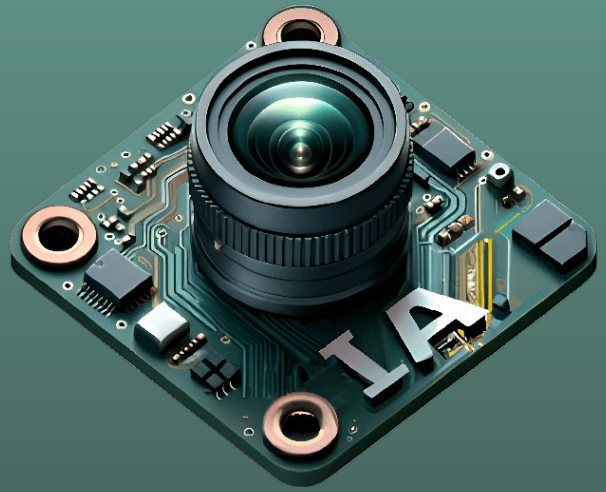
Universidad Industrial de Santander

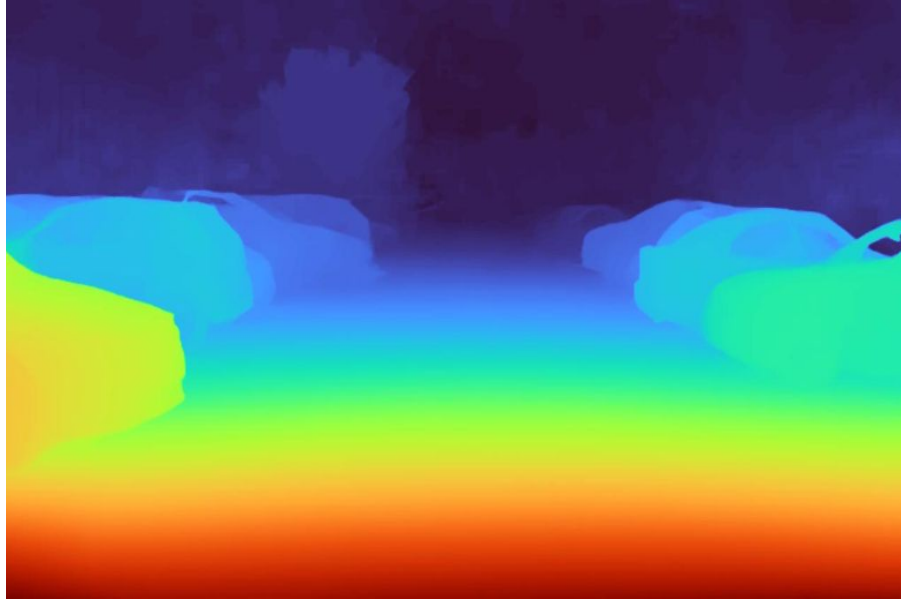


Semillero de Investigación
"Hands - on"
Computer Vision


SESIÓN 5:
PASSIVE DEPTH

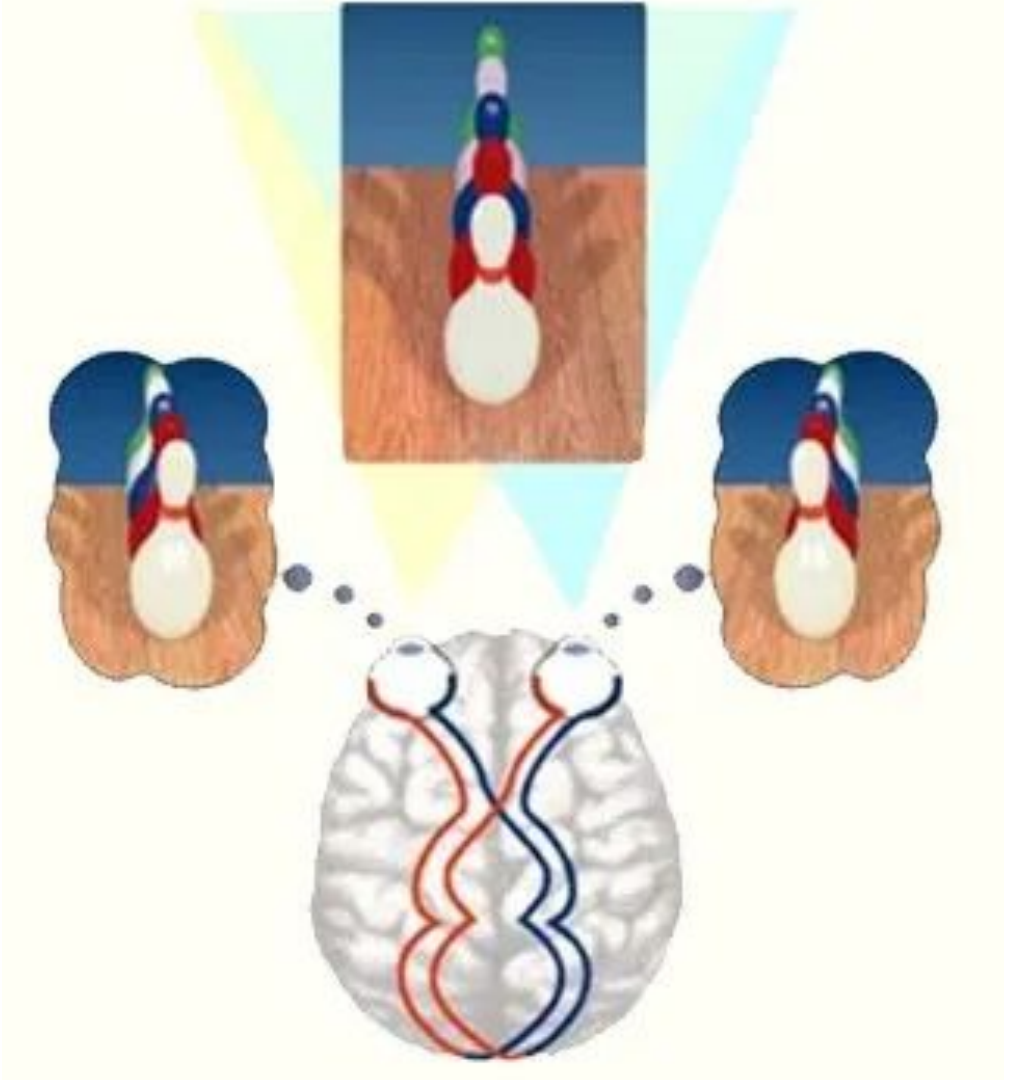
Contenidos

1. Percepción de profundidad
2. Técnicas de estimación de profundidad
   a. Métodos pasivos
   b. Métodos activos
3. Visión stereo
   a. Pipeline
   b. Disparidad y profundidad
   c. Rectificación de cámaras
   d. Triangulación y estimación 3D
4. Hands-on Stereo Vision

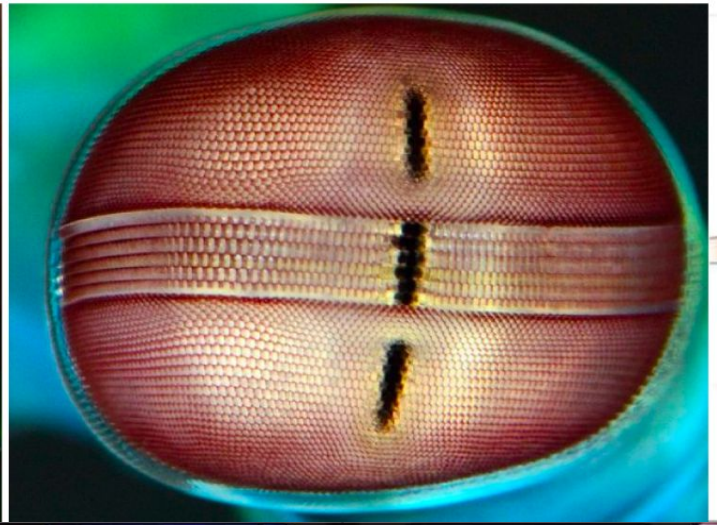


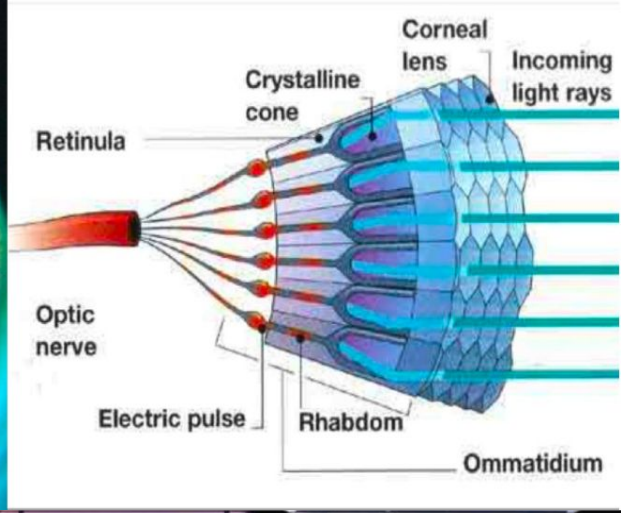
1. Percepción de profundidad

Percepción de profundidad

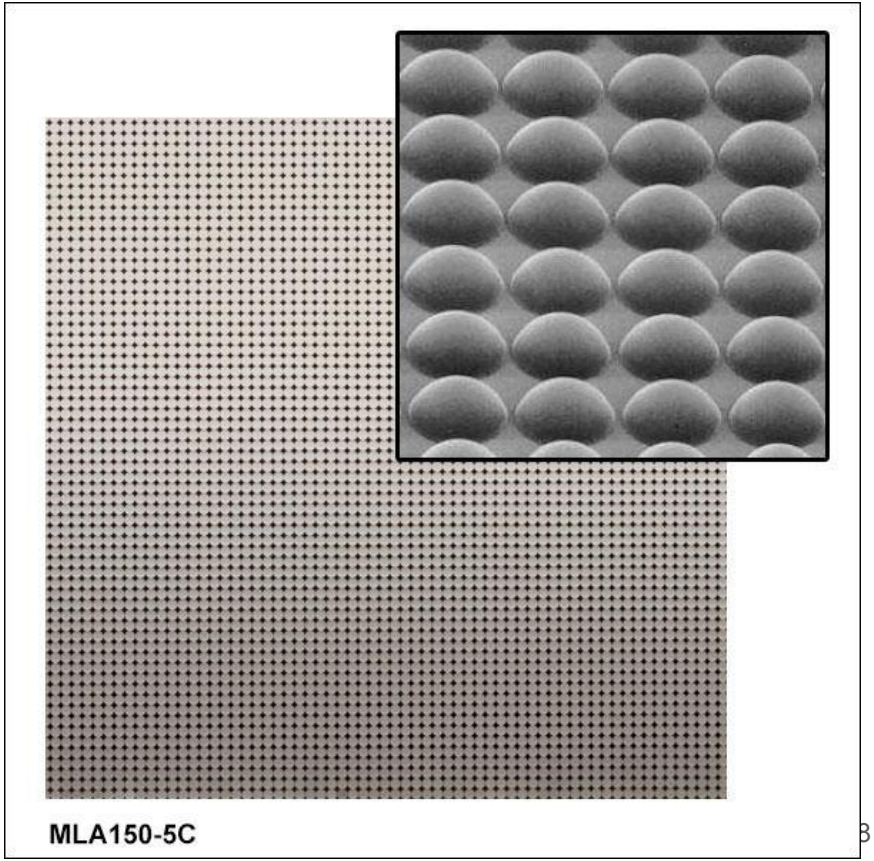


Más allá de Visión Binocular



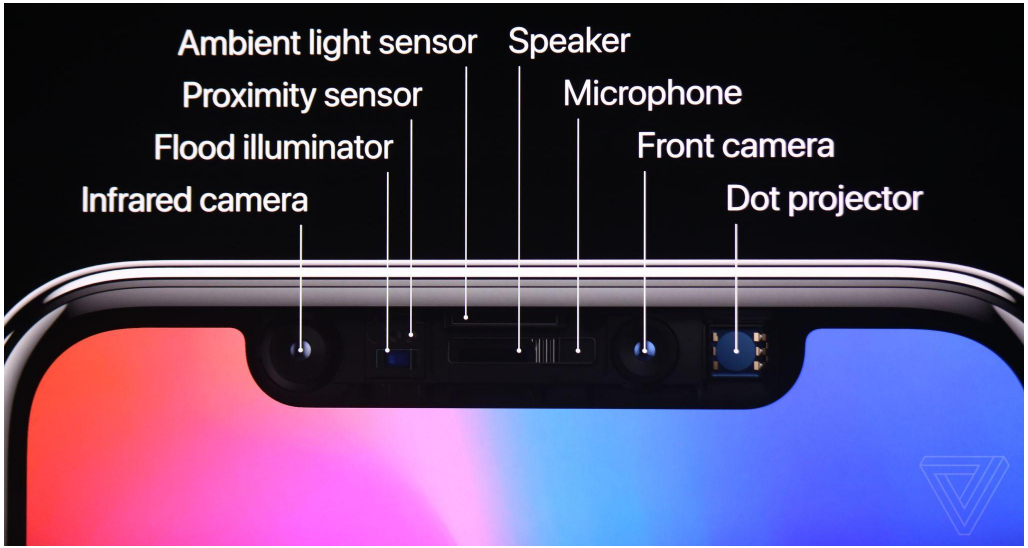


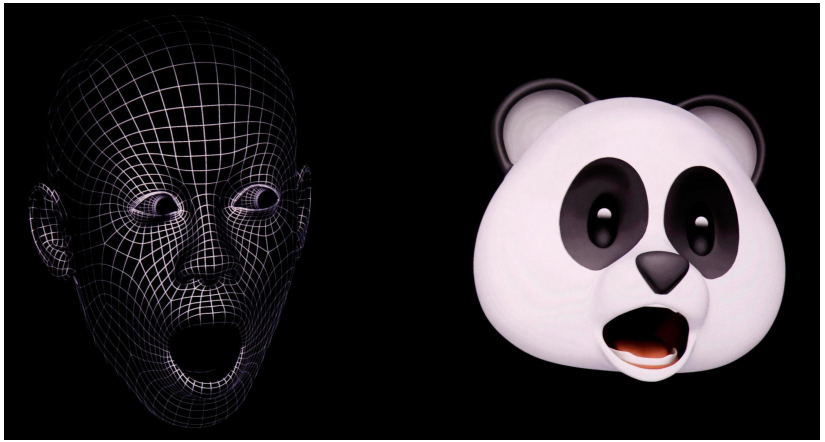
Múltiples “pares” de cámaras



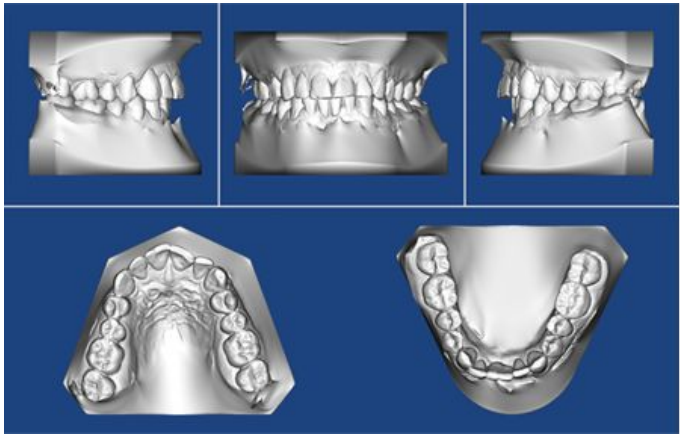
Imágenes de profundidad (x,y,z)

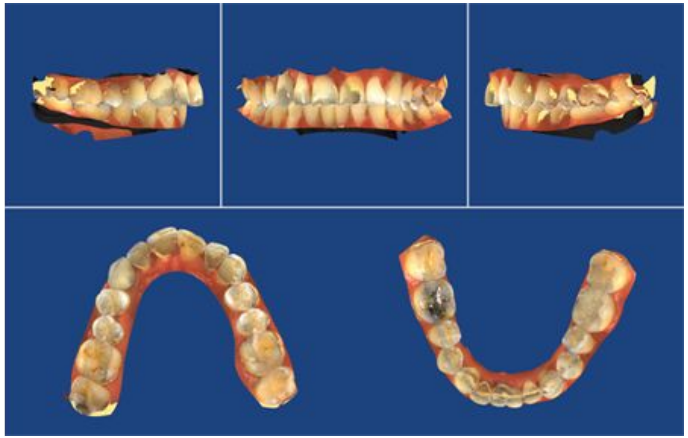
Aplicaciones

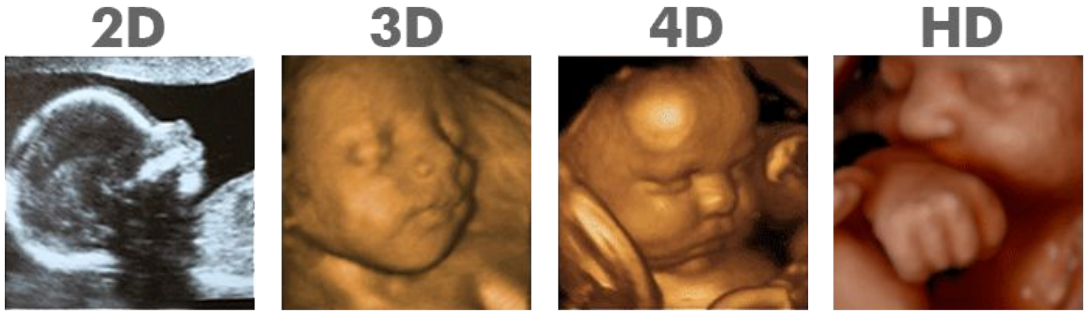




Aplicaciones



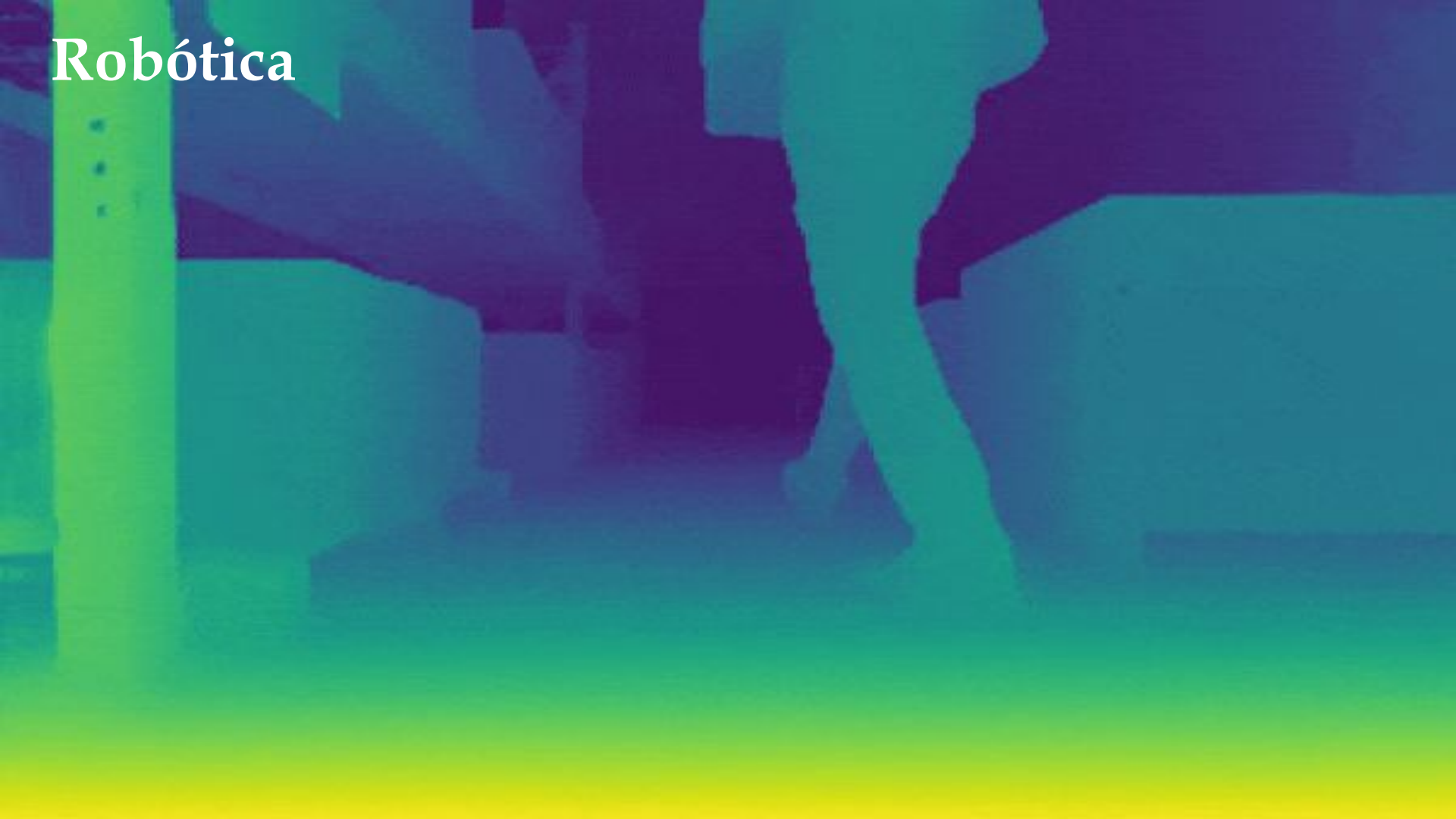




Input

Our depth predictions*

Robótica



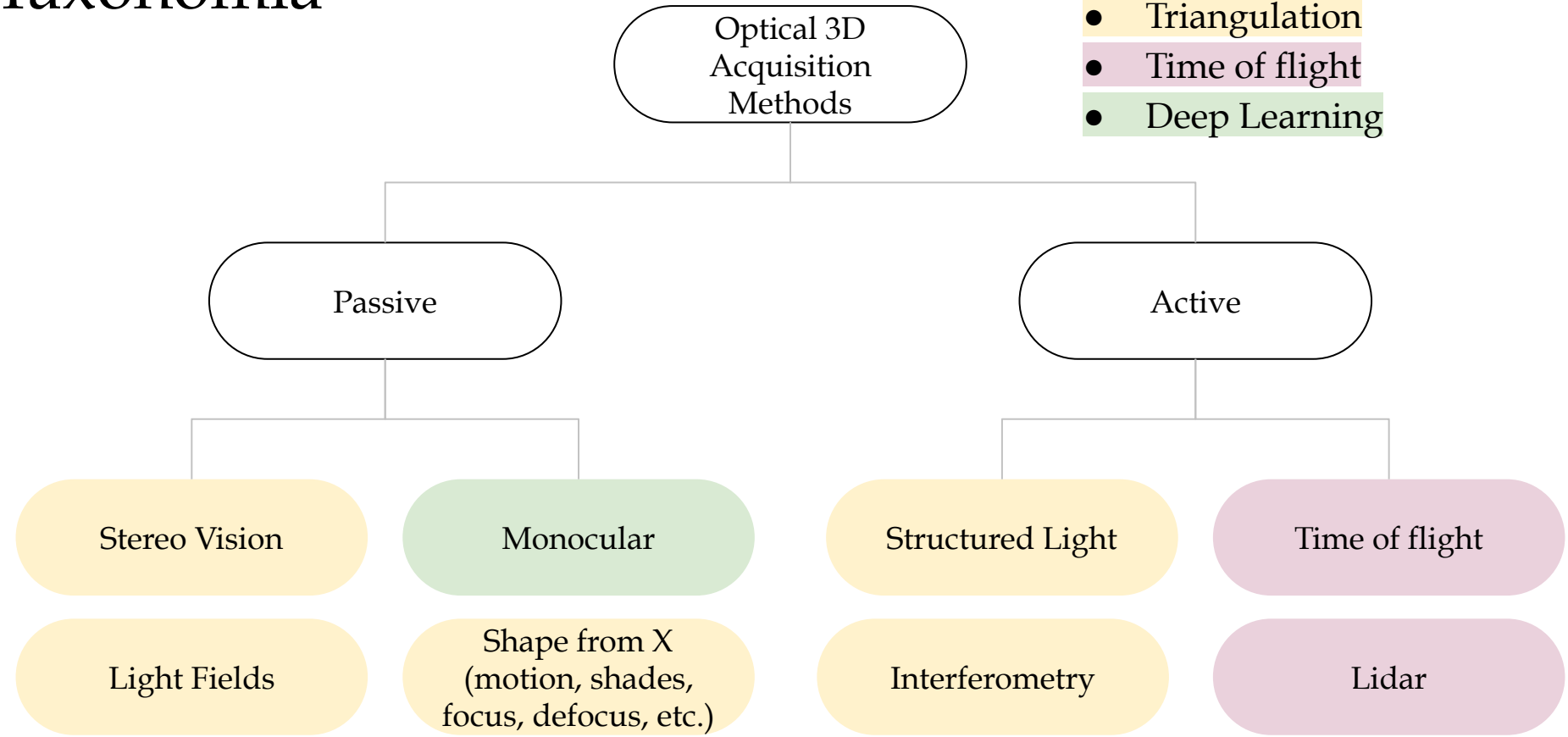
Vehículos autónomos

2. Técnicas de estimación de profundidad

Taxonomía

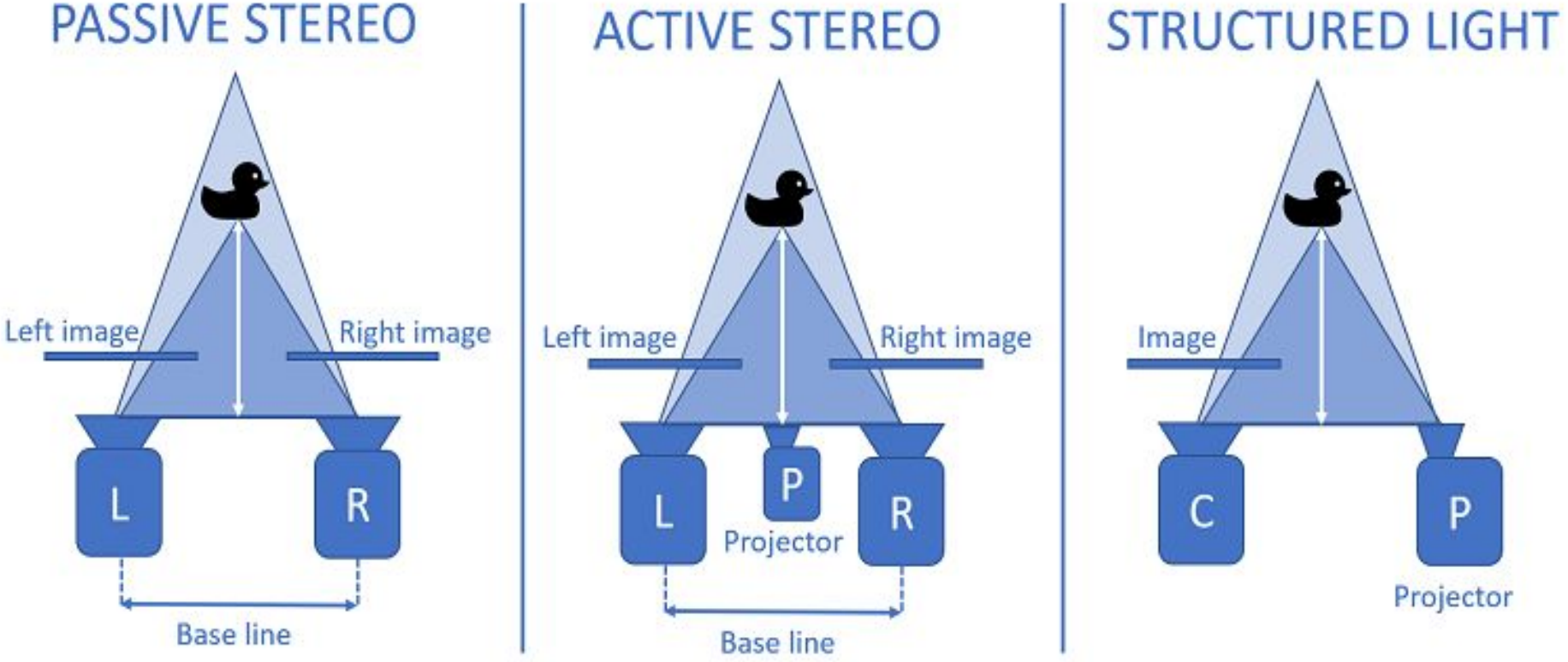
Depth estimation techniques:

- Triangulation
- Time of flight
- Deep Learning

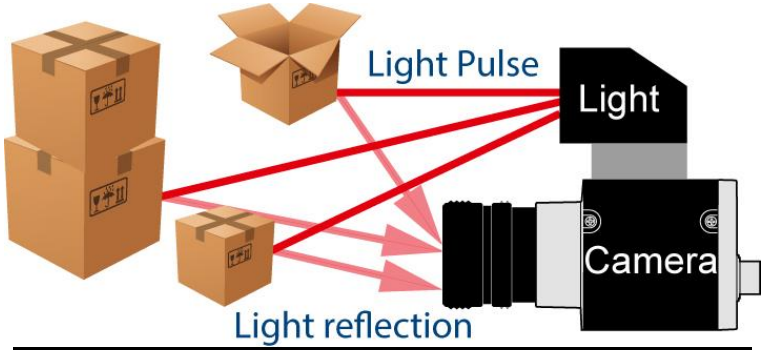


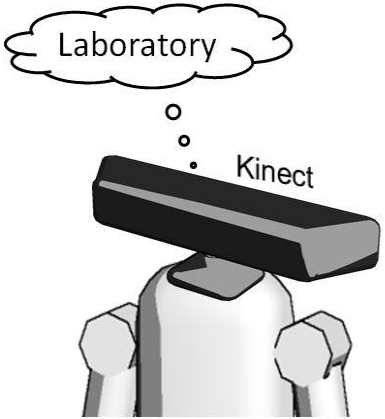
Pasivo: Estimación de la profundidad no requiere la emisión activa de señales o radiación hacia la escena

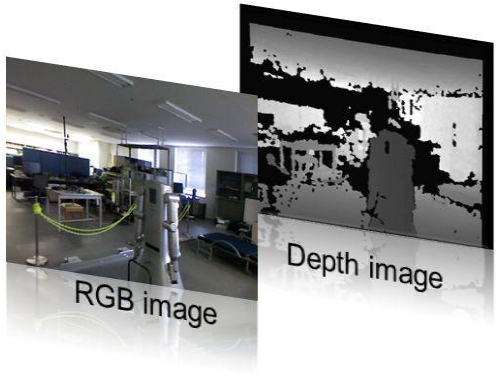
Basadas en Triangulación

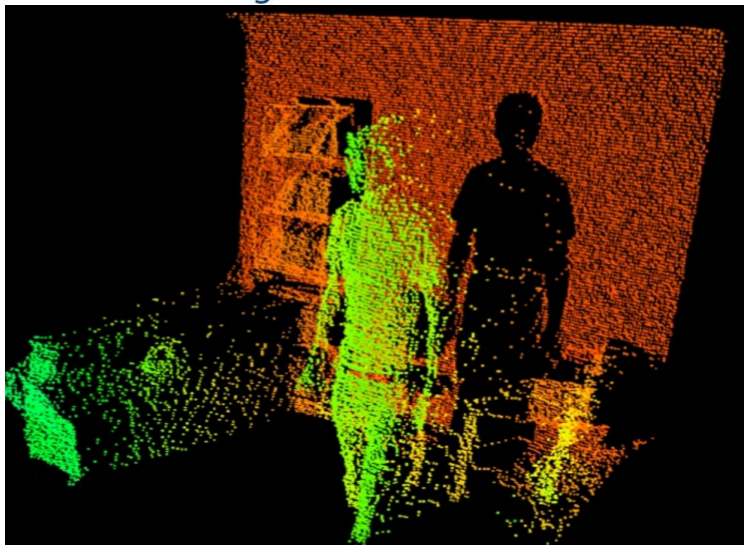


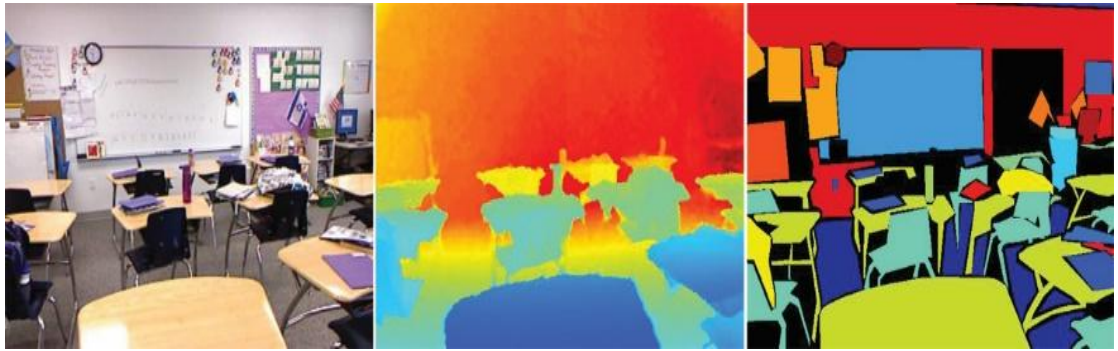
Basadas en Tiempo de Vuelo (Next HoCV session)





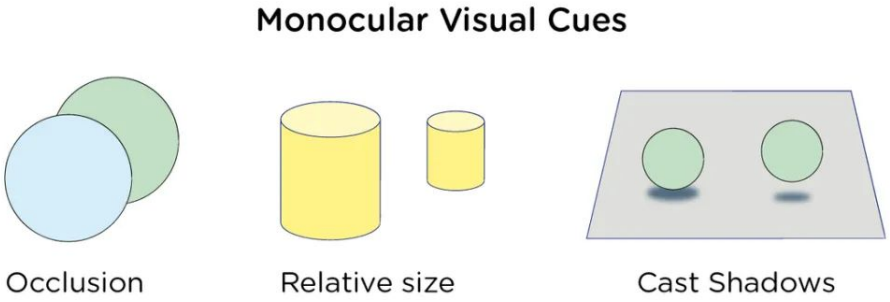


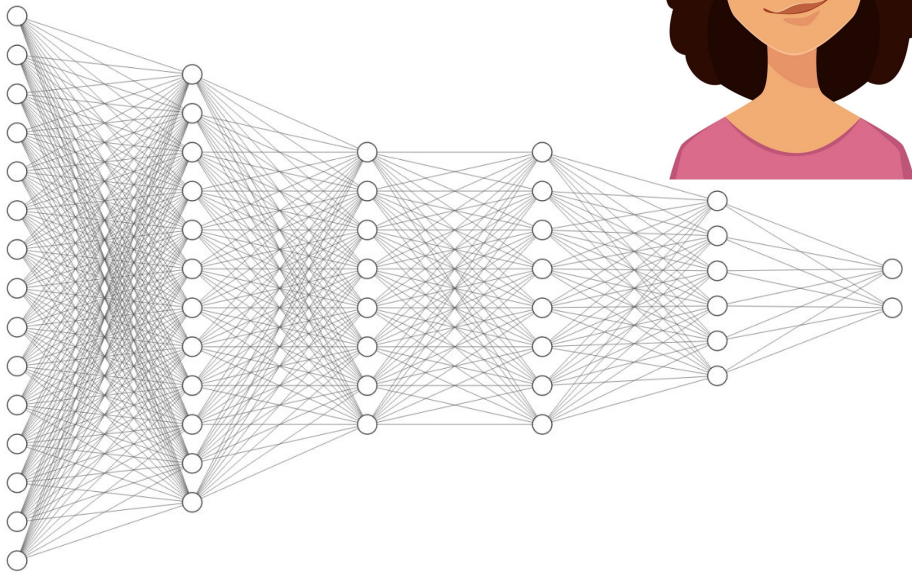


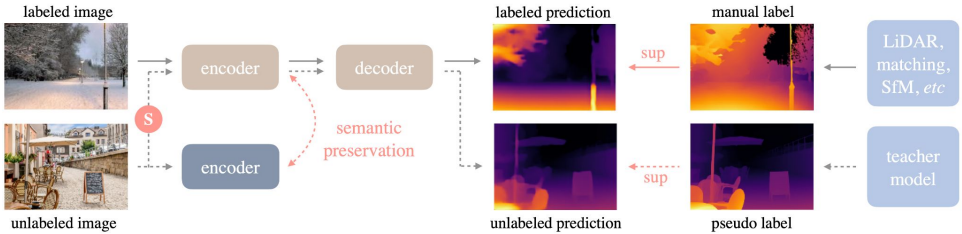


RGB Depth map Segmentation

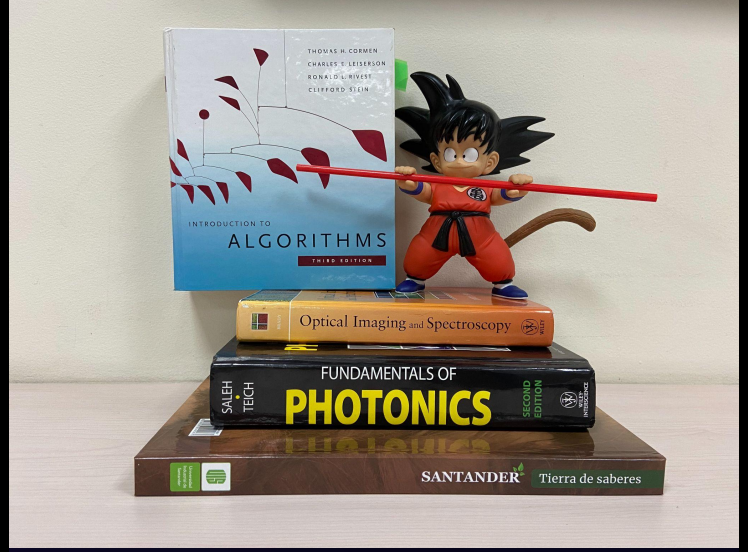
Basadas en Deep Learning





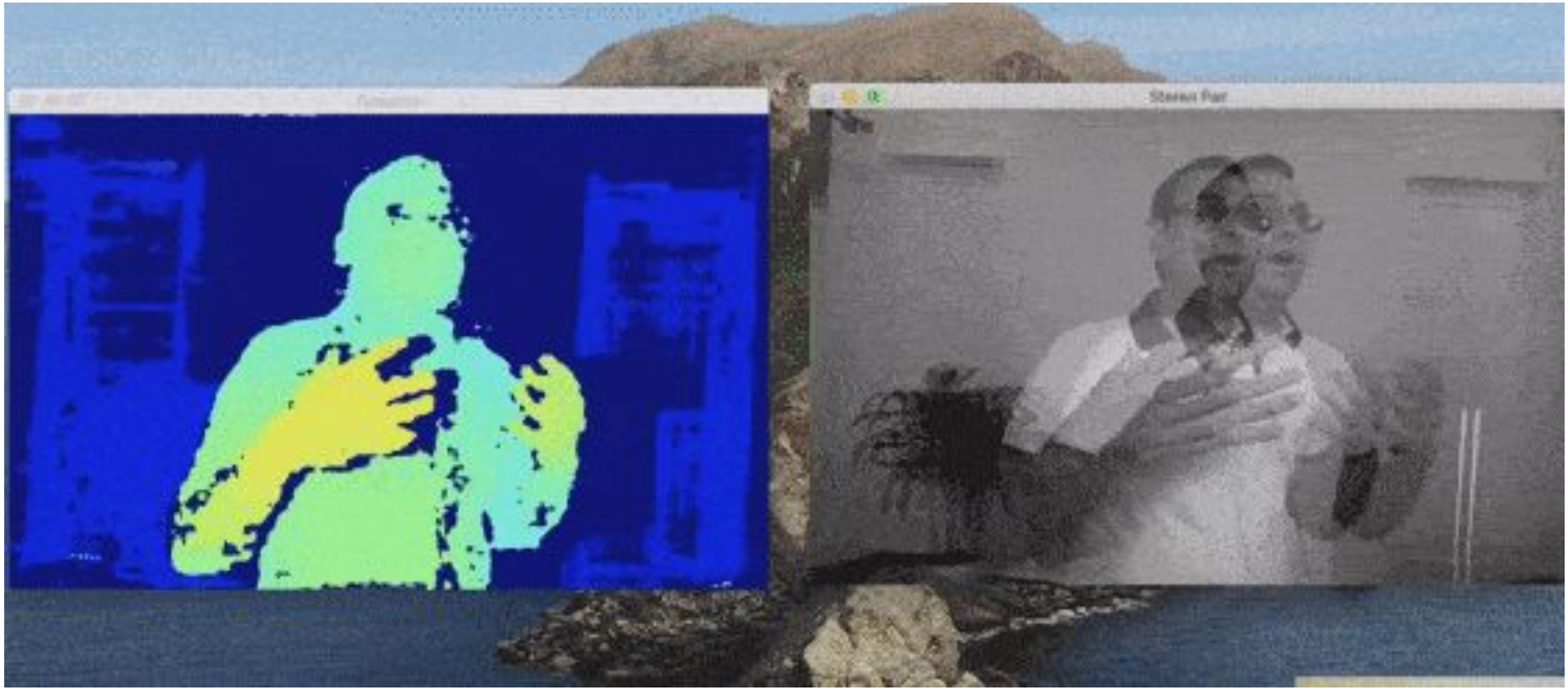


DEMO (Hugging Face)

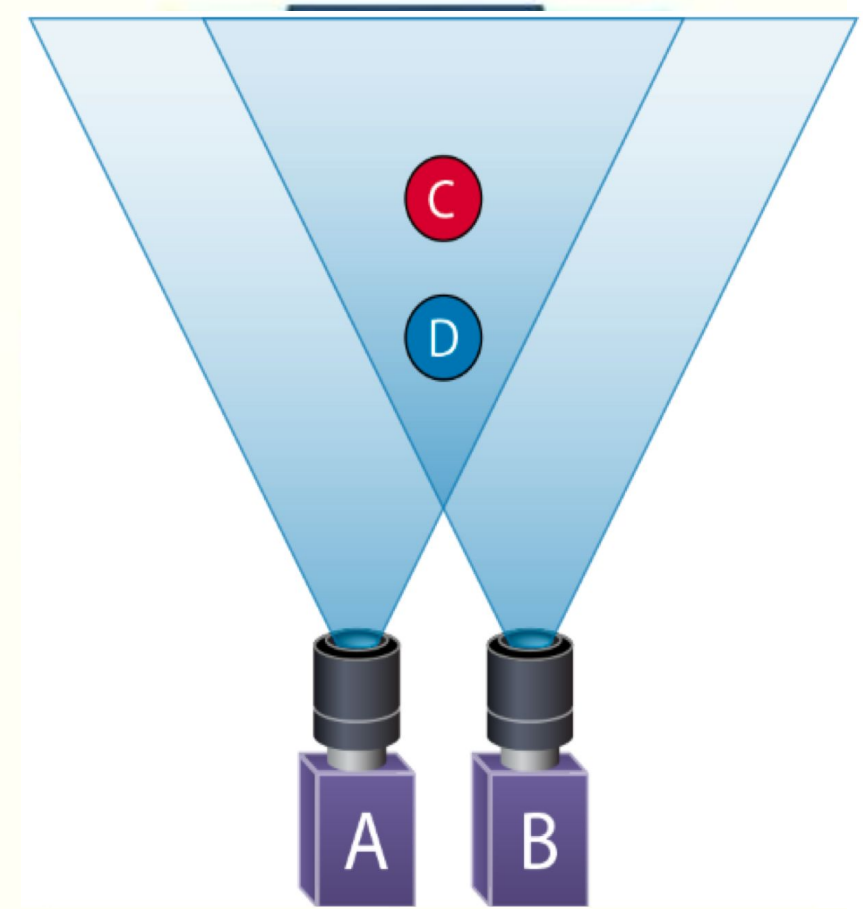


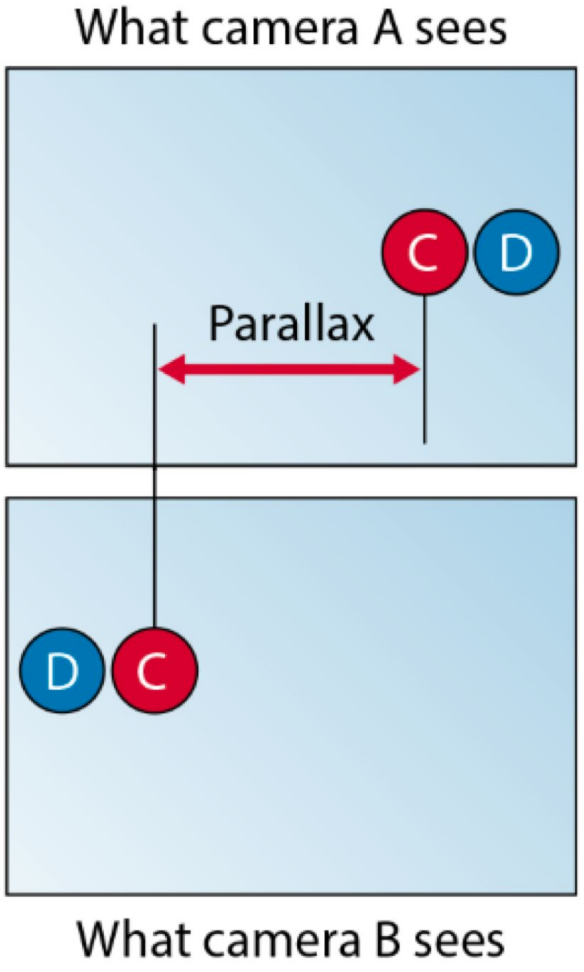
3. Visión Stereo (pasiva)

Visión Stereo (Estimar profundidad a partir de 2 imágenes)

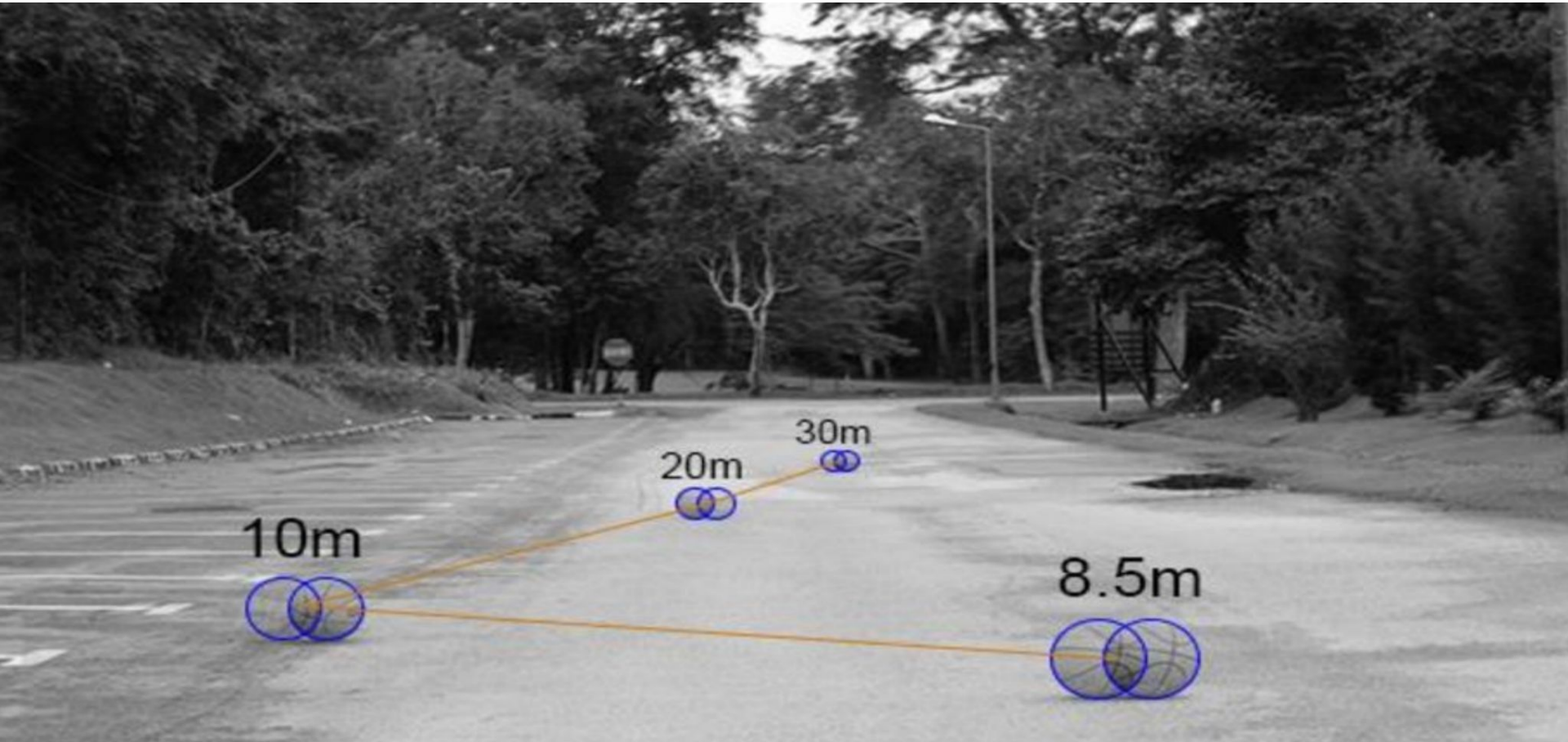


Parallax (Disparidad)

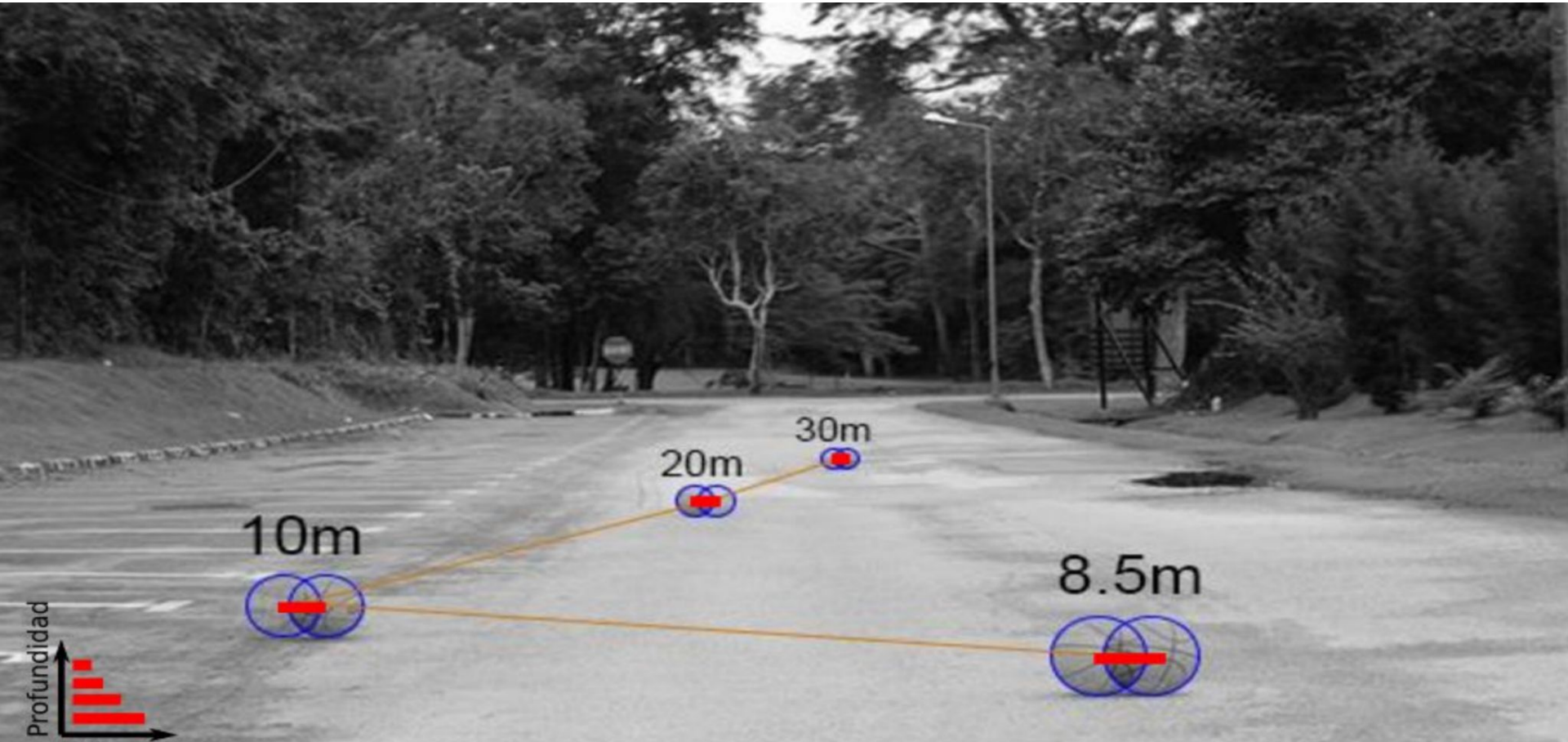




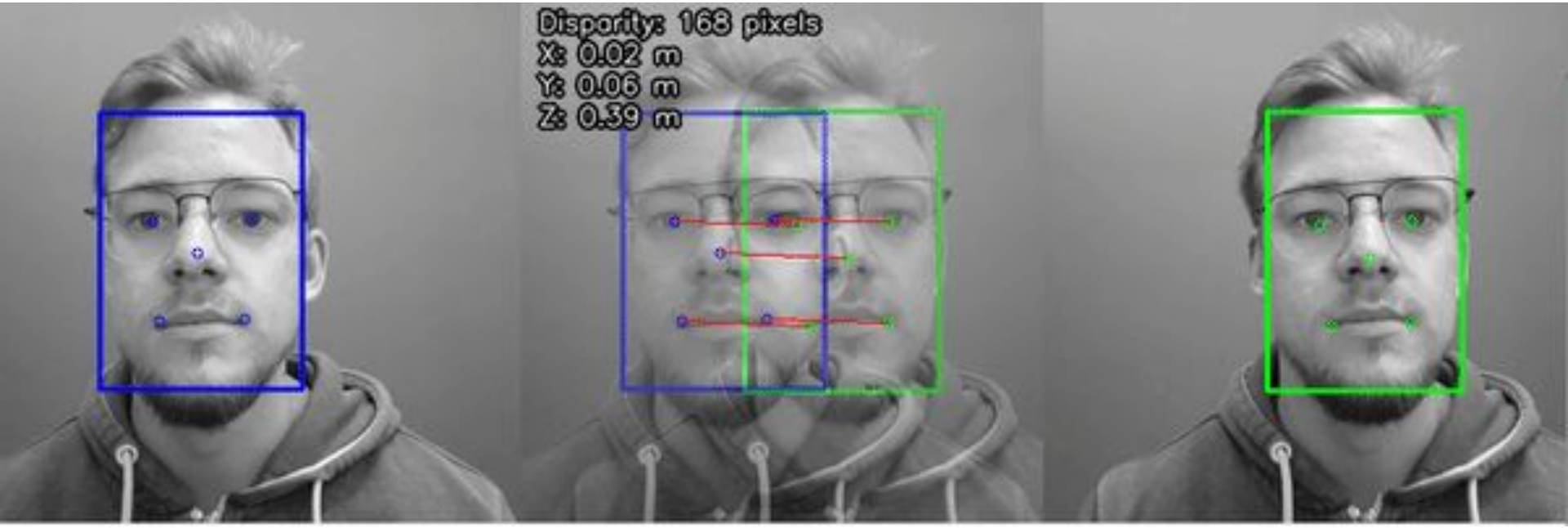
La disparidad **disminuye** con la distancia



La disparidad **disminuye** con la distancia



La disparidad **disminuye** con la distancia



Passive Stereo (Pipeline)

Calibration (offline)

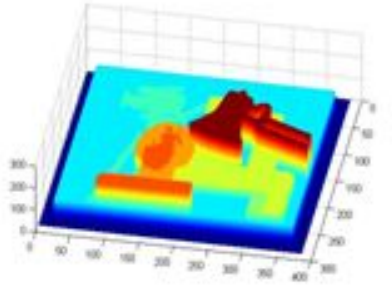
Rectification

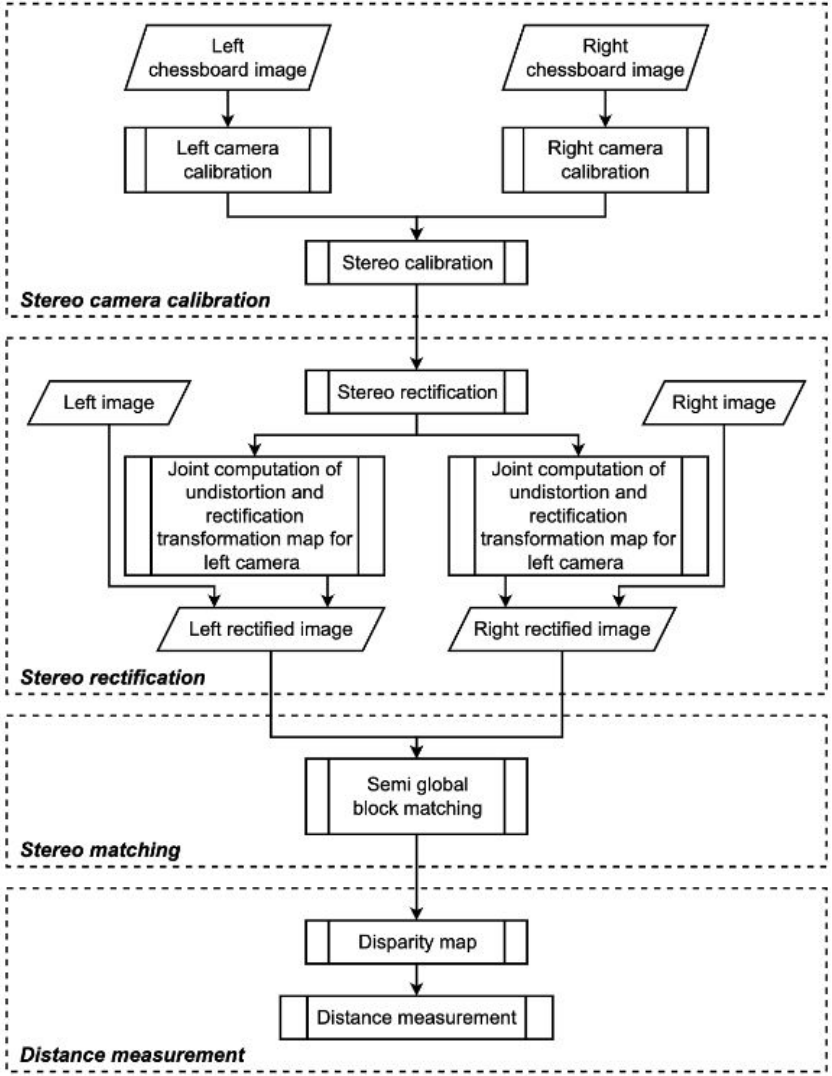
Stereo Correspondence

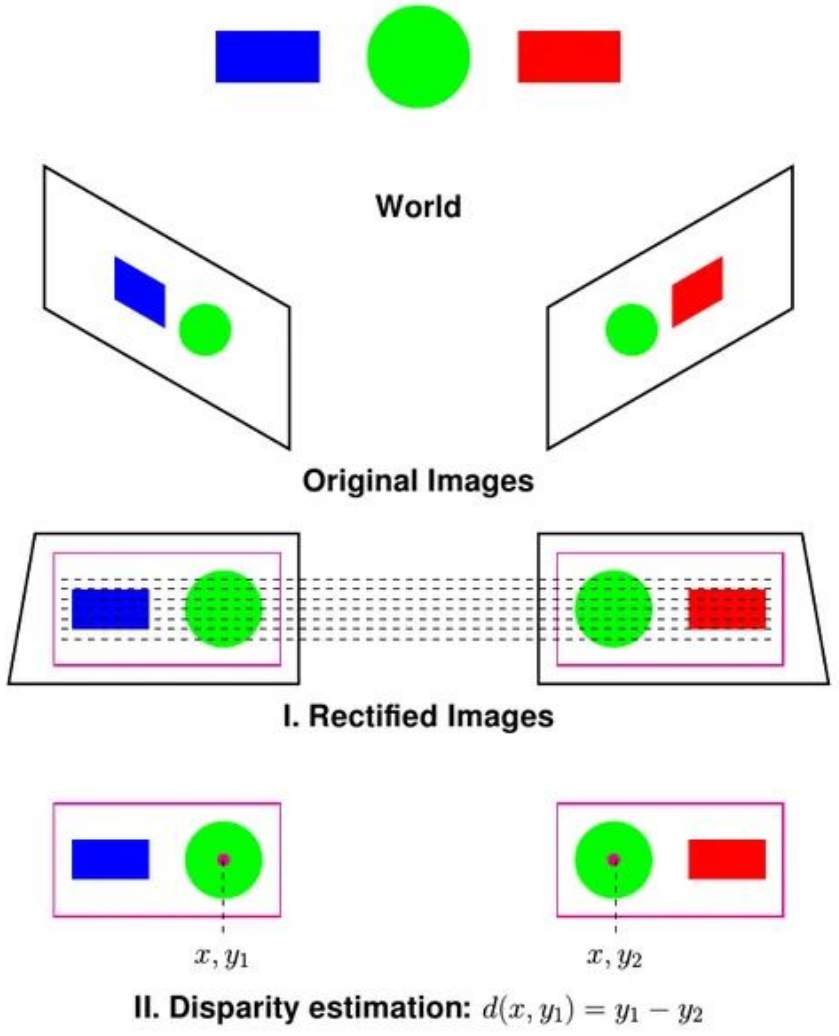
Triangulation

PC, FPGA, GPU

3D



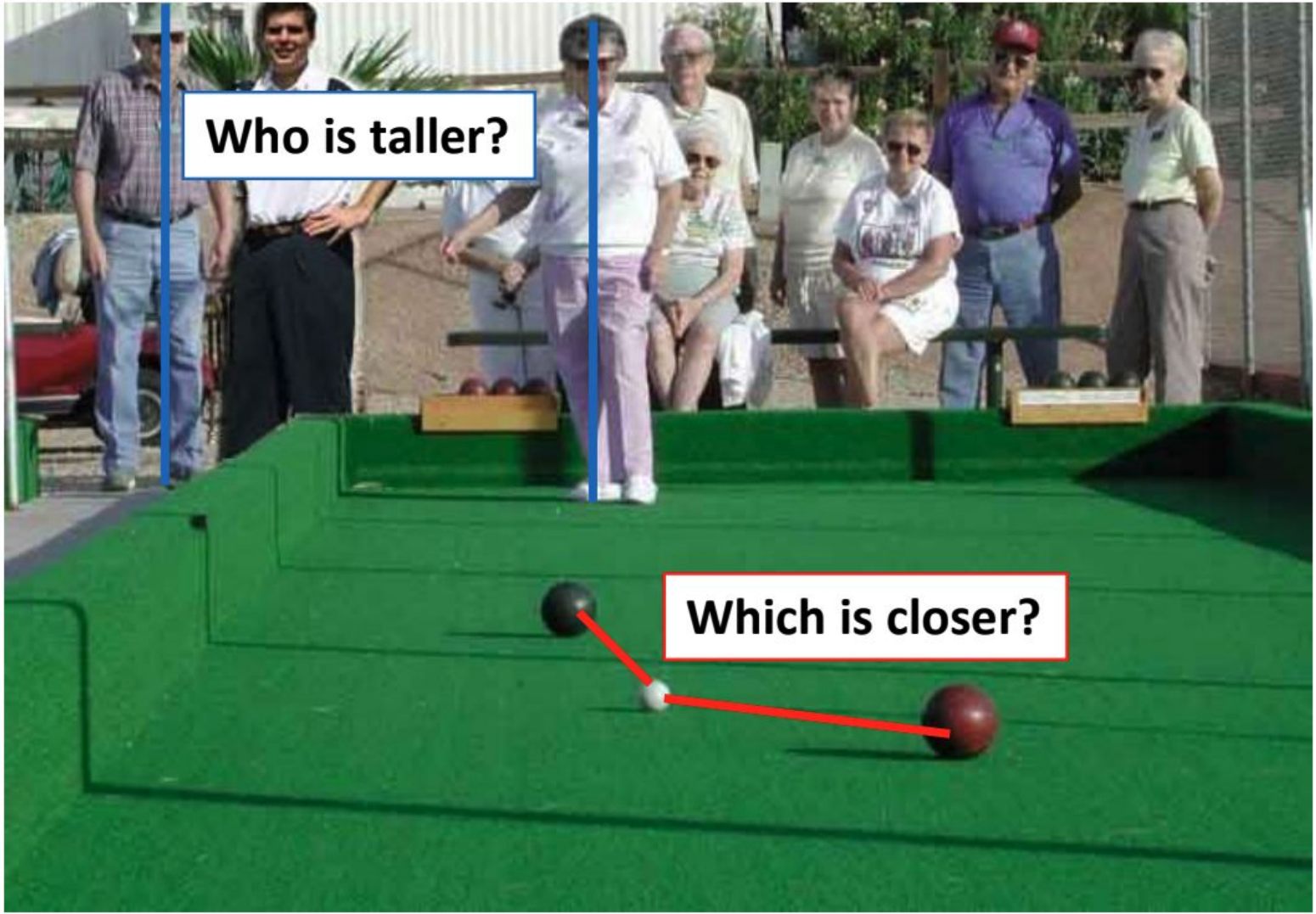

Left chessboard image
Right chessboard image
Left camera calibration
Right camera calibration
Stereo calibration
Stereo camera calibration
Stereo rectification
Left image
Right image
Joint computation of undistortion and rectification transformation map for left camera
Joint computation of undistortion and rectification transformation map for left camera
Left rectified image
Right rectified image
Stereo rectification
Semi global block matching
Stereo matching
Disparity map
Distance measurement
Distance measurement


World
Original Images
I. Rectified Images
x, y_1
x, y_2
II. Disparity estimation: $d(x, y_1) = y_1 - y_2$

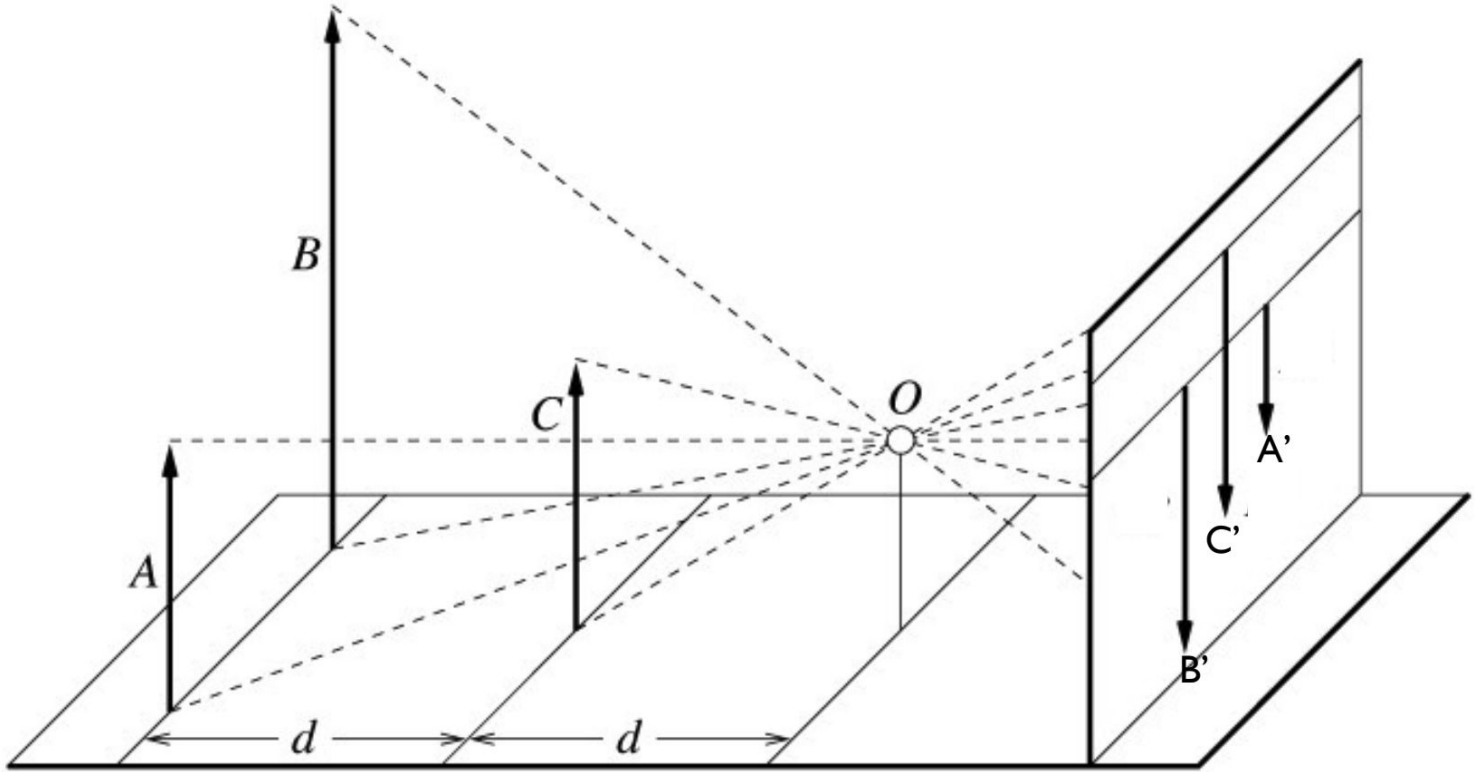
¿Cómo obtener la disparidad?

Proyección Perspectiva

Proyección
Perspectiva

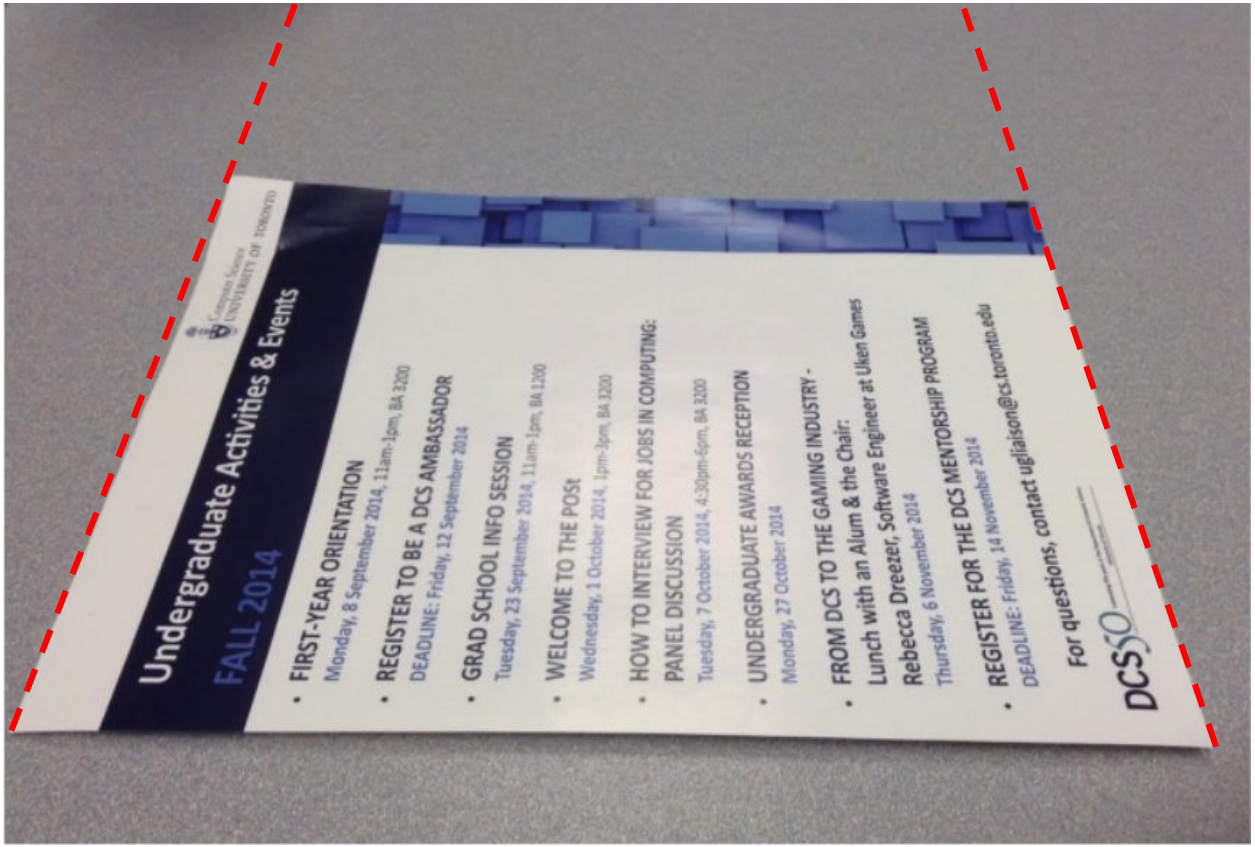

Who is taller?
Which is closer?

Geometría de Proyección Perspectiva



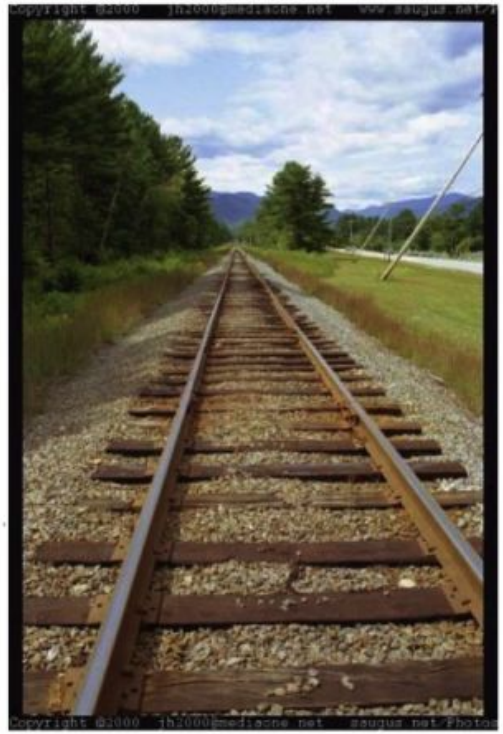
La longitud y el área no se conservan. A y C son del mismo tamaño, pero A se ve más pequeño en el sensor
De manera similar, B es mas grande que C, pero aparecen del mismo tamaño en el sensor

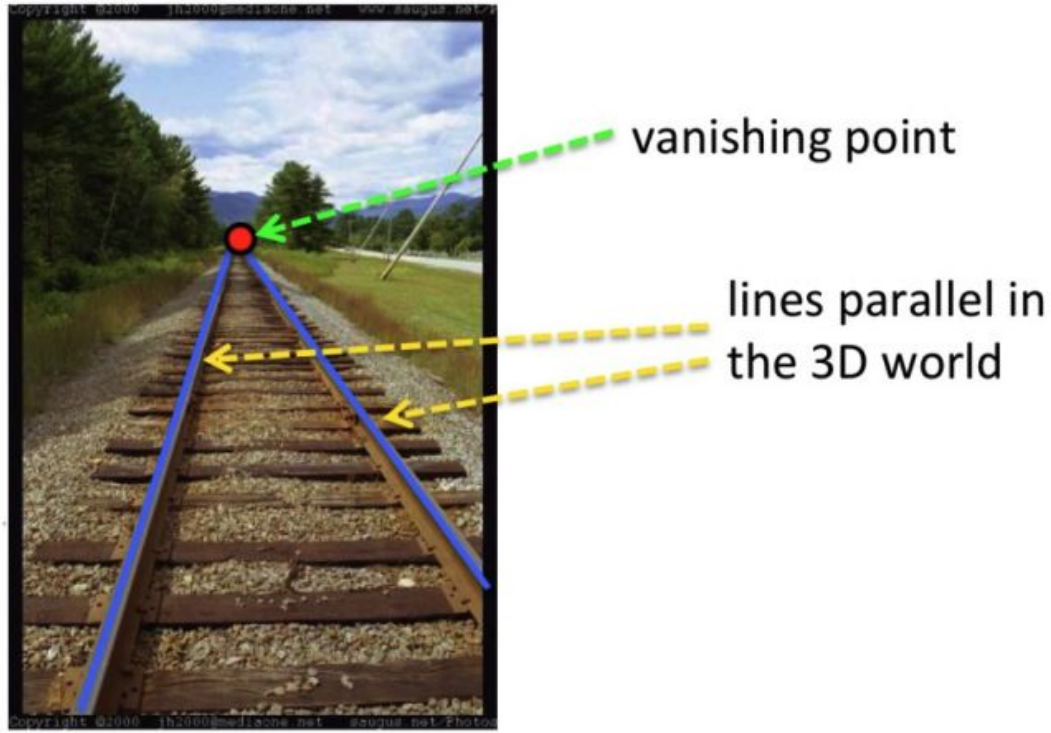
Vanishing Points



Líneas paralelas en el mundo 3D convergen a un punto en 2D (vanishing point)

Vanishing Points



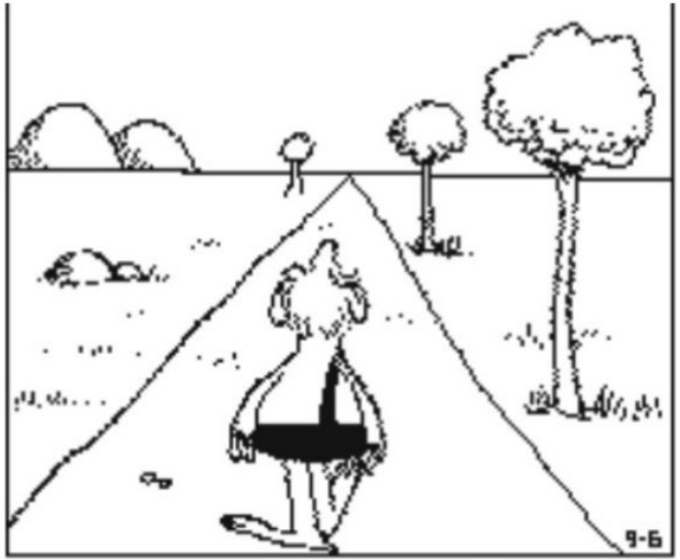


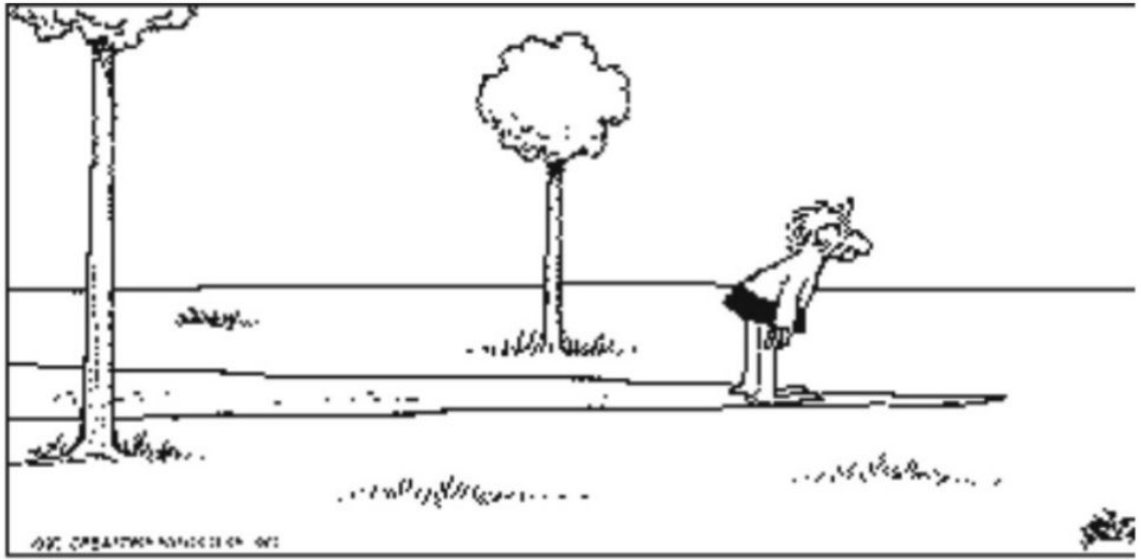
Cada ángulo de dirección diferente en el mundo tiene su propio “vanishing point”

Vanishing Points

Todas las líneas con la misma dirección en el mundo 3D se intersectan en el mismo vanishing point

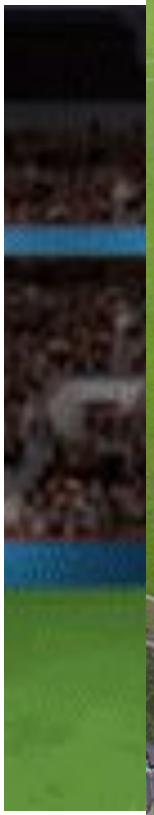
Vanishing Points

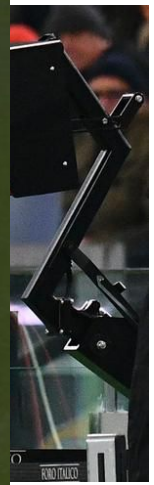




Lo contrario no siempre es verdad: Líneas que se intersectan en 2D no siempre corresponden a líneas paralelas en 3D

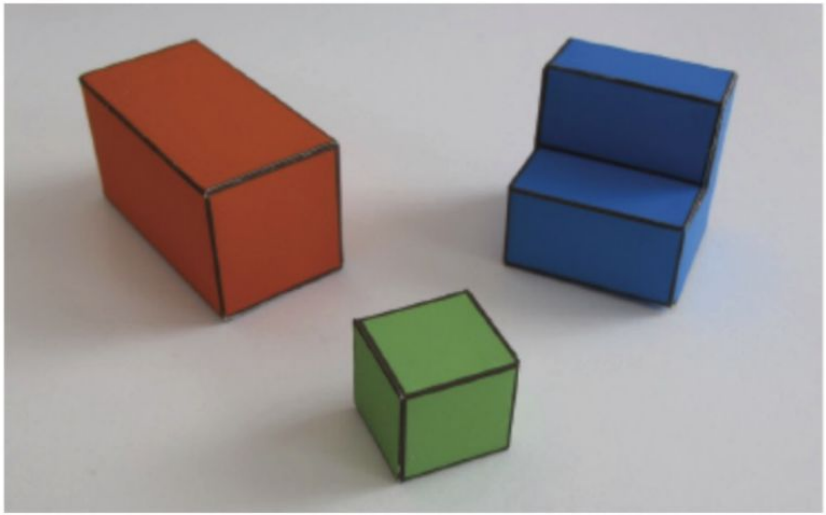
Inter



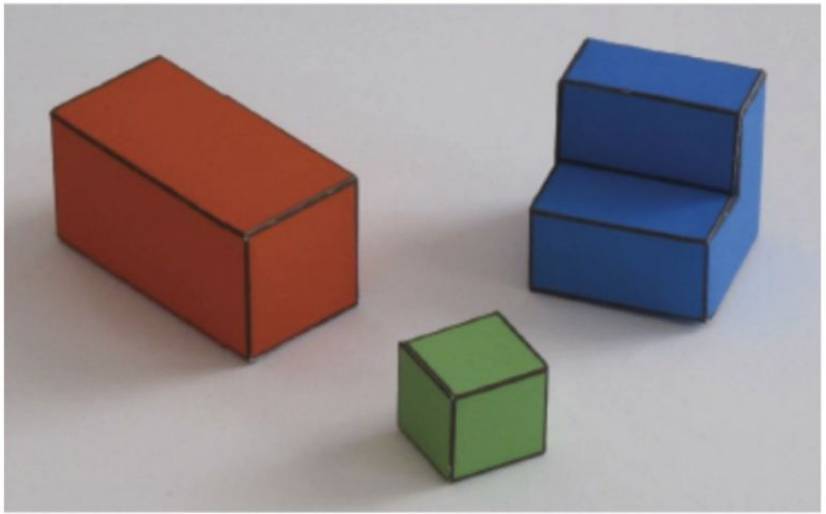



FIFA WORLD CUP
Qatar2022
MITOMA
9

Tipos de Proyecciones: Perspectiva vs Ortográfica



Perspective projection

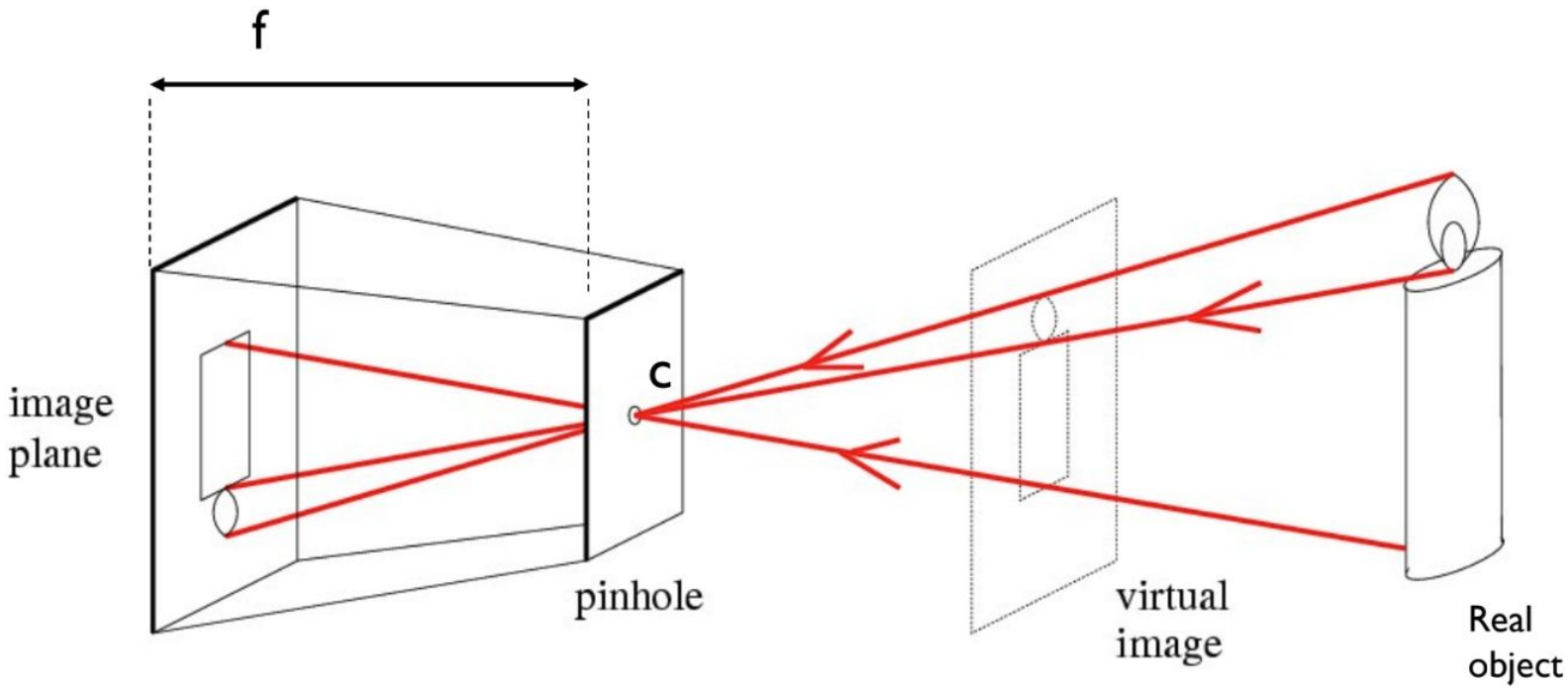


Parallel (orthographic) projection

- For perspective projection lines parallel in 3D **are not** parallel in the image.
- For orthographic projection lines parallel in 3D **are** parallel in the image.

[Source: A. Torralba]

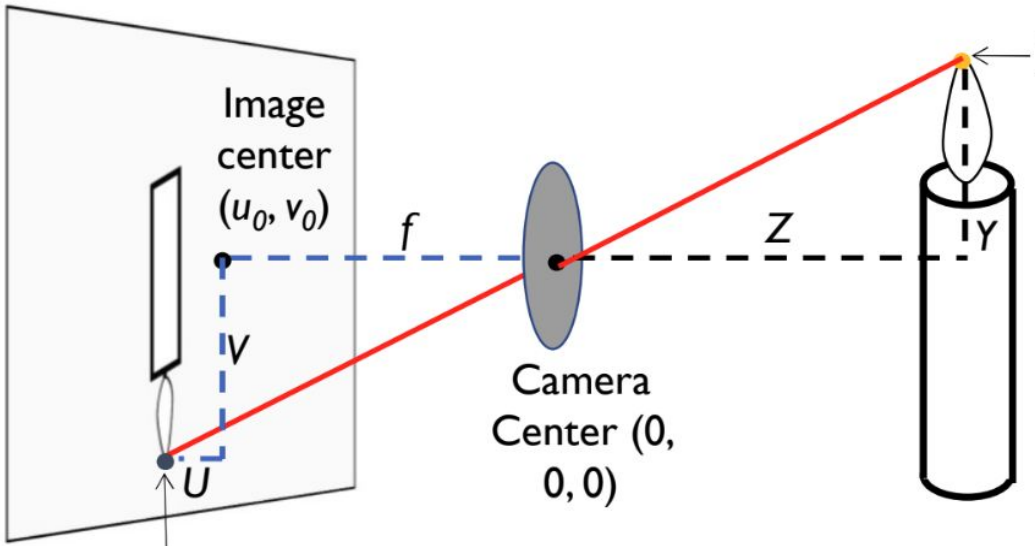
Pinhole Camera Model



f = Focal length

c = Optical center of the camera

Geometría de Proyección Perspectiva



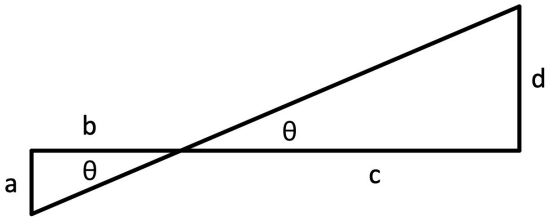
$$\mathbf{P} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$\mathbf{p} = \begin{bmatrix} U \\ V \end{bmatrix}$$

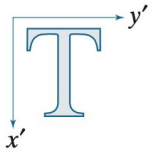
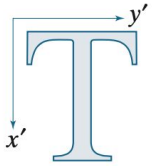
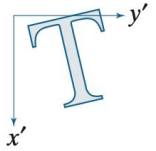
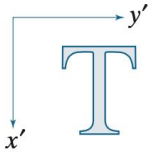
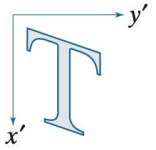
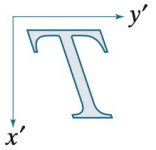
p = distance from image center

$$U = -X * \frac{f}{Z} \qquad V = -Y * \frac{f}{Z}$$

Recordando triángulos



$$\frac{a}{b} = \frac{d}{c} \longrightarrow a = \frac{bd}{c}$$

Transformation Name	Affine Matrix, A	Coordinate Equations	Example
Identity	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x$ $y' = y$	
Scaling/Reflection (For reflection, set one scaling factor to –1 and the other to 0)	$\begin{bmatrix} c_x & 0 & 0 \\ 0 & c_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = c_x x$ $y' = c_y y$	
Rotation (about the origin)	$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x\cos\theta - y\sin\theta$ $y' = x\sin\theta + y\cos\theta$	
Translation	$\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x + t_x$ $y' = y + t_y$	
Shear (vertical)	$\begin{bmatrix} 1 & s_v & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x + s_v y$ $y' = y$	
Shear (horizontal)	$\begin{bmatrix} 1 & 0 & 0 \\ s_h & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x$ $y' = s_h x + y$	

Geometría de Proyección Perspectiva: Ecuaciones

Una cámara y su posición en el espacio está descrita por varios parámetros:

- Translación (T) del centro óptico de la imagen al de las coordenadas reales
- Rotación (R) del plano imagen
- Distancia focal (f), punto principal (u_0, v_0), tamaño de pixel (s_x, s_y)
- Azul = Extrínsecos; Rojo = Intrínsecos

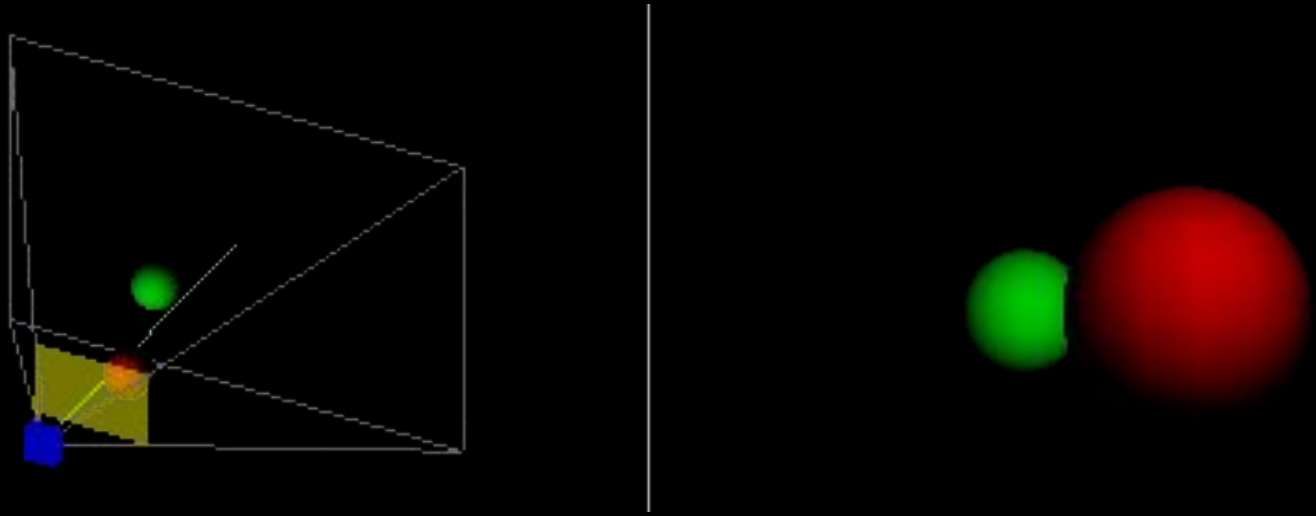
$$\mathbf{p} = \mathbf{K}[\mathbf{R} \;\; \mathbf{t}]\mathbf{P}$$

$$w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

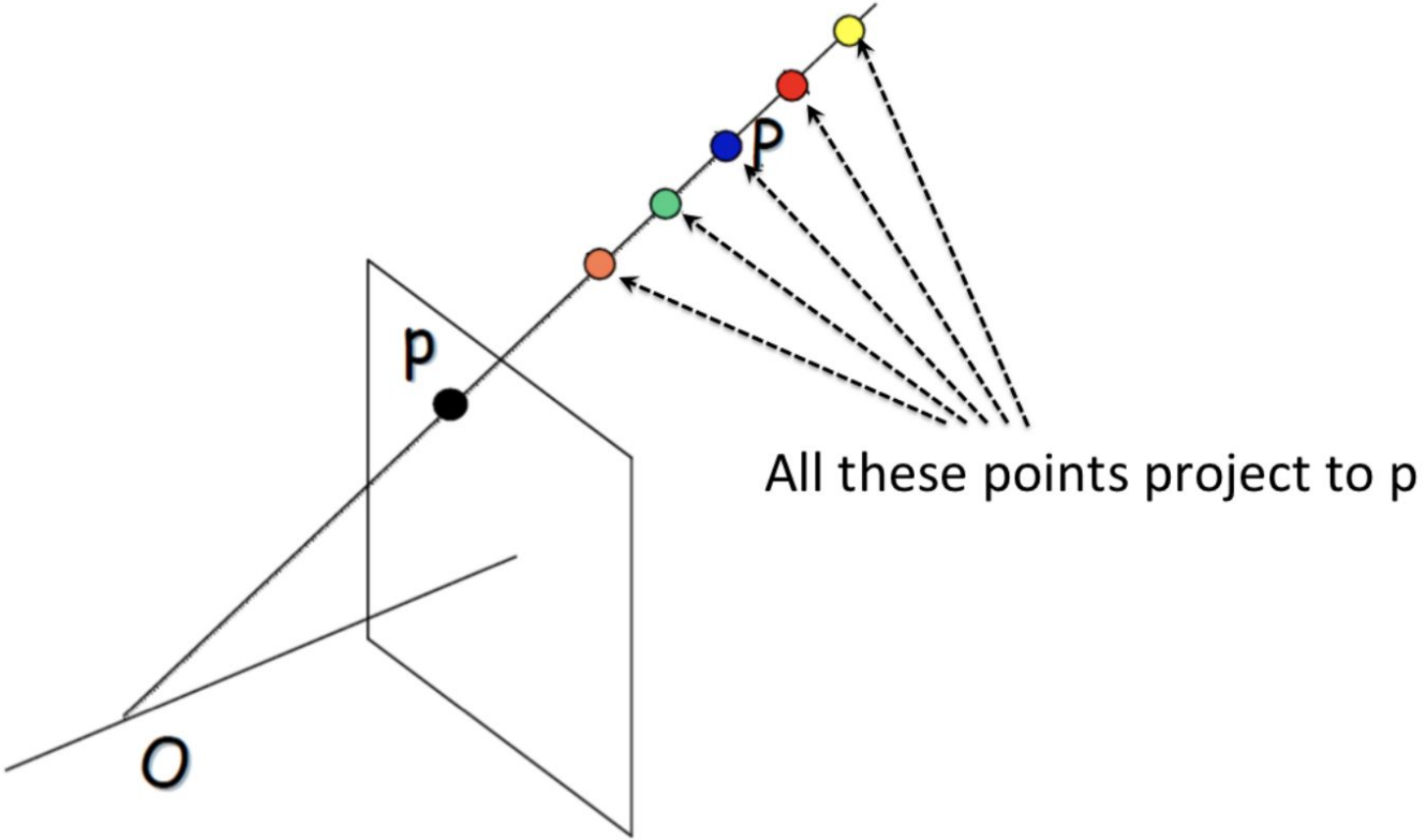
Intrínsecos Extrínsecos

Jugando con los parámetros Intrínsecos y Extrínsecos

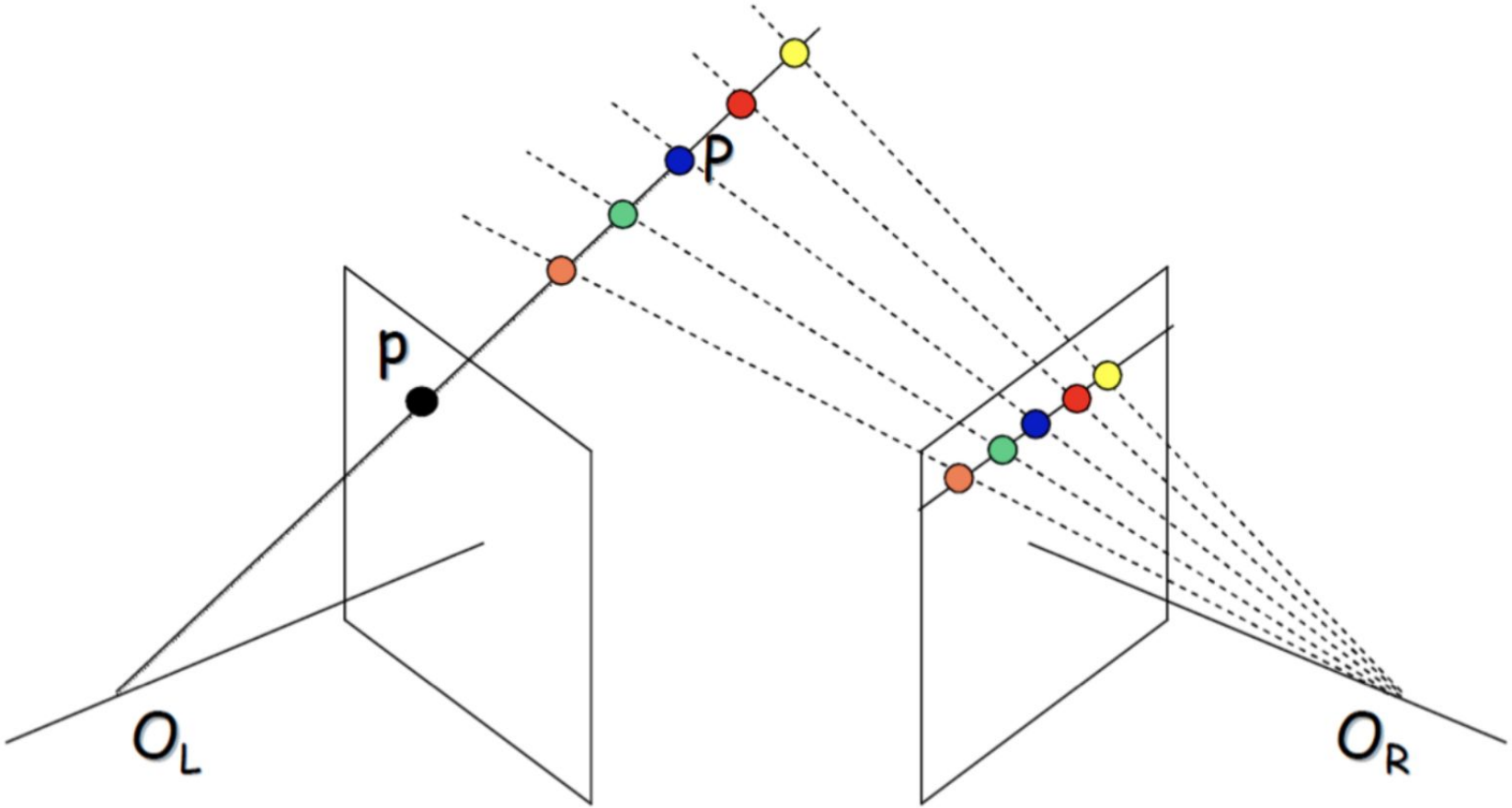
https://ksimek.github.io/perspective_camera_toy.html



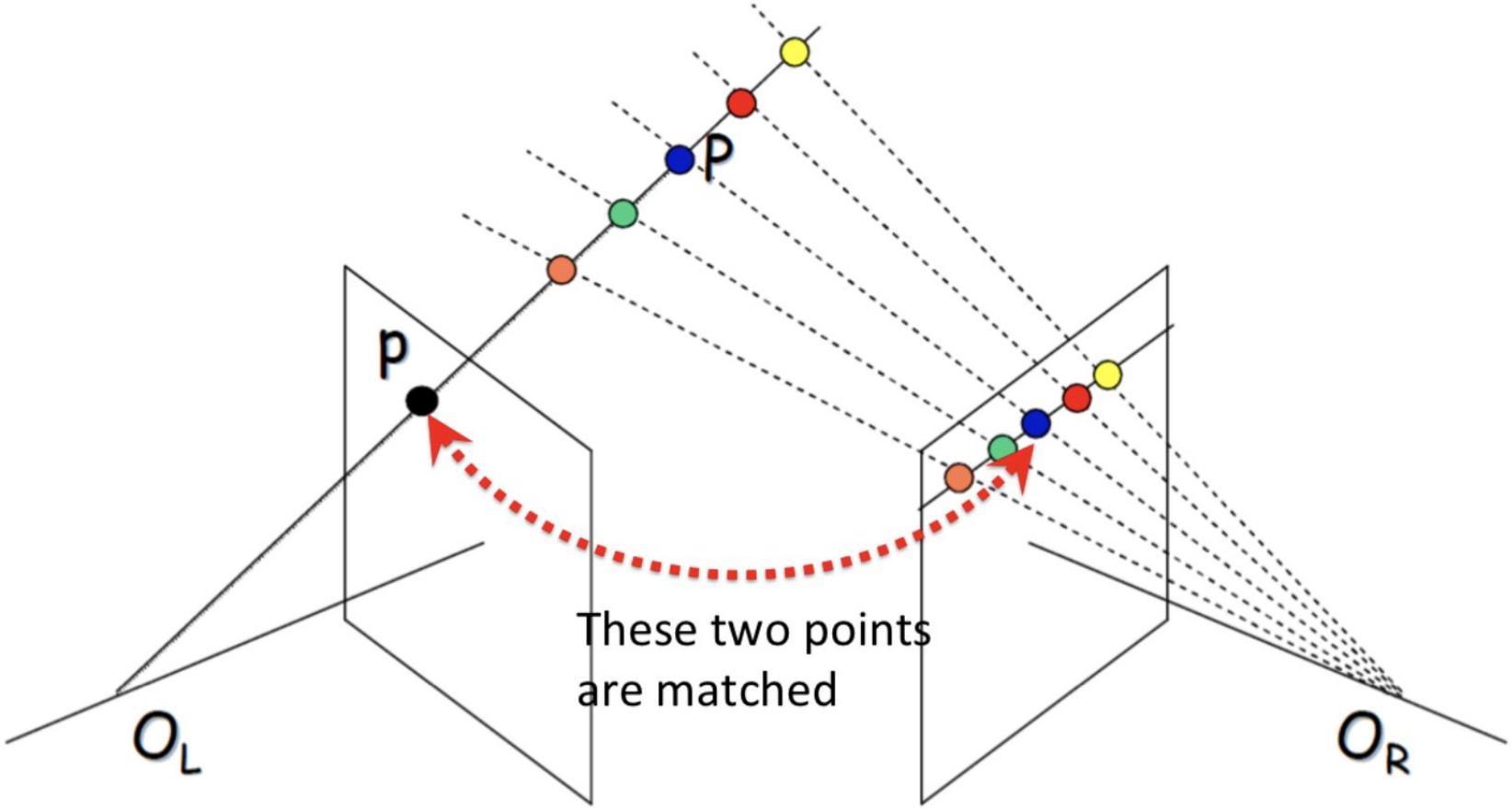
Geometría de Visión Stereo



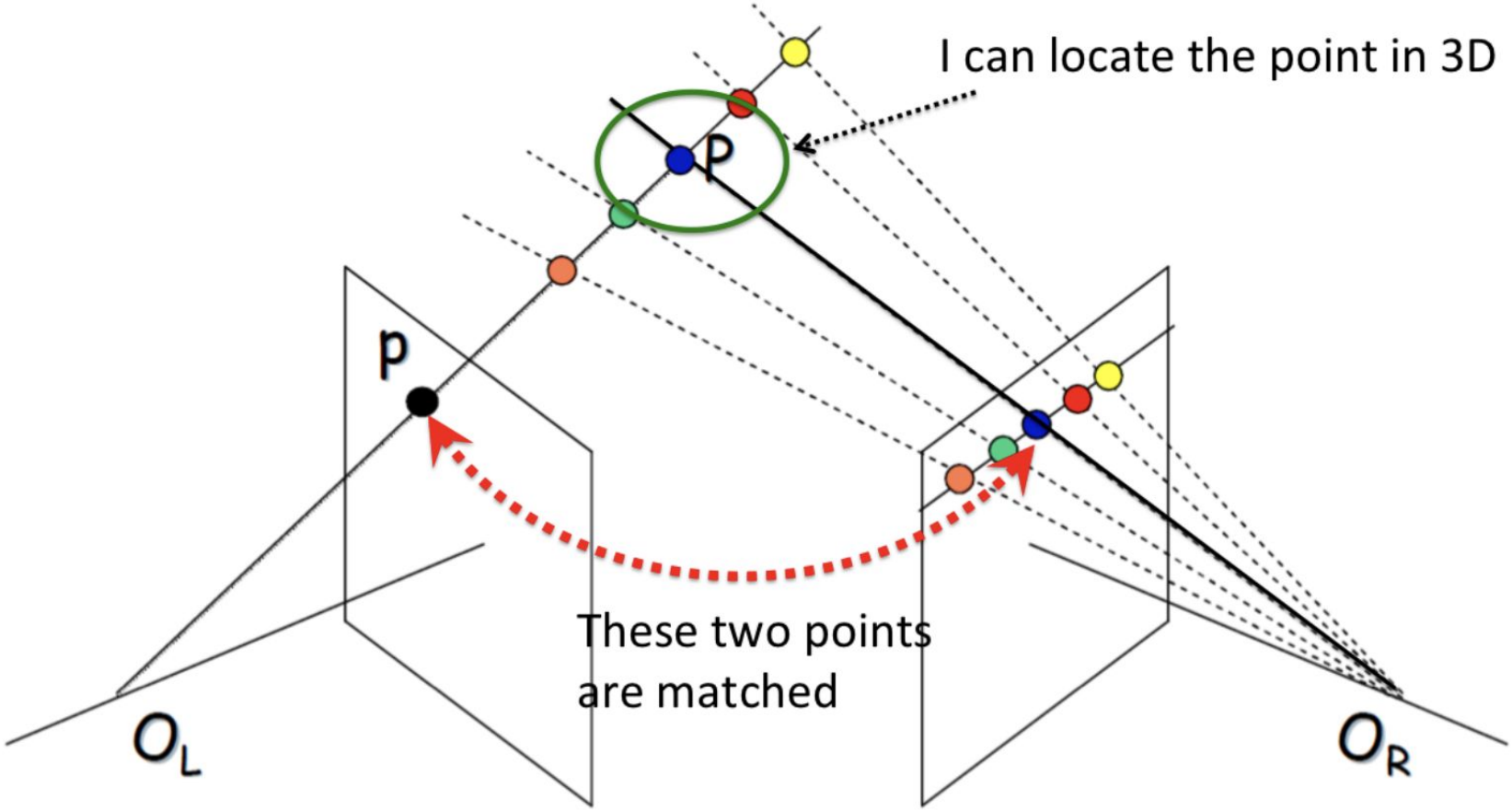
Geometría de Visión Stereo



Geometría de Visión Stereo

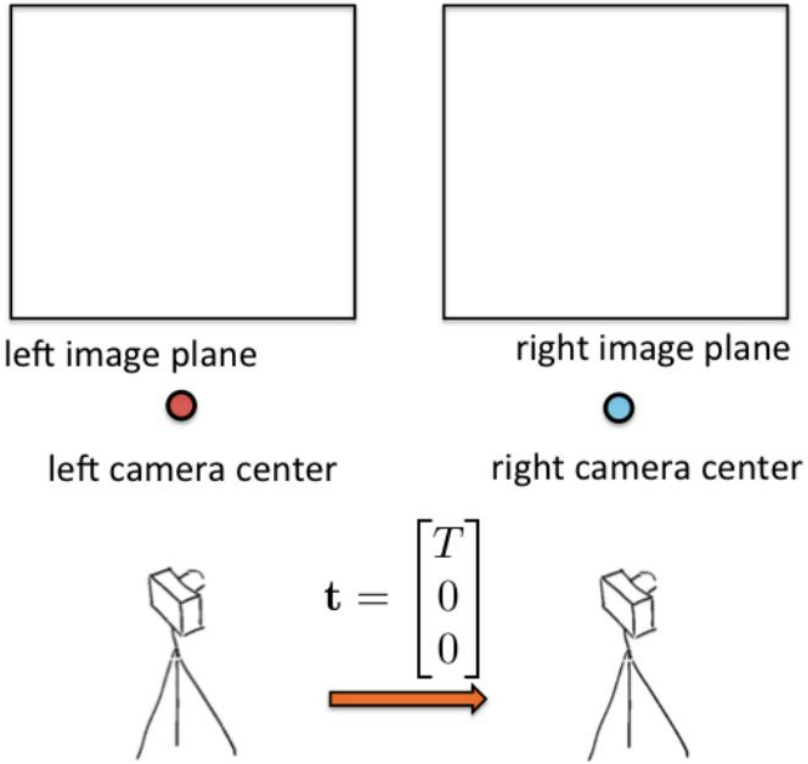


Geometría de Visión Stereo

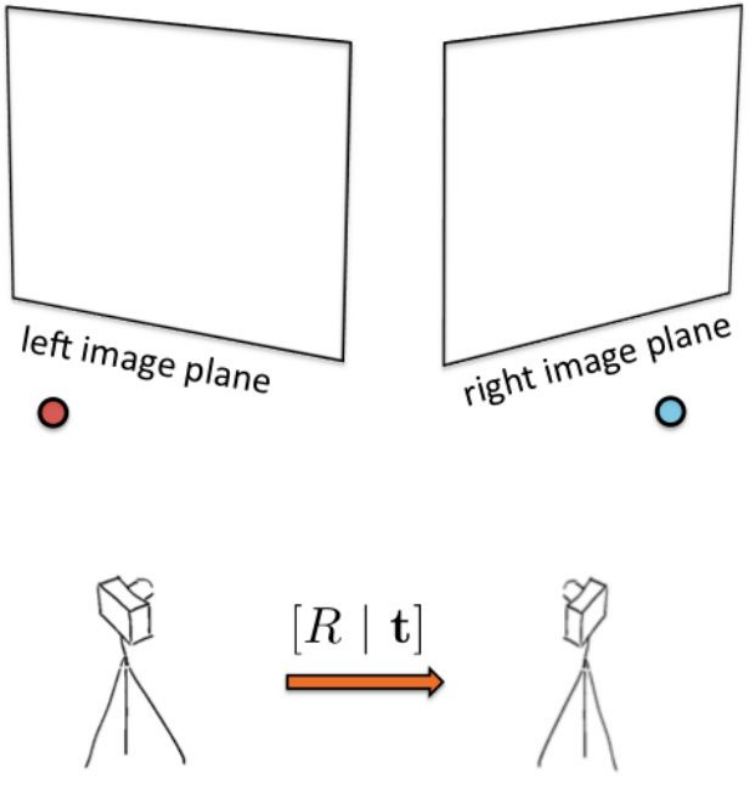


Geometría de Visión Stereo

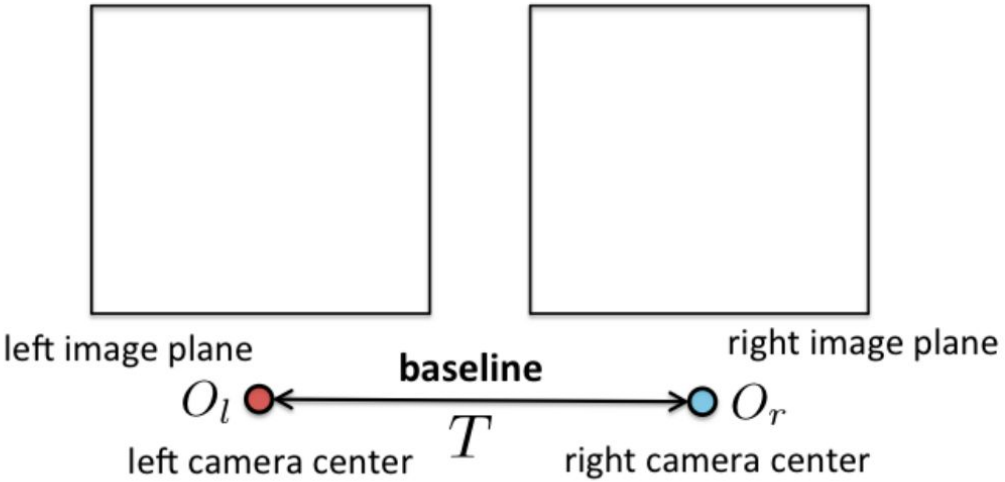
Paralelas (Caso simple)

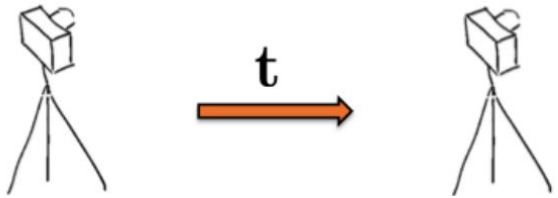


General (Rectificar)



Caso simple

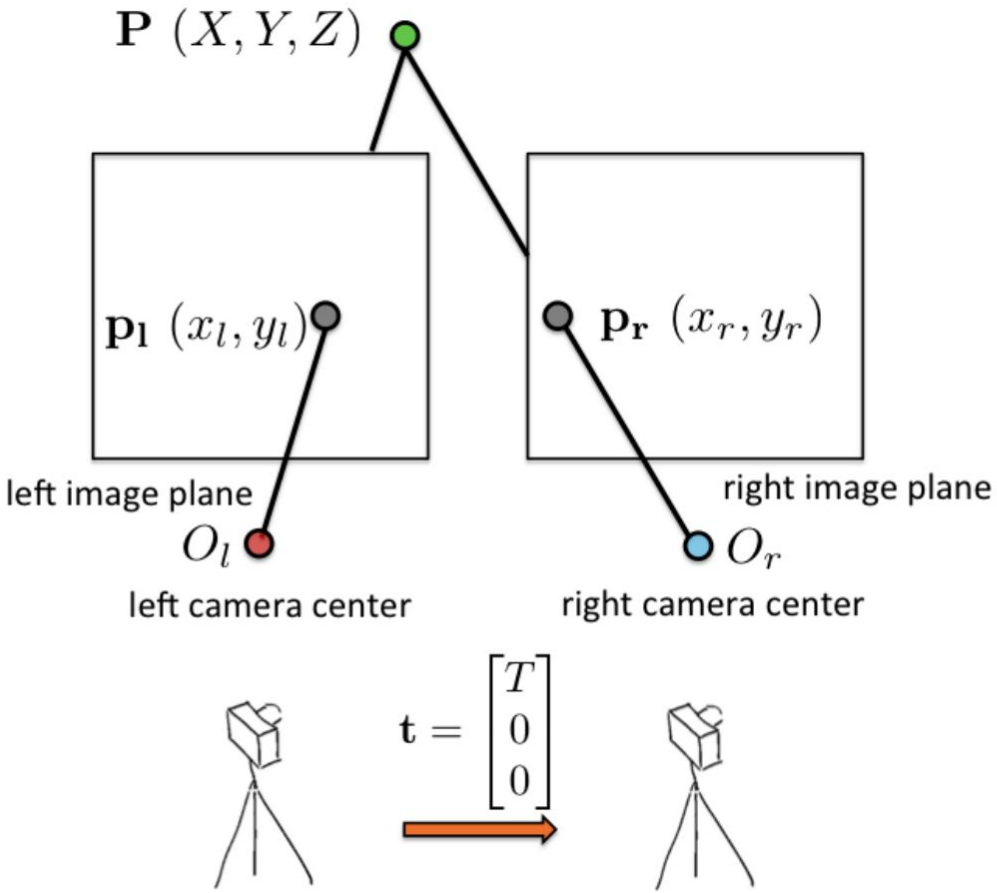




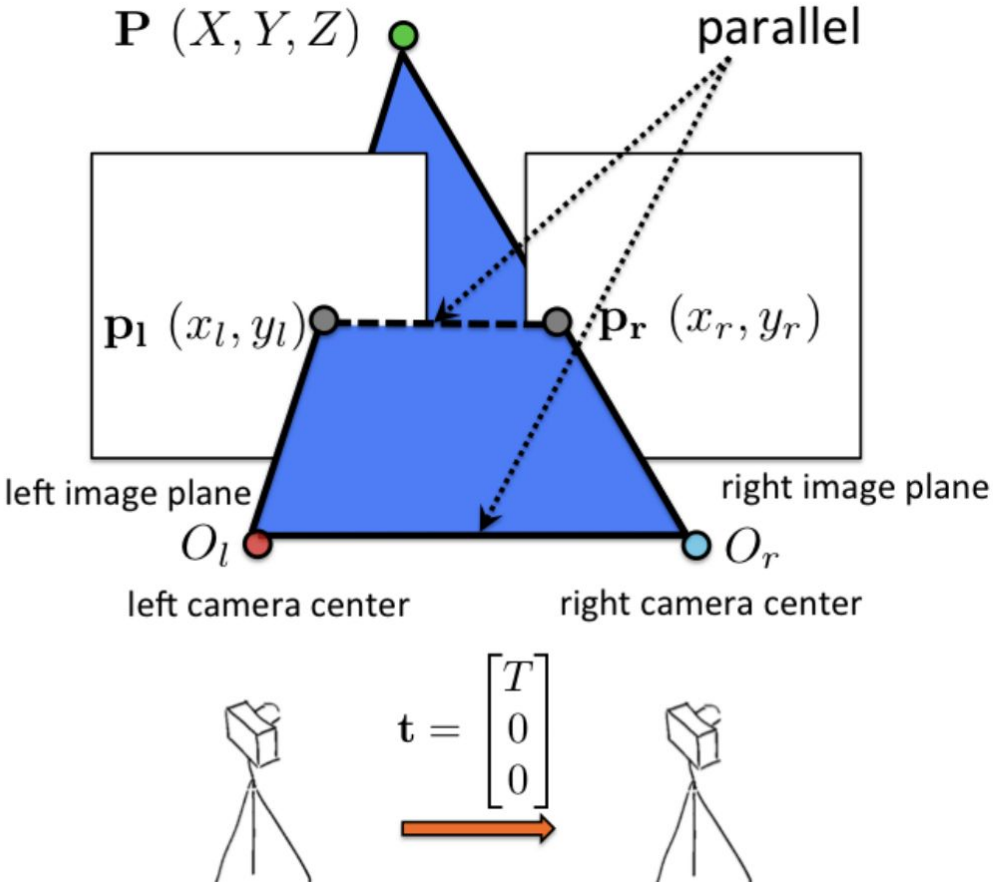
$$\mathbf{t} = \begin{bmatrix} T \\ 0 \\ 0 \end{bmatrix}$$

The right camera is shifted to the right in X direction

Caso simple


$\mathbf{P}\ (X, Y, Z)$
$\mathbf{p_l}\ (x_l, y_l)$
$\mathbf{p_r}\ (x_r, y_r)$
left image plane
right image plane
O_l
O_r
left camera center
right camera center
$\mathbf{t} = \begin{bmatrix} T \\ 0 \\ 0 \end{bmatrix}$

Caso simple

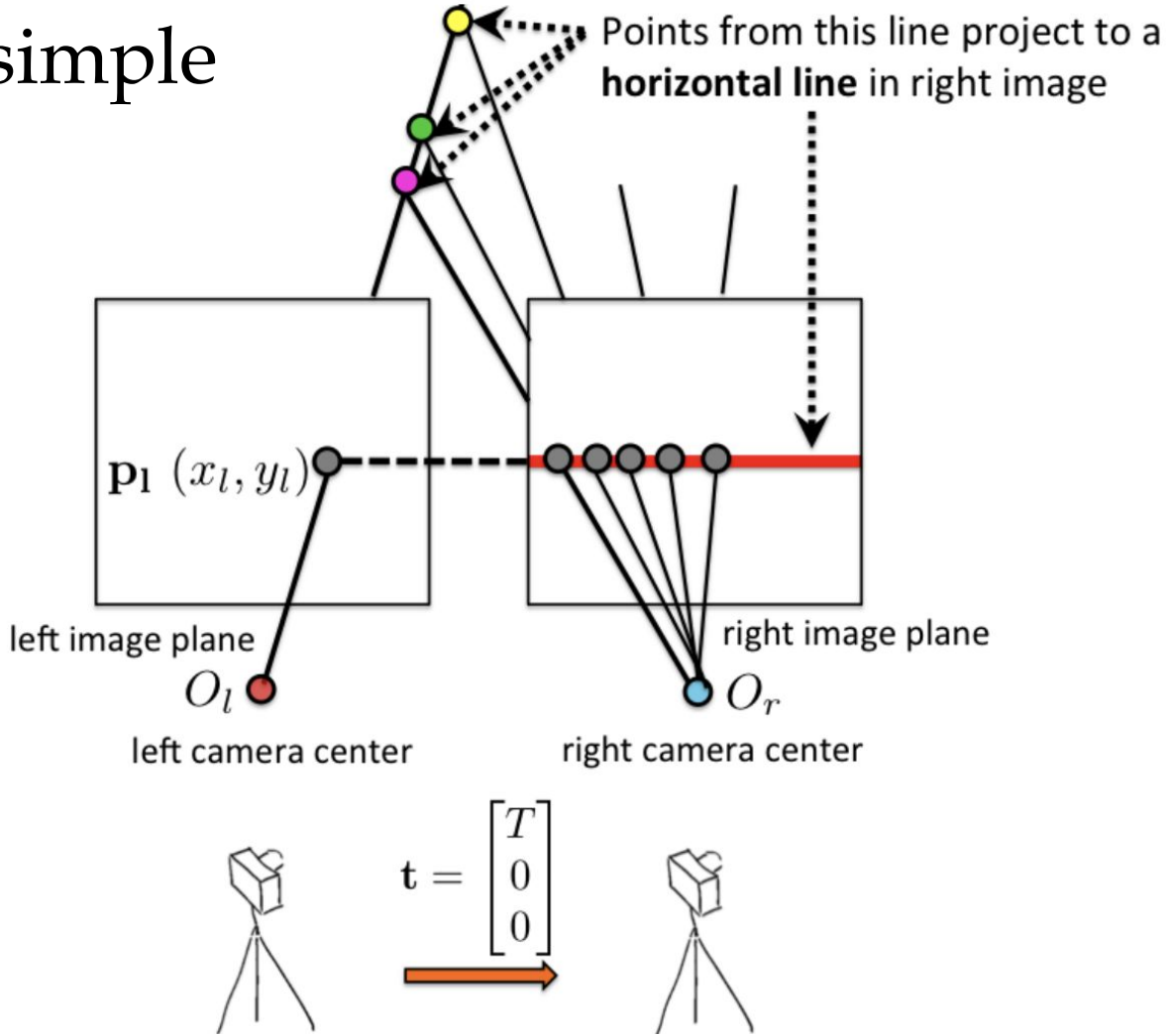


So: $y_r = y_l$

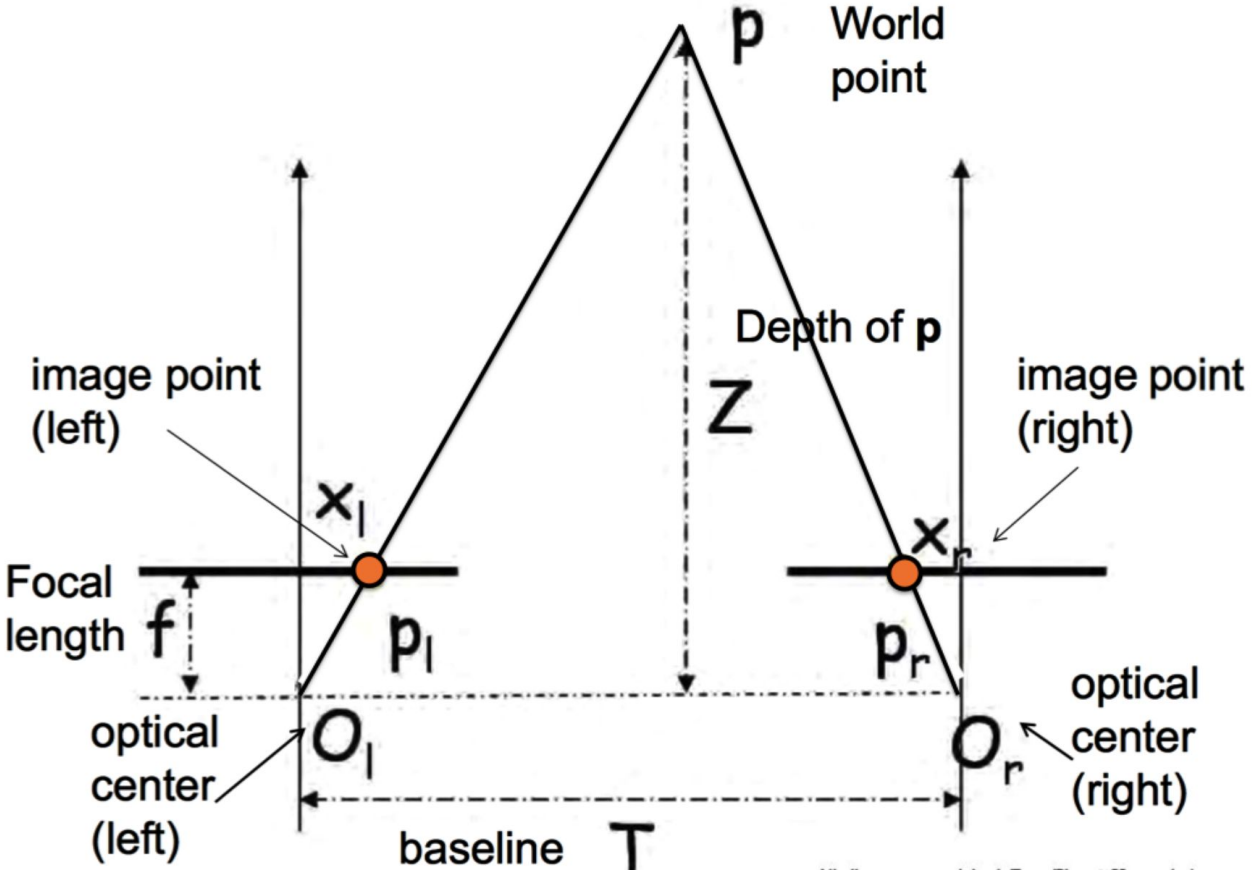
Nota:

- Olvidémonos de él
- Para poder olvidarnos de él, las cámaras deben estar **rectificadas**

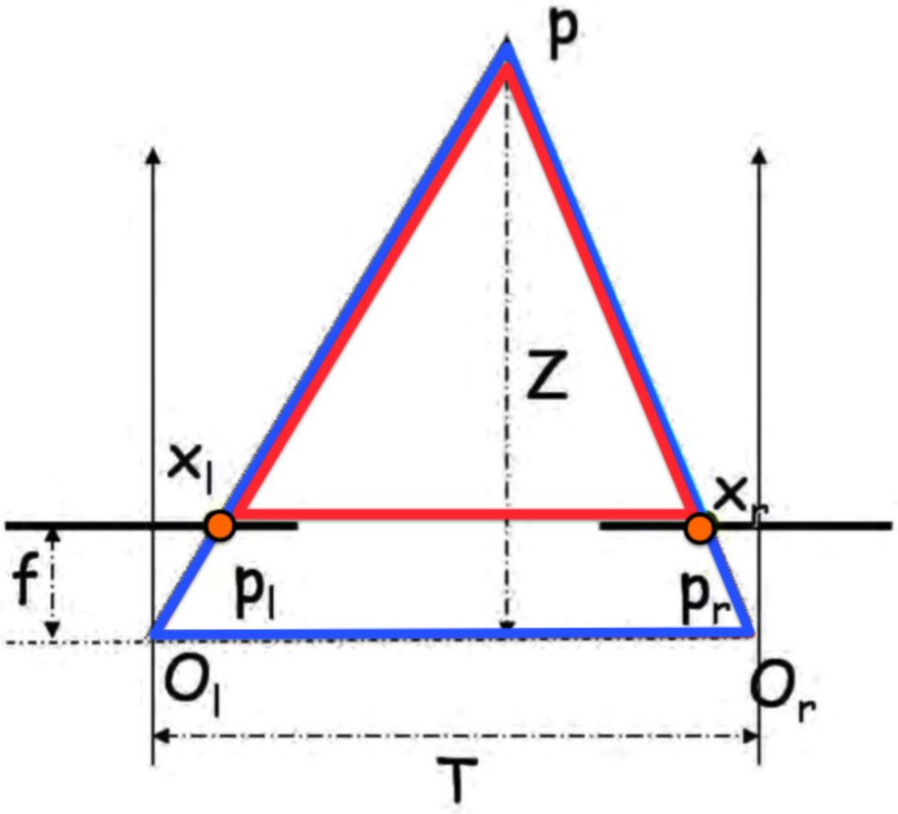
Caso simple


Points from this line project to a
horizontal line in right image
$\mathbf{p_l}$ (x_l, y_l)
left image plane
right image plane
O_l
O_r
left camera center
right camera center
$\mathbf{t} = \begin{bmatrix} T \\ 0 \\ 0 \end{bmatrix}$

Calculando Profundidad



Calculando Profundidad



Similar triangles:

$$\frac{T}{Z} = \frac{T + x_r - x_l}{Z - f}$$

$$Z = \frac{f \cdot T}{x_l - x_r}$$

baseline

focal length

disparity

Buscando (x_l, x_r)

Buscando (x_l, x_r)

left image

x_l

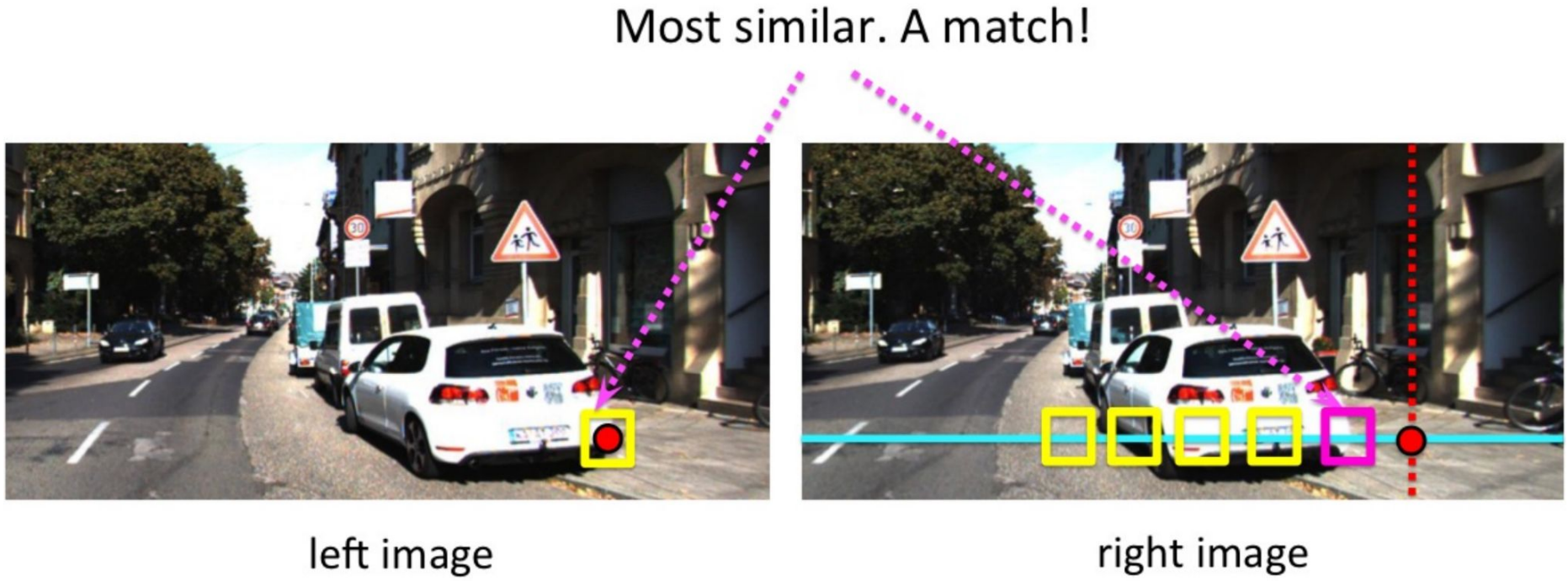
(x_r)

We are looking for this point

right image

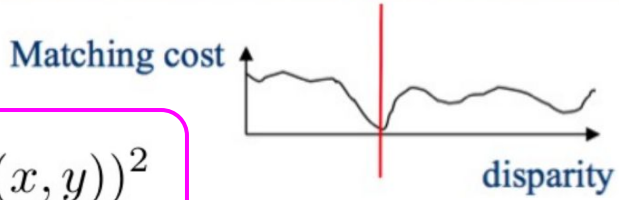
x_l

Buscando (x_l, x_r)



Buscando (x_l, x_r)

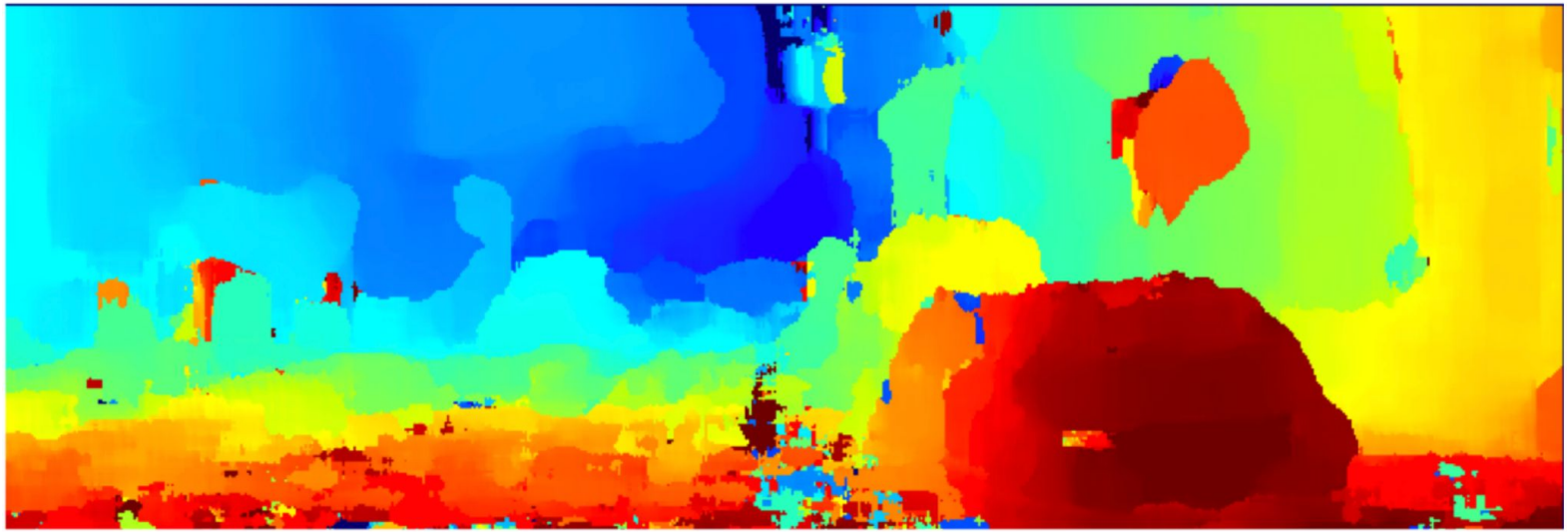
left image



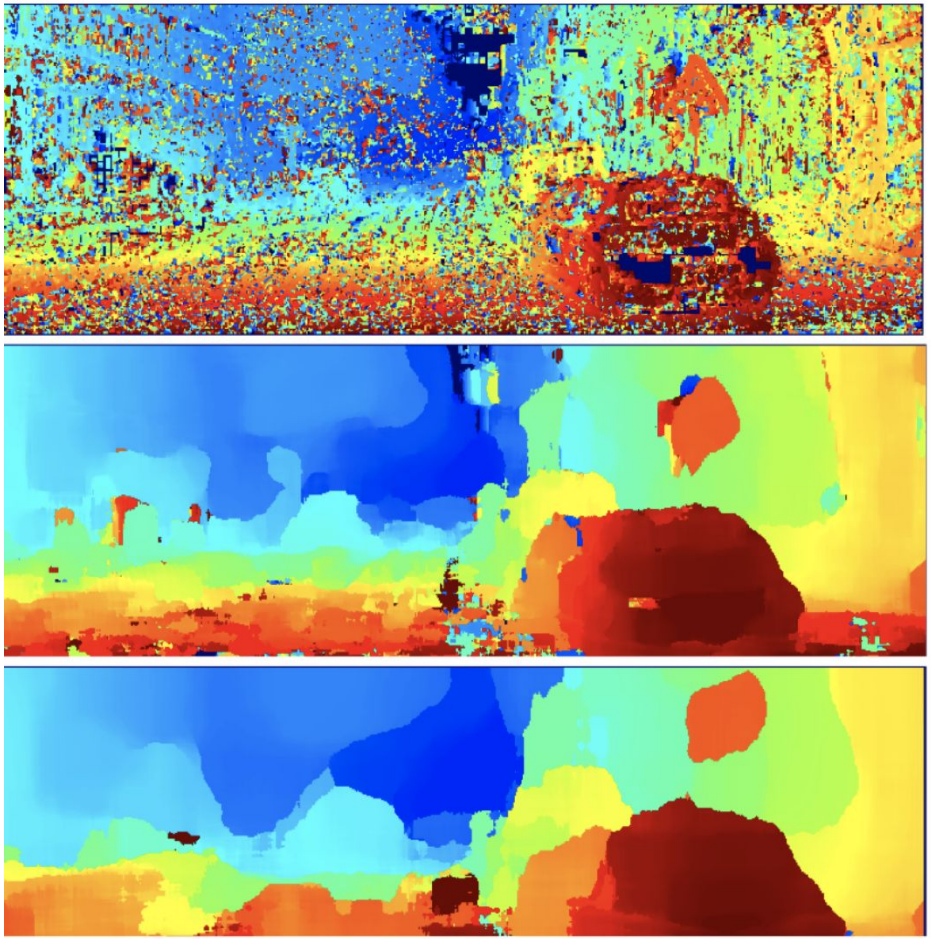
$$SSD(\text{patch}_l, \text{patch}_r) = \sum_x \sum_y (I_{\text{patch}_l}(x,y) - I_{\text{patch}_r}(x,y))^2$$

Buscando (x_l, x_r): Para todos los **“parches”**

Buscando (x_l, x_r): Para todos los **"parches"**



Disparidad para diferentes tamaños de **"parche"**



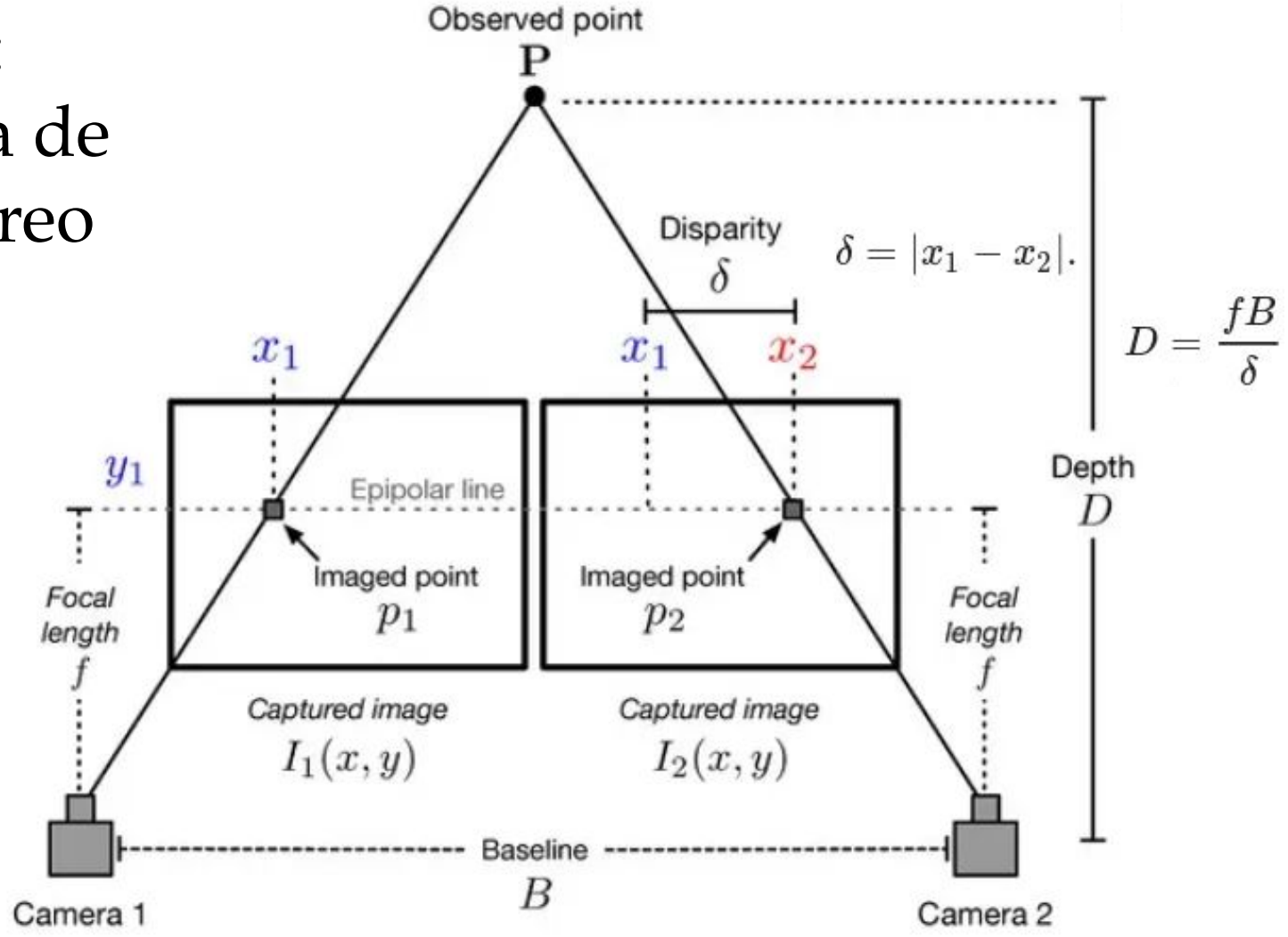
patch size = 5

patch size = 35

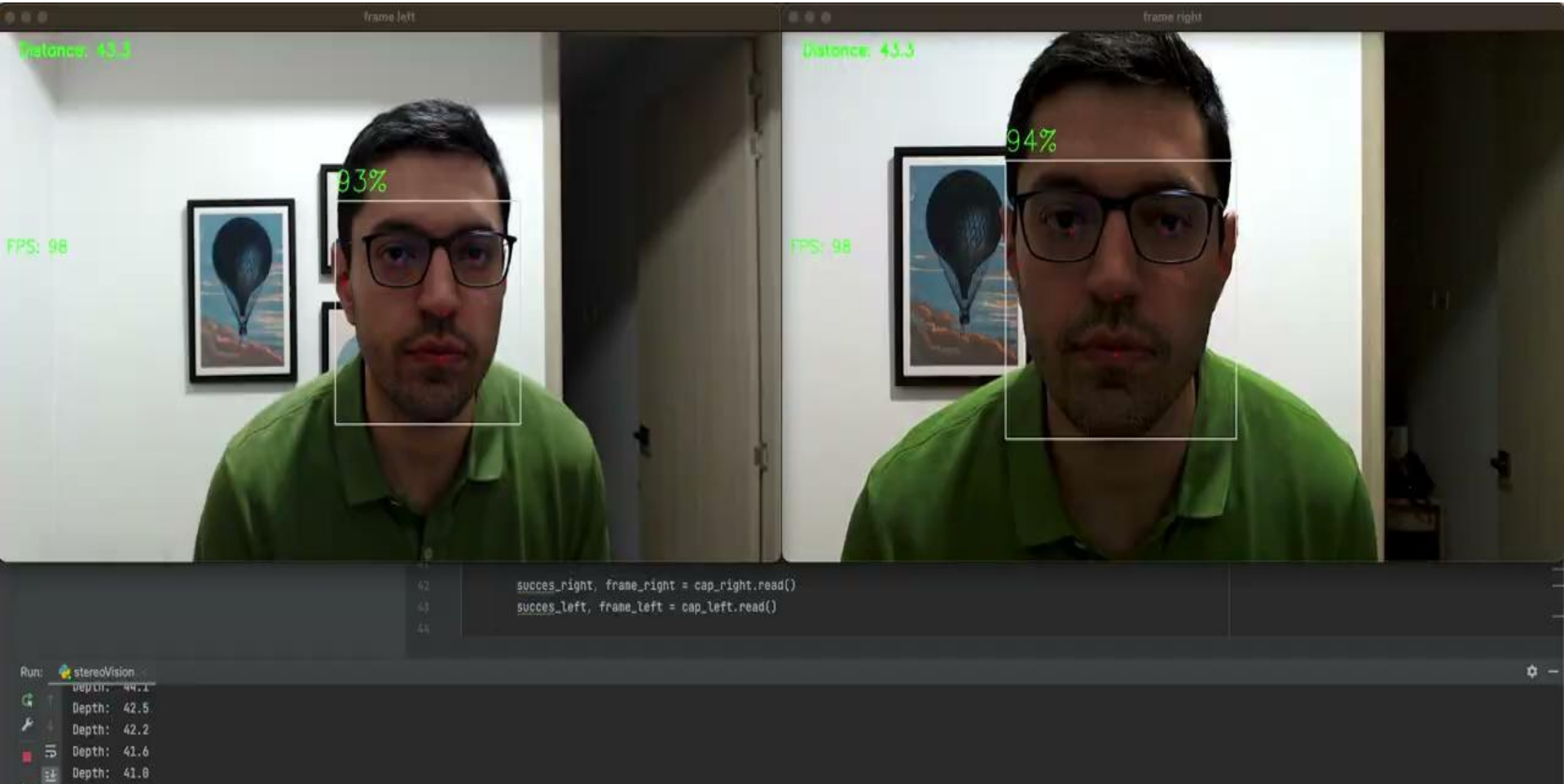
patch size = 85

Más grande = menos detalles (ruido)

Resumen:
Geometría de Visión Stereo



Demo: Stereo Vision + Tracking

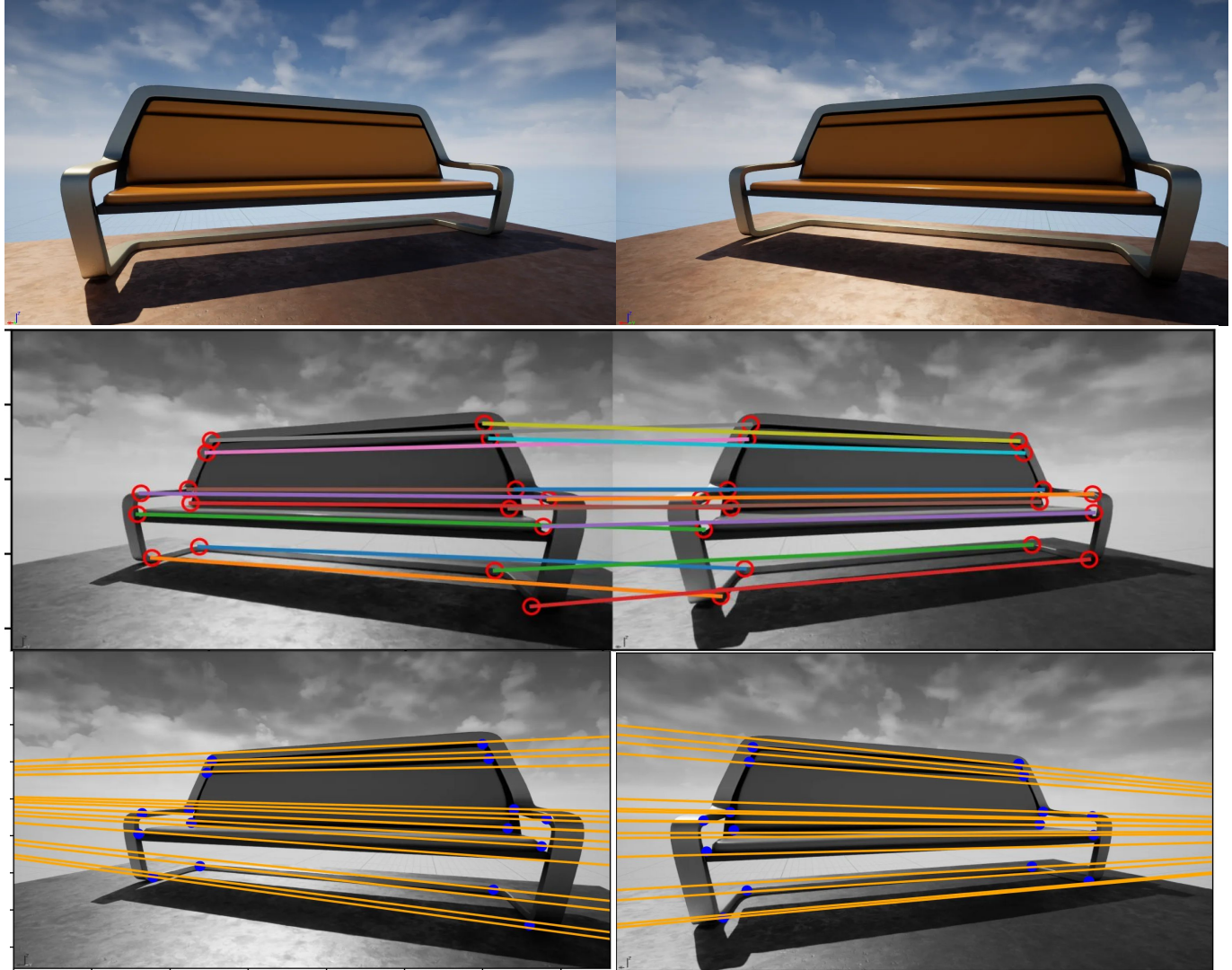


Caso NO ideal
(Haciendo rectificación)

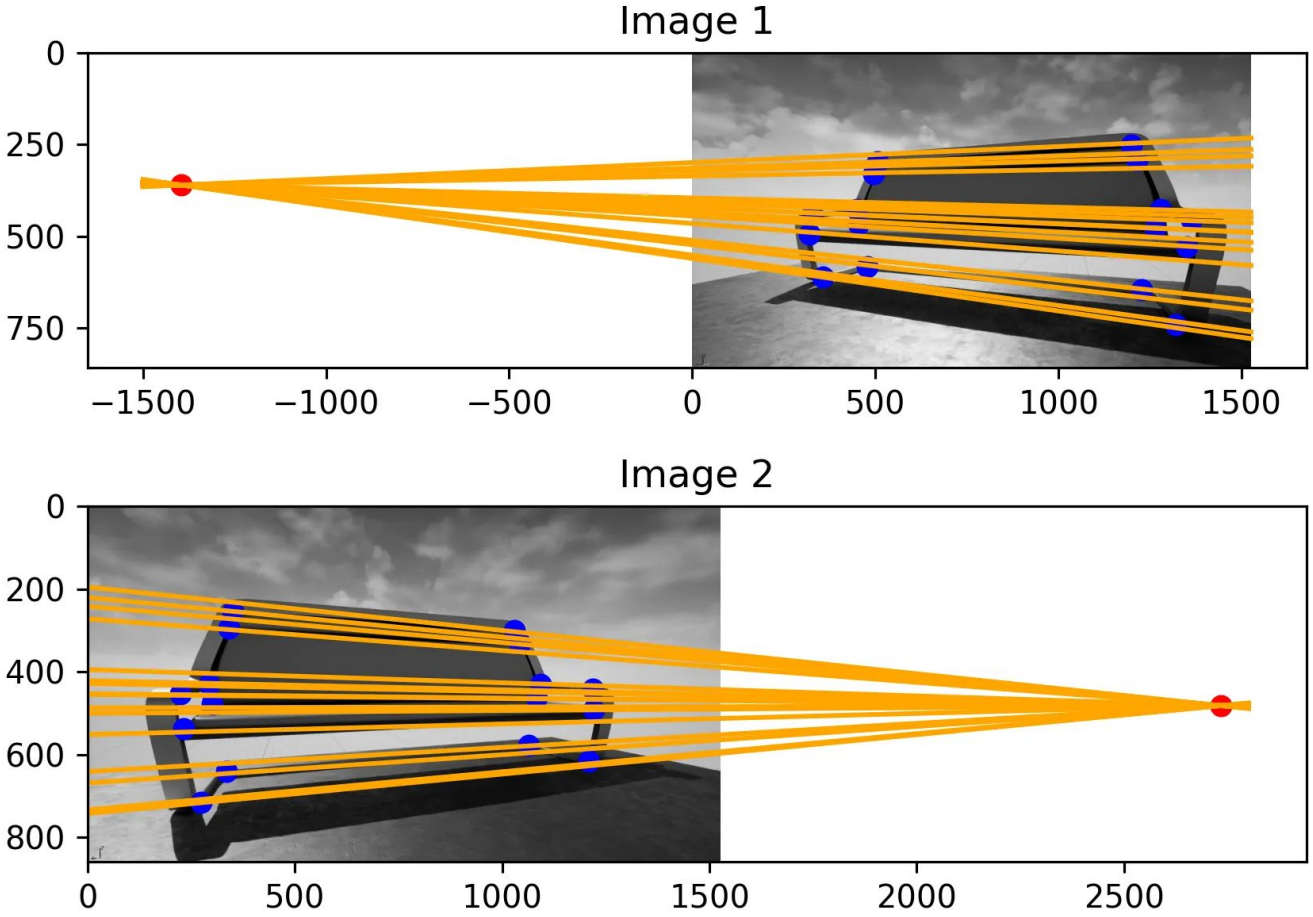
Adquirir imágenes

Buscar “features” (SIFT, ORB, etc.)

Calcular “vanishing point”

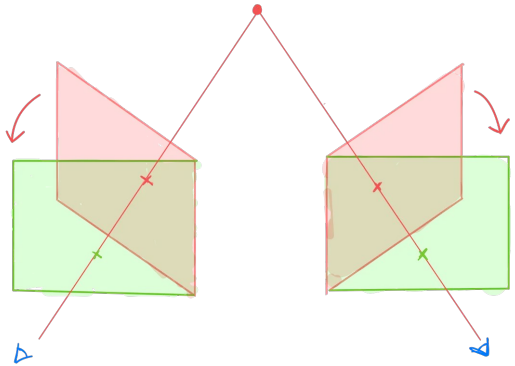


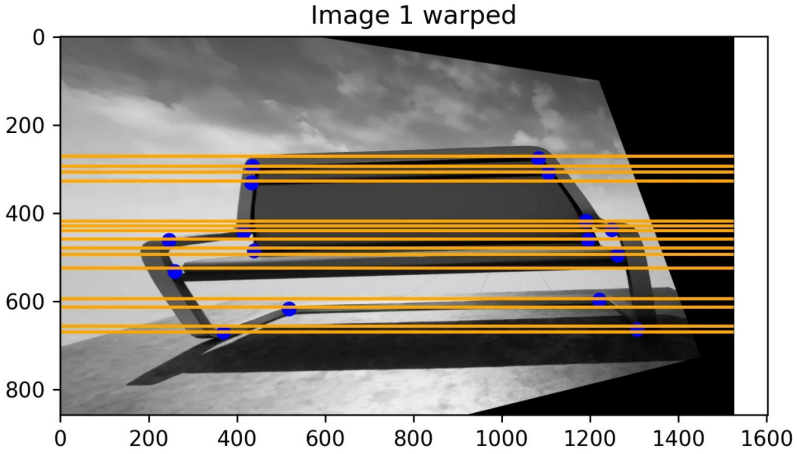
Calculando Vanishing Points

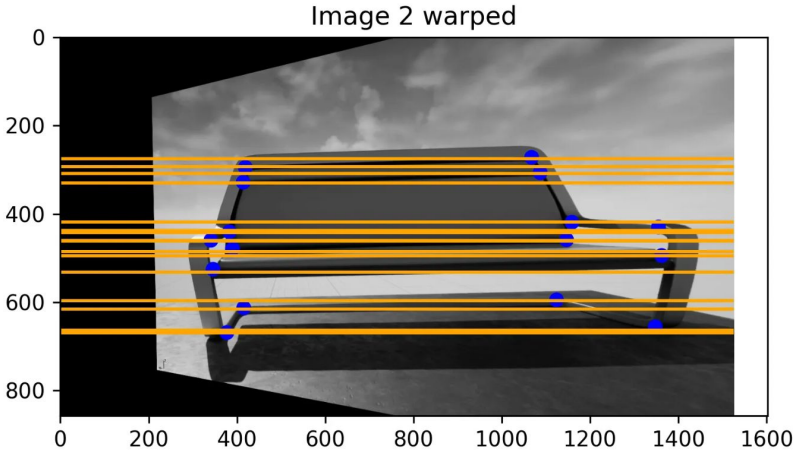


Rectificando

- Se busca la matriz de proyección **(Matriz Esencial y Fundamental)** usando la restricción epipolar
- Se hace una transformación (Warping) de las imágenes para obtener líneas paralelas







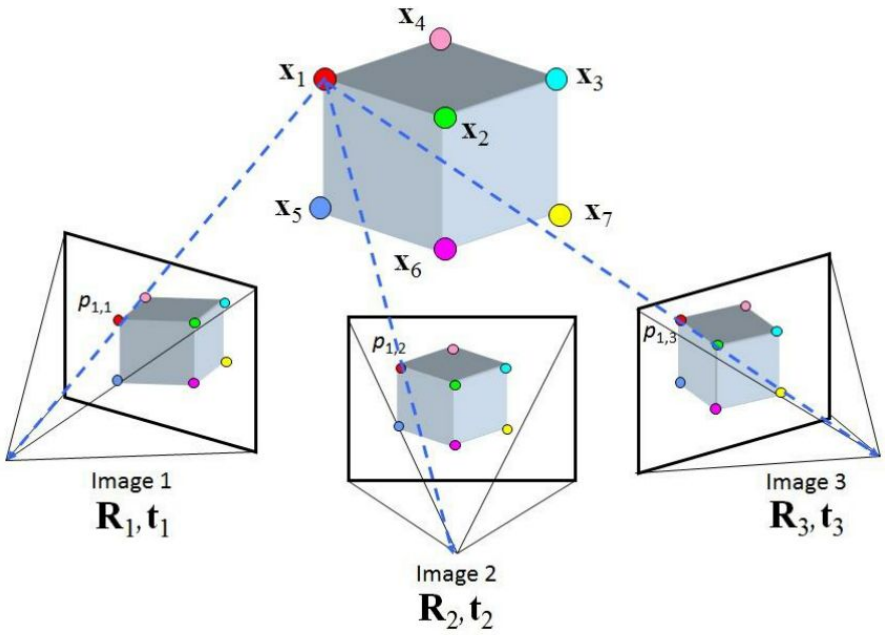
Nota: Con las imágenes rectificadas podemos calcular disparidad

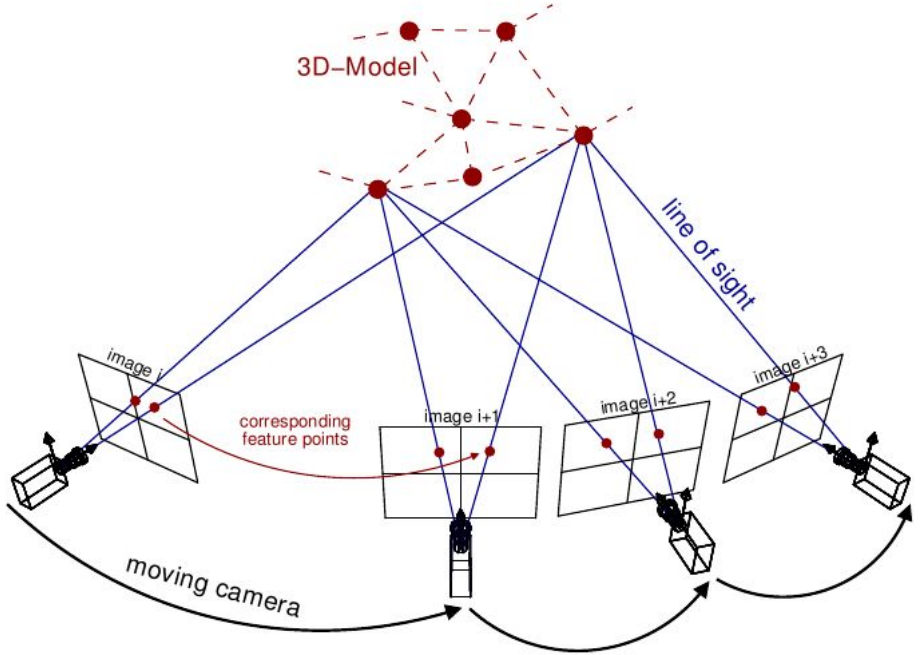
Rectificando (Ejemplo 2)

Nota: Con las imágenes rectificadas podemos calcular disparidad

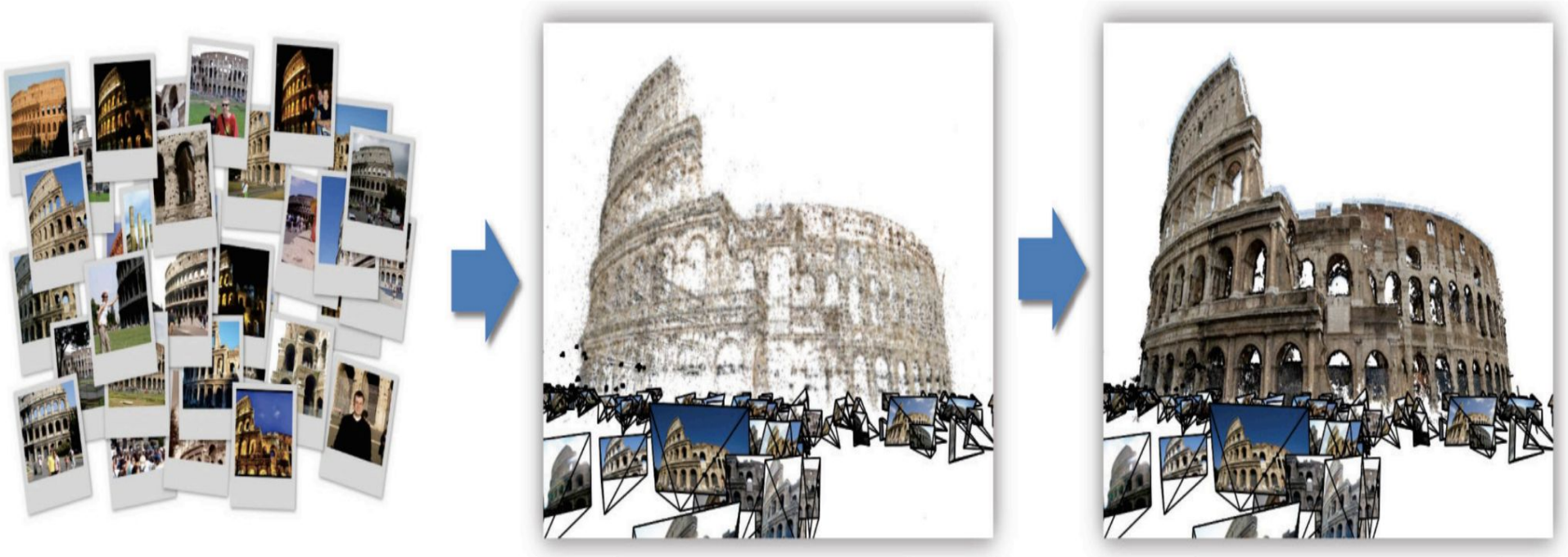
Caso más complejo (Structure from Motion)

Estructura a partir del movimiento (SfM)



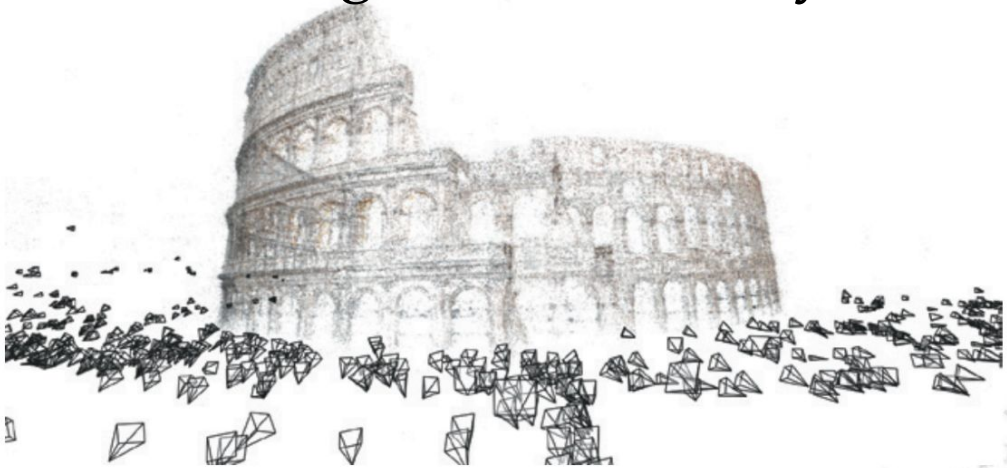


SfM: Building Rome in a Day

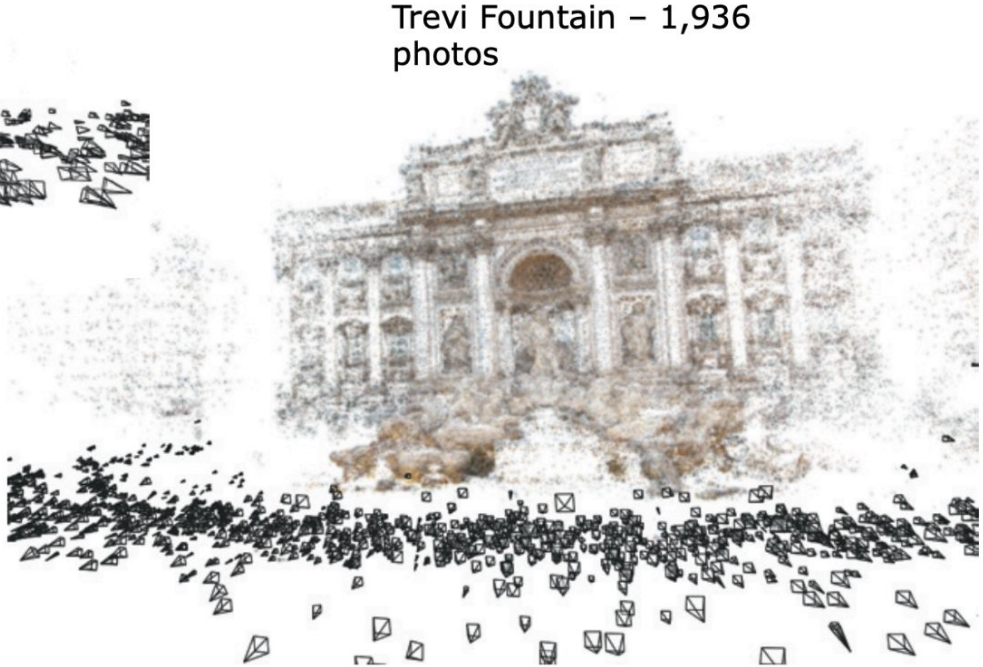


https://grail.cs.washington.edu/rome/

SfM: Building Rome in a Day



Colosseum – 2,106 photos



Trevi Fountain – 1,936 photos

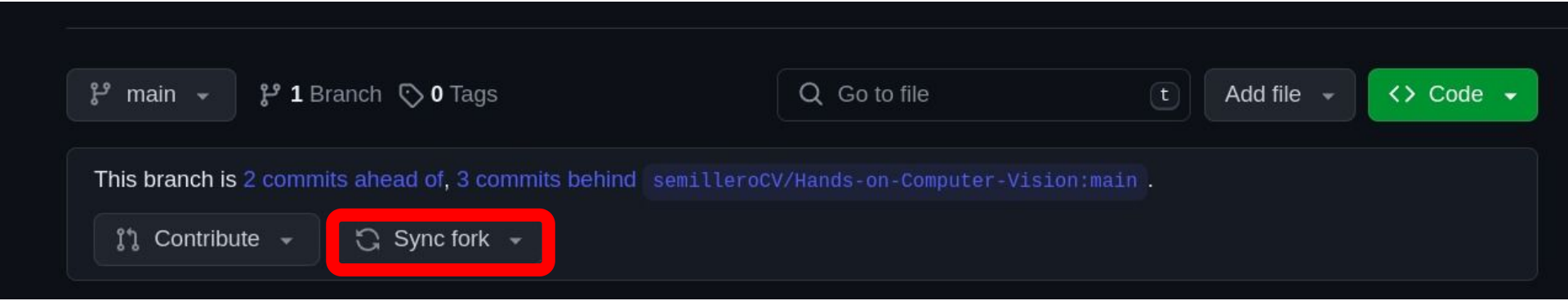
La reconstrucción de Roma (150.000 fotografías) tomó 26 horas (18 horas para hacer *matching* y 8 horas para la reconstrucción) utilizando 496 procesadores.

Visión Estereo *in the wild*

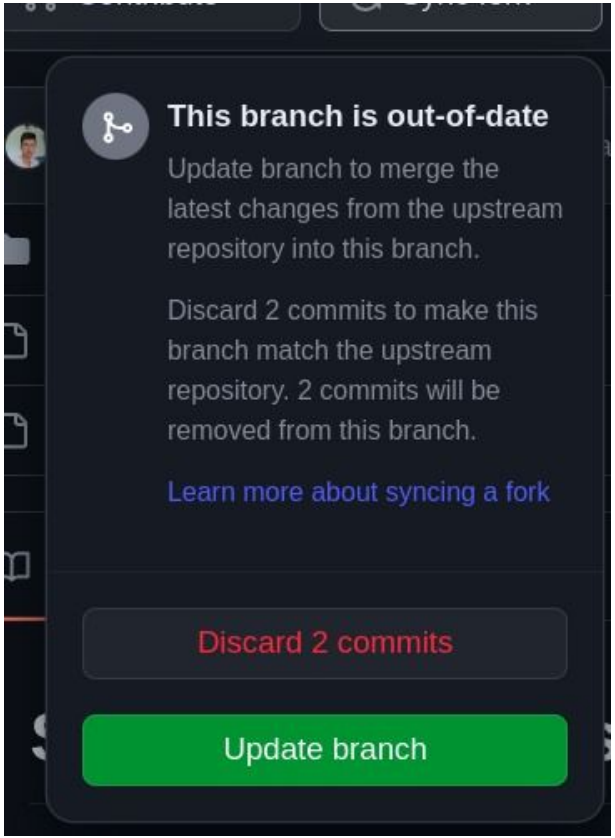
SfM

4. Hands-on: Stereo Vision

Actualicemos el repositorio!



Actualicemos el repositorio!



Actualicemos el repositorio!