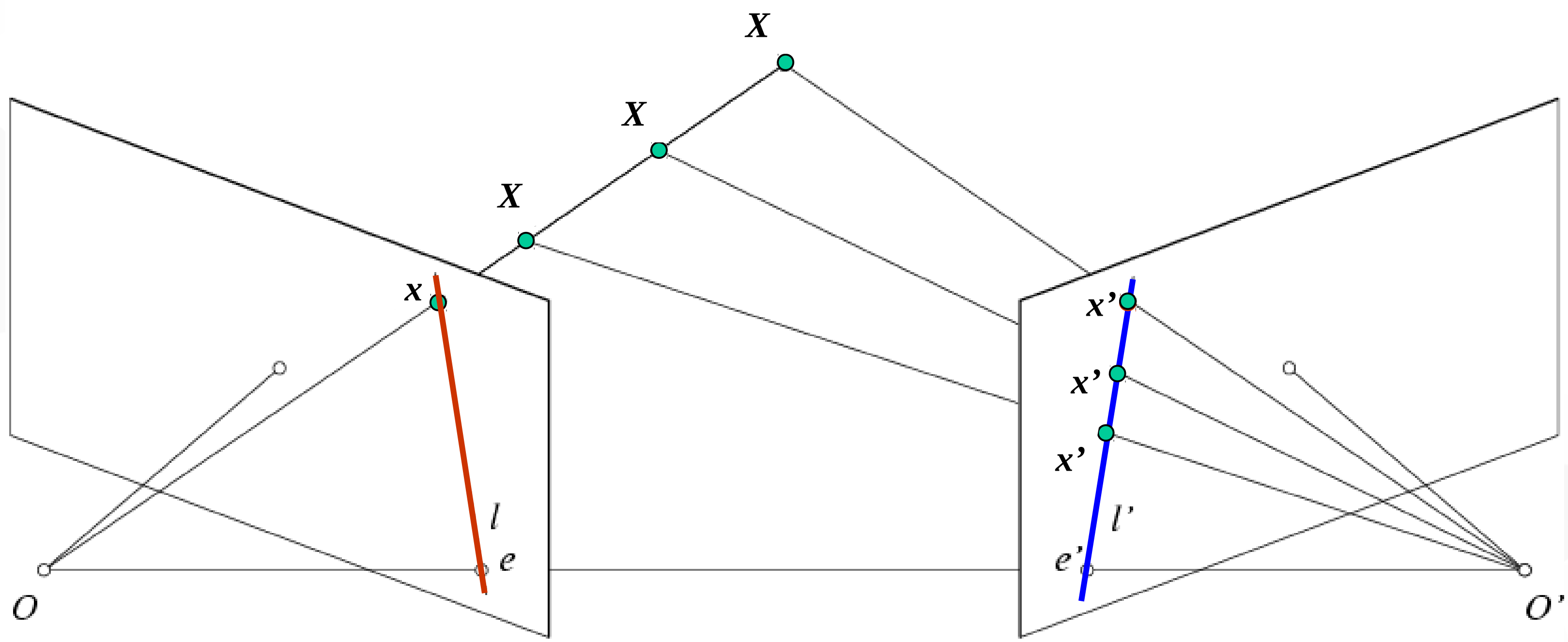


- P: punto 3D.
- P_l: P en el sistema de referencia de la cámara izquierda.
- P_r: P en el sistema de referencia de la cámara derecha.
- p_l: proyección de P en la cámara izquierda.
- p_r: proyección de P en la cámara derecha.
- e_l y e_r: epipolos.
- O_l y O_r: centros de proyección.

Geometría epipolar

- Todas las líneas epipolares pasan por el epipolo de la cámara.
- Si la línea base es paralela al plano de imagen, el epipolo está situado en el infinito.
- Restricción epipolar:
  - Dado p_l , P está situado en cualquier punto del rayo que pase por O_l y p_l.
  - La proyección de este rayo en la imagen derecha es la línea epipolar de p_l.
  - El punto p_r homólogo a p_l tiene que pertenecer a la línea epipolar de p_l.
  - La búsqueda de correspondencias se reduce a una línea.

Geometría epipolar



Geometría epipolar

- Dado un punto en una imagen, ¿cómo estimar la línea epipolar correspondiente en la otra imagen?
- Es necesario estimar la geometría epipolar, codificada en dos matrices:
  - **La matriz esencial**
  - **La matriz fundamental**

La matriz esencial

- Fundamentos:
  - Los sistemas de referencia de las cámaras izquierda y derecha están relacionados por los parámetros extrínsecos: vector de traslación $T = (O_r - O_l)$ y matriz de rotación R.
  - Dado un punto 3D P, la relación entre P_l y P_r viene dada por: $P_r = R(P_l - T)$.
  - La relación entre un punto 3D y sus proyecciones viene descrita por:

$$p_l = \frac{f_l}{Z_l} P_l \qquad\qquad p_r = \frac{f_r}{Z_r} P_r$$

La matriz esencial

- Matriz esencial **E**:

$$S=\begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \longrightarrow E=RS \longrightarrow P_r^T E P_l = 0$$

- Utilizando las expresiones de los puntos proyectados (p_l y p_r), se obtiene una ecuación que permite relacionar ambos puntos:

$$P_r^T E P_l = 0 \longrightarrow p_r^T E p_l = 0$$

- Definiendo $u_r = Ep_l$, se obtiene la expresión $p_r^T u_r = 0$ de pertenencia de un punto a una recta, por lo que u_r codifica la línea epipolar en la cámara derecha de un punto p_l en la cámara izquierda.

La matriz fundamental

- La matriz fundamental (*F*) es similar a la esencial, pero relaciona puntos de imagen en lugar de puntos de cámara.
- Dadas las matrices de intrínsecos de las dos cámaras (K_l y K_r), la relación entre puntos en coordenadas de cámara (p_l y p_r) y sus correspondientes en coordenadas de píxel ($\bar{p}_l$ y $\bar{p}_r$), viene dada por:

$$p_l = K_l^{-1}\,\bar{p}_l \qquad\qquad p_r = K_r^{-1}\,\bar{p}_r$$

- Utilizando la restricción impuesta por la matriz esencial:

$$p_r^T E\, p_l = 0 \longrightarrow \bar{p}_r^{\,T} F\, \bar{p}_l = 0$$

- La matriz *F* se denomina matriz fundamental y viene dada por:

$$F = K_r^{-T} E\, K_l^{-1}$$

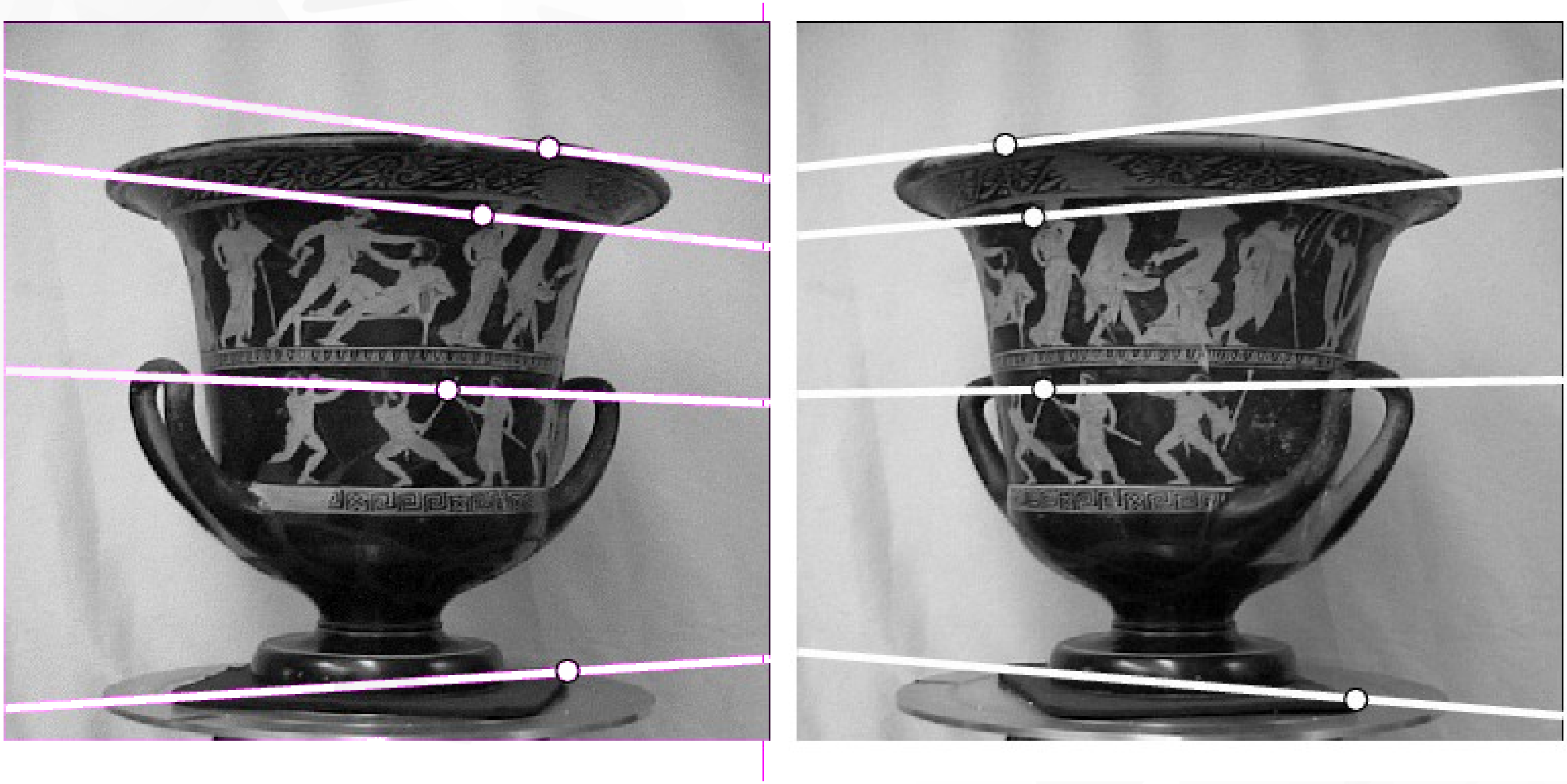
La matriz fundamental

- En coordenadas de píxel, la línea epipolar en la imagen derecha ($\bar{u}_r$) asociada con un punto de la imagen izquierda ($\bar{p}_l$), se obtiene como:

$$\bar{u}_r = F\,\bar{p}_l$$

- Una vez estimada la matriz fundamental, se obtiene una codificación de la geometría epipolar del sistema estéreo sin necesidad de conocer los parámetros intrínsecos y extrínsecos del sistema.
- La matriz esencial proporciona una relación entre los sistemas de coordenadas de las dos cámaras.
- La matriz fundamental proporciona una relación entre las dos imágenes.

La matriz fundamental

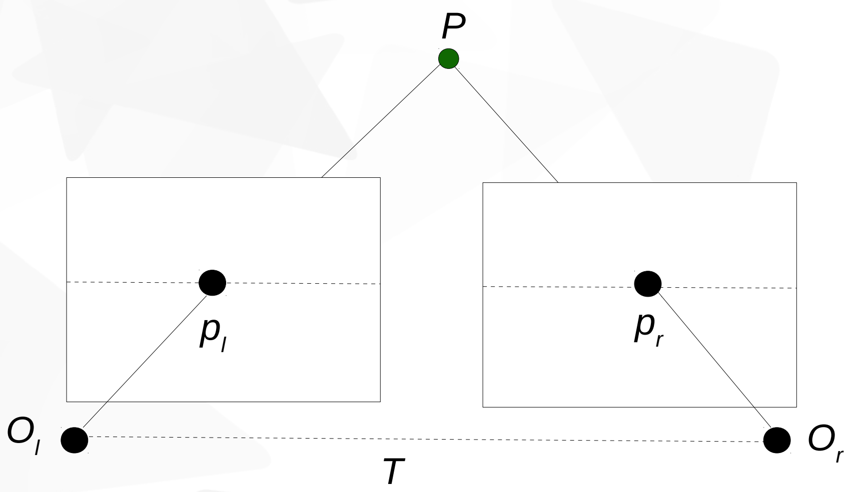


Rectificación de imágenes

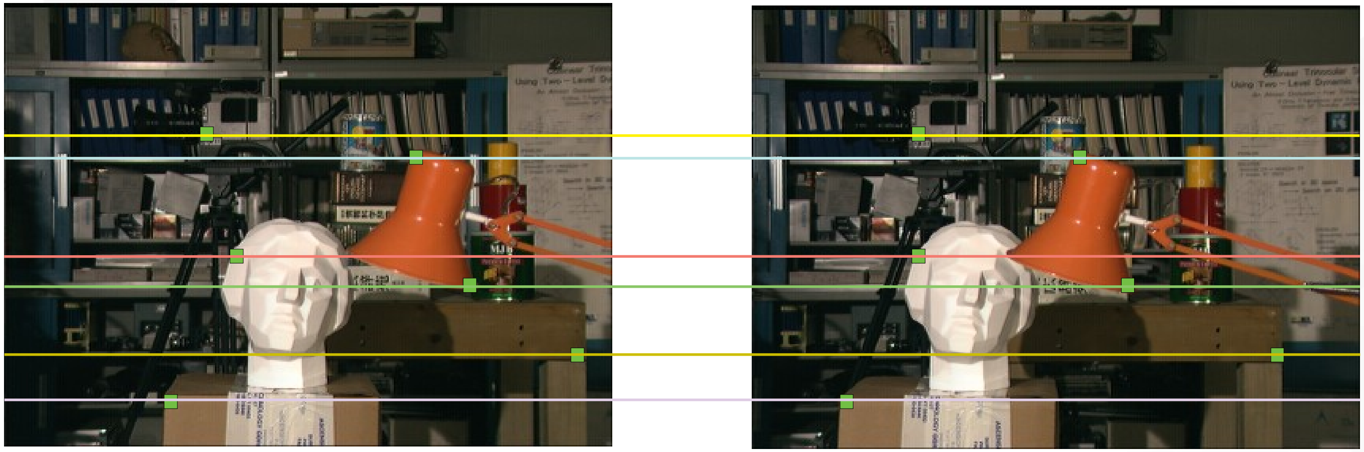
- La rectificación de imágenes es una transformación que convierte pares conjugados de líneas epipolares en colineales y paralelas al eje horizontal (línea base).

- La búsqueda de correspondencias se simplifica si las imágenes están rectificadas (la coordenada *y* de dos puntos en correspondencia es la misma)

Rectificación de imágenes



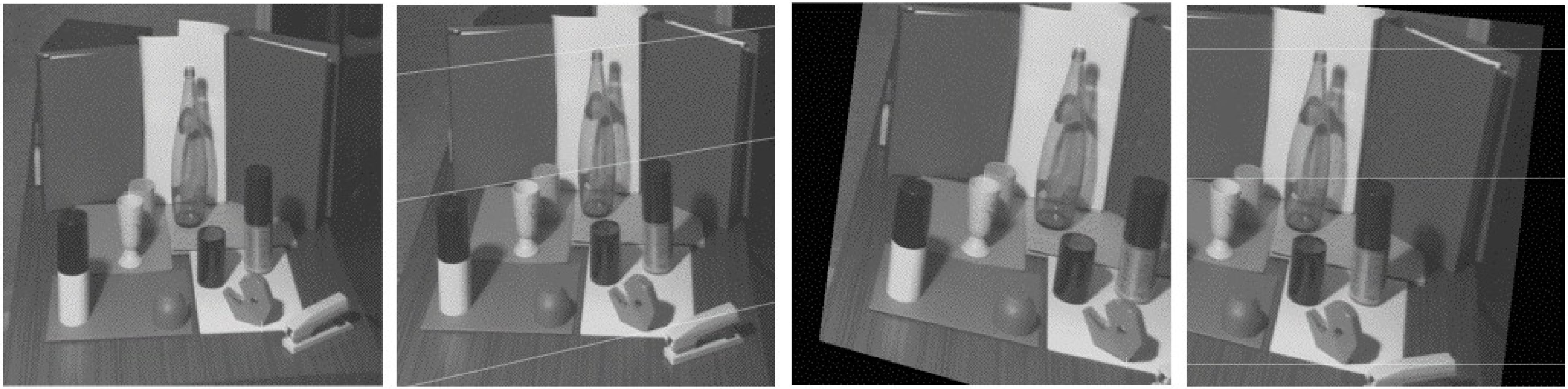
- Si no existe rotación entre cámaras, los epipolos se sitúan en el infinito.
- Las líneas epipolares son horizontales.
- El homólogo de un punto $p_l = (x_l, y_l)$ está situado en la fila y_l de la otra imagen.



Rectificación de imágenes

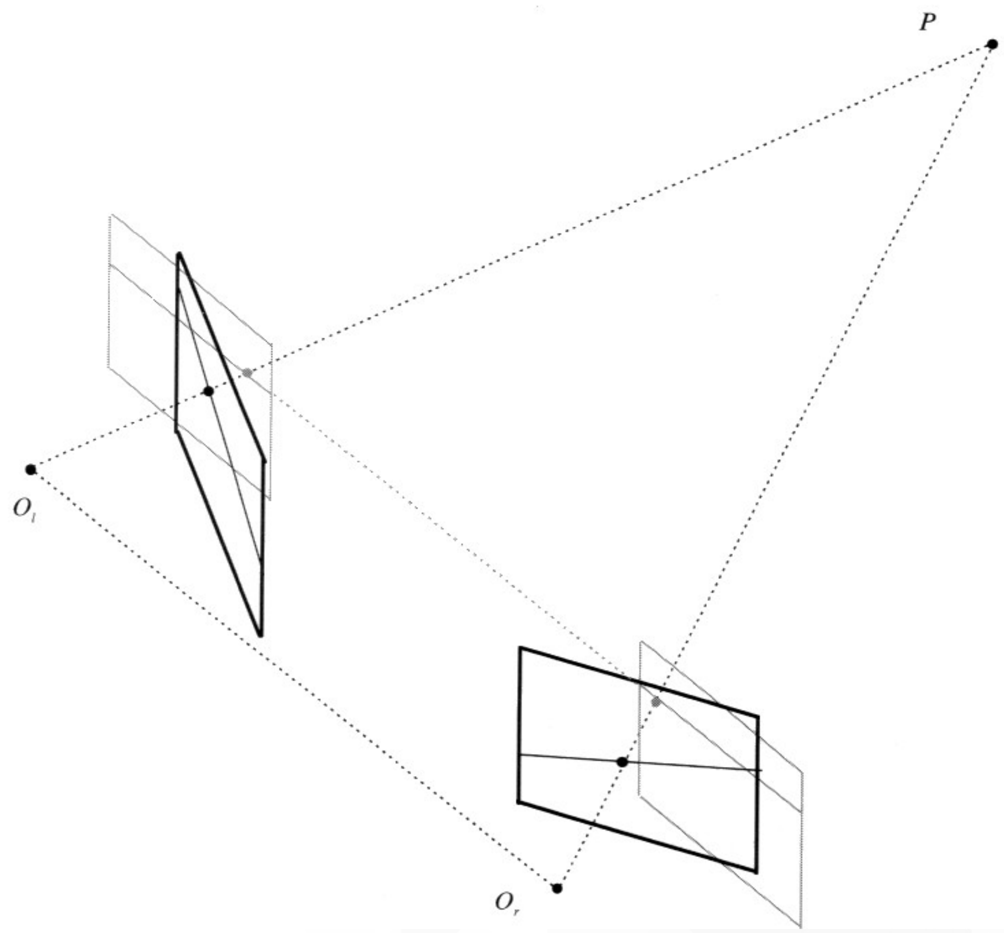
- La posición original de las cámaras puede ser cualquiera.

- La rectificación alinea las líneas epipolares con las filas de la imagen: la disparidad sólo se produce en el eje *x*.



Rectificación de imágenes

- La rectificación puede entenderse como una rotación de las cámaras alrededor de sus centros ópticos que consigue situar los epipolos en el infinito.
- Las imágenes rectificadas son el resultado de aplicar las rotaciones anteriores a los puntos de las imágenes originales.



Rectificación de imágenes

- Suponiendo conocidos los parámetros extrínsecos (*R* y *T*) e intrínsecos (*f*) del sistema estéreo:
  - Rotar la cámara izquierda de manera que el epipolo se sitúe en el infinito:

$$p'_l = R_{rect}\, p_l$$

  - Aplicar la misma rotación a la cámara derecha para recuperar la geometría original.
  - Rotar la cámara derecha mediante *R* (rotación del par estéreo).

$$p'_r = R\, R_{rect}\, p_r$$

  - Ajustar la escala de los puntos en ambas cámaras.

$$p'_l = [x', y', z']^T \longrightarrow \bar{p}'_l = \frac{f}{z'}\, p'_l$$

Reconstrucción 3D

- Reconstrucción mediante triangulación

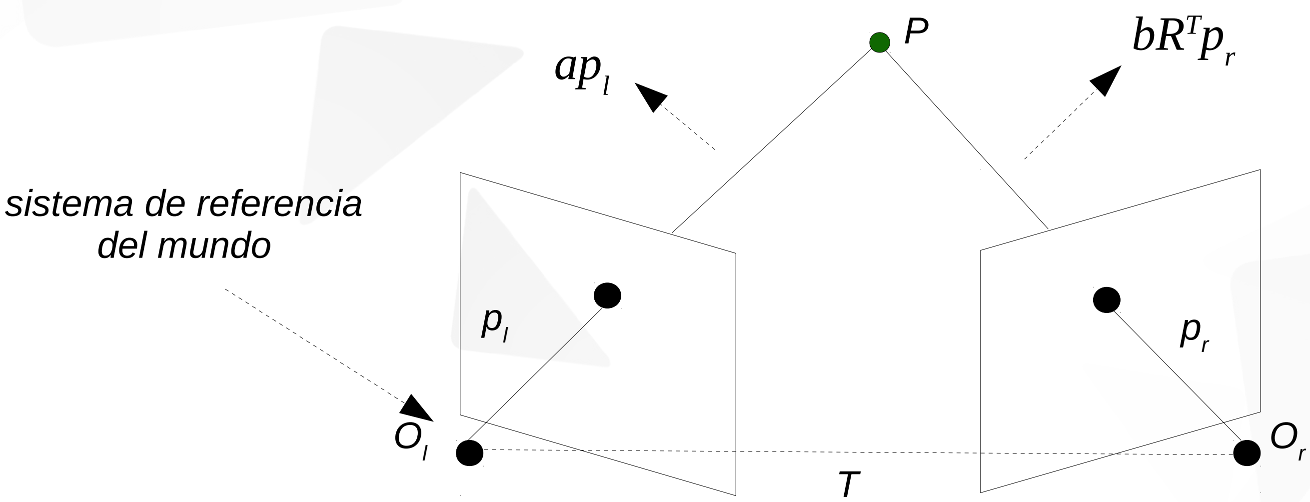
Reconstrucción 3D

- El método para determinar las coordenadas de un punto 3D a partir de una correspondencia estéreo depende de la información disponible.
- Si los parámetros intrínsecos y extrínsecos son conocidos, se obtiene una estimación completa.
- Si sólo se conocen los parámetros intrínsecos, la estimación del 3D depende de un factor de escala.
- Si tantos los parámetros intrínsecos como los extrínsecos son conocidos, la estimación depende de una proyección 3D.

Reconstrucción mediante triangulación

- Si los parámetros intrínsecos y extrínsecos son conocidos, es posible calcular el punto 3D asociado a una correspondencia $\bar{p}_l$-$\bar{p}_r$ (en posiciones de píxel) a partir de sus proyecciones:

$$p_l = K_l^{-1}\,\bar{p}_l \qquad p_r = K_r^{-1}\,\bar{p}_r$$

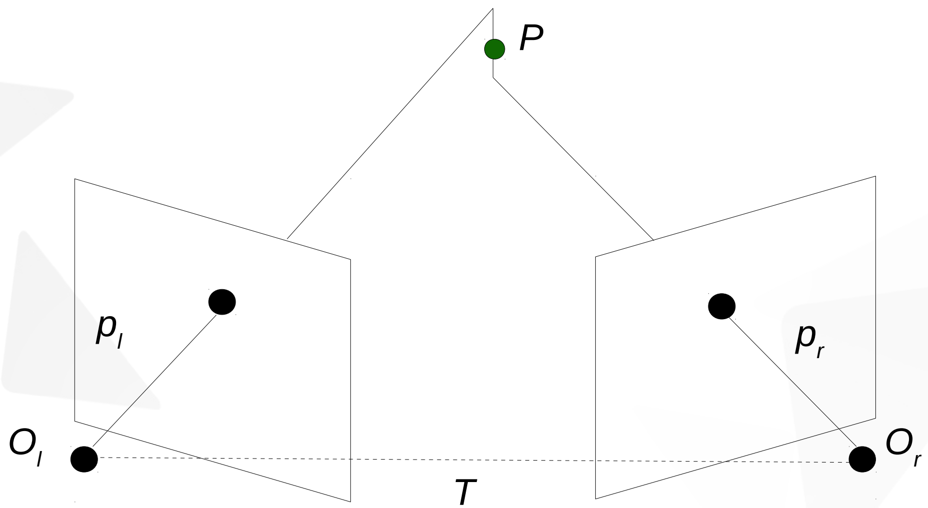


$$P = a\,p_l$$
$$P = b\,R^T\,p_r + T$$

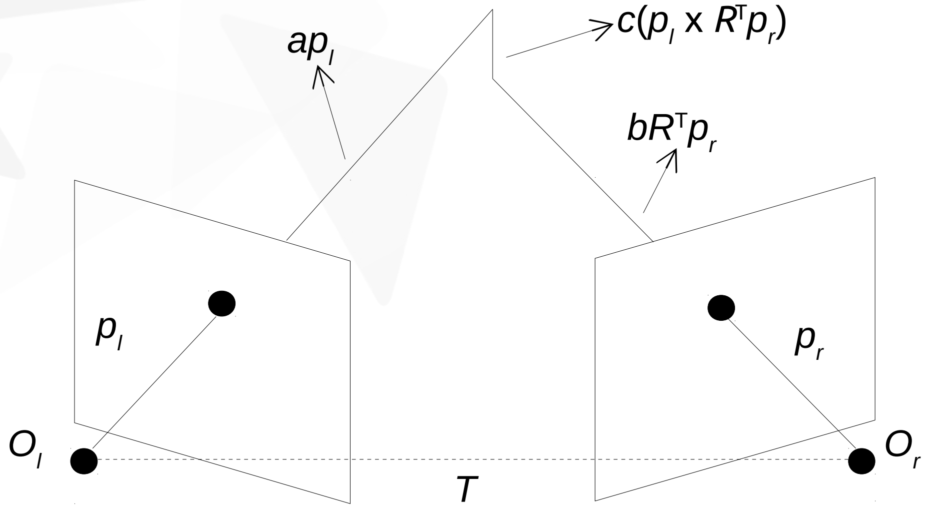
$$\boxed{a\,p_l - b\,R^T\,p_r = T}$$

Reconstrucción mediante triangulación

- Puesto que los parámetros y las posiciones de los puntos homólogos son conocidas sólo de manera aproximada, es posible que las proyecciones no se corten.
- *P* se aproxima al punto medio de la perpendicular común a las dos proyecciones.



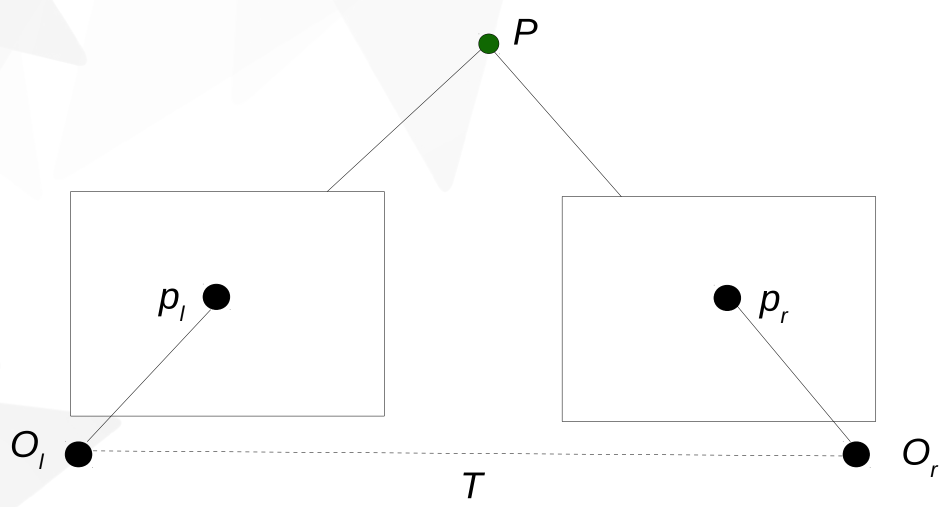
Reconstrucción mediante triangulación



- Debe cumplirse: $a\,p_l - b\,R^T p_r + c\left(p_l \times R^T p_r\right) = T$
- a_0 y b_0 son las soluciones del sistema de ecuaciones anterior.
- El punto 3D se determina como el punto medio del segmento definido por $(a_0 p_l)$ y $(b_0 R^T p_r + T)$.

Reconstrucción mediante triangulación

Imágenes rectificadas (R = I)



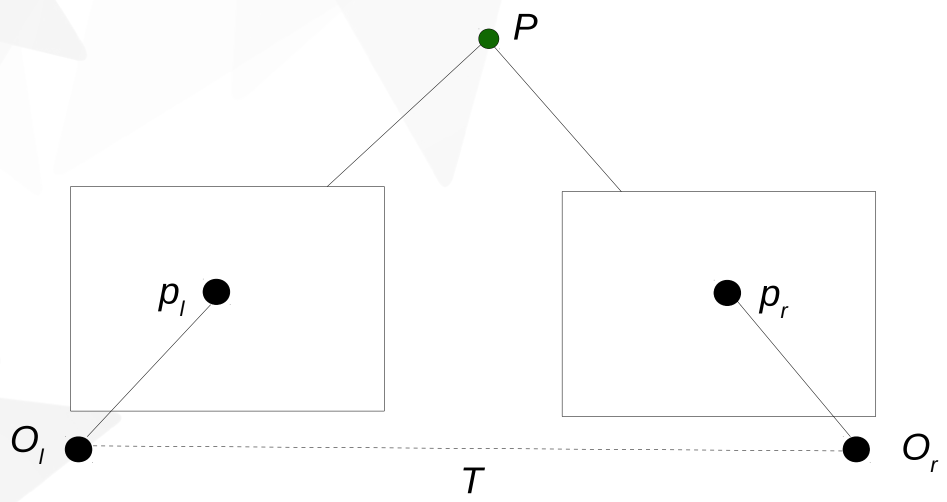
$$P = a\, p_l$$
$$P = b\, p_r + T$$

$$\boxed{a\, p_l - b\, p_r = T}$$

$$p_l = \begin{bmatrix} -(x_l - c_x)/f \\ -(y - c_y)/f \\ 1 \end{bmatrix} \qquad p_r = \begin{bmatrix} -(x_r - c_x)/f \\ -(y - c_y)/f \\ 1 \end{bmatrix} \qquad T = \begin{bmatrix} -T_x \\ 0 \\ 0 \end{bmatrix}$$

Reconstrucción mediante triangulación

Imágenes rectificadas (R = I)



$$P = a\,p_l$$
$$P = b\,p_r + T$$

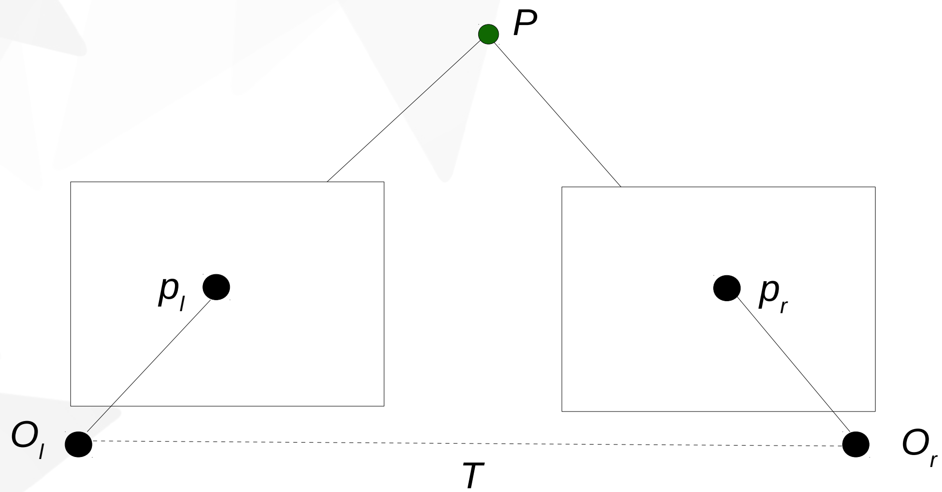
$$a\,p_l - b\,p_r = T$$

$$-a(x_l - c_x)/f + b(x_r - c_x)/f = -T_x$$
$$-a(y - c_y)/f + b(y - c_y)/f = 0$$
$$a - b = 0$$

$$a = b$$

Reconstrucción mediante triangulación

Imágenes rectificadas (R = I)



$$P = a\,p_l$$
$$P = b\,p_r + T$$

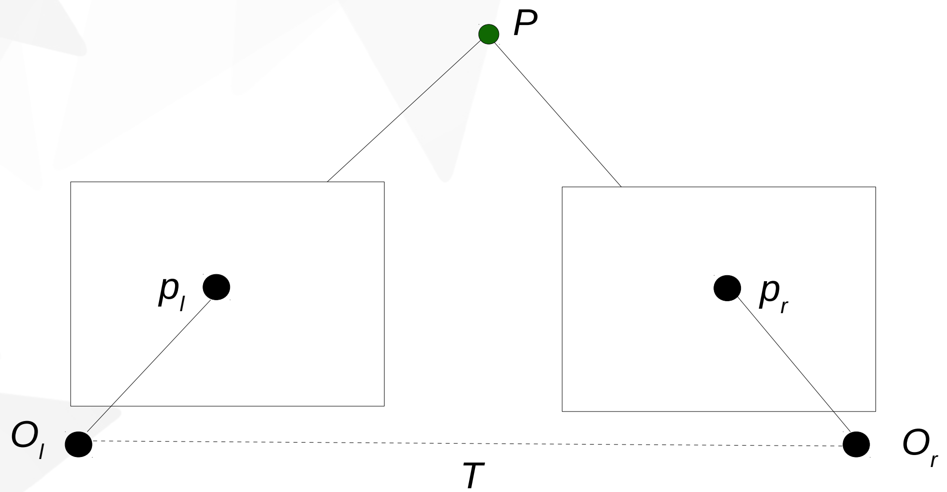
$$\boxed{a\,p_l - b\,p_r = T}$$

$$-a(x_l - c_x)/f + a(x_r - c_x)/f = -T_x \longrightarrow a\underbrace{(x_l - x_r)}_{d} = fT_x$$

disparidad horizontal

Reconstrucción mediante triangulación

Imágenes rectificadas (R = I)



$$P = a\,p_l$$
$$P = b\,p_r + T$$

$$\boxed{a\,p_l - b\,p_r = T}$$

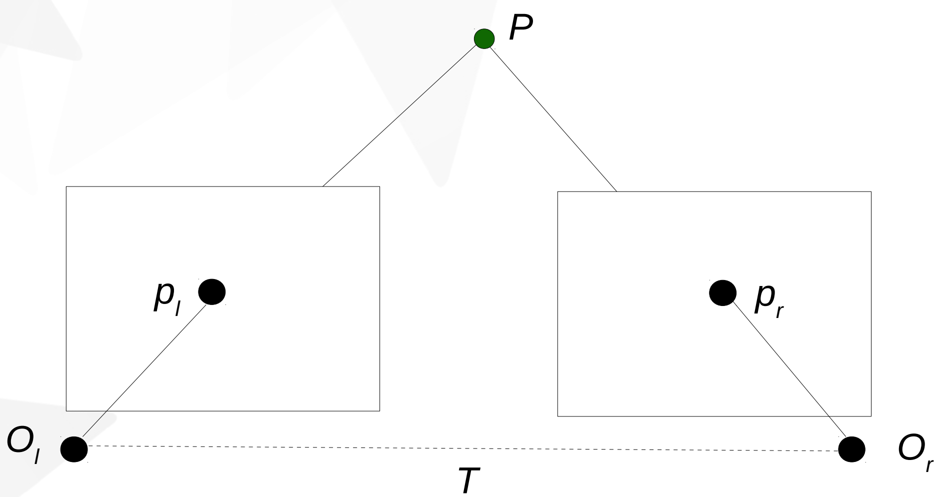
$$a = f\,\frac{T_x}{d}$$

$$p_l = \begin{bmatrix} -(x_l - c_x)/f \\ -(y - c_y)/f \\ 1 \end{bmatrix}$$

$$P = a\,p_l$$

Reconstrucción mediante triangulación

Imágenes rectificadas (R = I)



$$P = a\,p_l$$
$$P = b\,p_r + T$$

$$a\,p_l - b\,p_r = T$$

$$P = \begin{bmatrix} -(x_l - c_x)T_x/d \\ -(y - c_y)T_x/d \\ fT_x/d \end{bmatrix}$$