
Visión 3D en robots móviles

José María Cañas Plaza

<http://gsyc.escet.urjc.es/jmplaza>



III Jornadas de robótica URJC, 17 octubre 2006

Contenidos

- Introducción
- Localización de objetos
- Reconstrucción 3D
- Fusión sensorial
- Atención visual
- Conclusiones

Introducción

¿Qué es un robot?



Sistema informático con:

- Sensores
- Actuadores
- Computadores

Hay que **programarlo** para que consiga sus objetivos y sea sensible a la situación

¿Qué es la visión computacional?

- Cámara, imágenes
- Procesamiento 2D: bordes, transformadas, color, segmentación en objetos...
- Reconocimiento de formas, caras...
- Pares estéreo, línea epipolar...



Visión dentro de robótica

- Es un sensor más: proporciona información del entorno
- Potencialmente muy *rico*
- MUY *barato* (cmos, webcam)
- Extraer información útil es *complejo*
- Flujo desbordante de datos

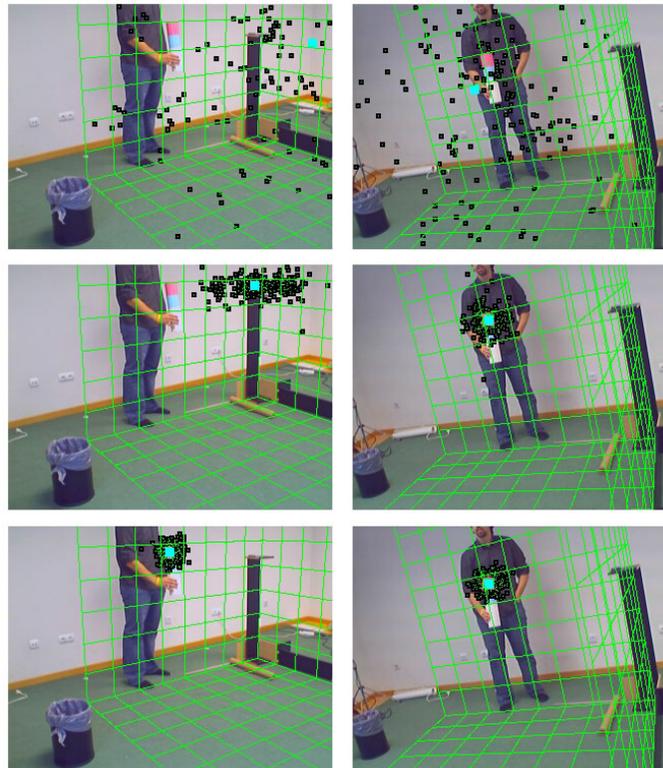


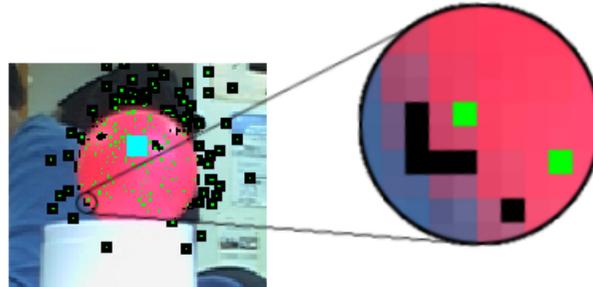
Escena 3D rica

- Una **representación 3D rica de los alrededores del robot** facilita la navegación
 - la construiremos fundamentalmente desde información visual
 - dinámica y local
 - El procesamiento visual ha de encajar en el planteamiento de percepción que ofrece la arquitectura cognitiva
- **Localizar** 3D objetos de interés
- No hace falta **reconstruir 3D** para navegar, pero ayuda
- **Fusión sensorial** para estímulos cotidianos
- **Atención visual** para alimentar la escena

Localización visual 3D de objetos

Con filtro de partículas





- Población de partículas se mueve por el espacio de estados
- Cada partícula tiene un peso probabilístico
- **MonteCarlo**: *Modelo de observación, de movimiento y remuestreo*
- Iterativo, eficiente (40-150 ips)
- Modelos de observación simples: de color y movimiento
- Multiobjeto se complica

Watcher JDE

Virtual Camera X:14777.3 Y:-2093.2 Z:4610.7 x:0.0 y:4425.0 z:-25.0 Camera Select

real cams simulated cams
 virtual camera on

Color Filter and Image Selectors

colorA 9.0 fps
 colorB 9.0 fps
 colorC 9.0 fps
 colorD 8.0 fps
 hsi filter
 motion filter

Camera A Camera B Virtual Camera Camera C Camera D Cameras Control

A B C D All VC

Room Grid OptAxis Axis

black ambient flip image
Hue Tolerance 10.00 Saturation Tolerance 6.10
Item Tolerance 30.00 Motion Tolerance 40.00
Hmax Hmin Smax Smin

Files Filter Control Particles Filter Control Simulated Object Control Main Options

Files Filter Control: color motion
Eitism 30 Abductive 60 Thermalnoise 5 Random Mutation 5
NO OBJECT DETECTED NEAR TARGET
files fps 0.0 fps Estimated: X:0.0 Y:0.0 Z:0.0 Diff: X:0.0 Y:0.0 Z:0.0

Particles Filter Control: color motion
100 Std 0 Mean
WARNING! OBJECT DETECTED NEAR TARGET
particles fps 27.0 fps Estimated: X:3707.5 Y:2363.6 Z:961.7 Diff: X:0.0 Y:0.0 Z:0.0

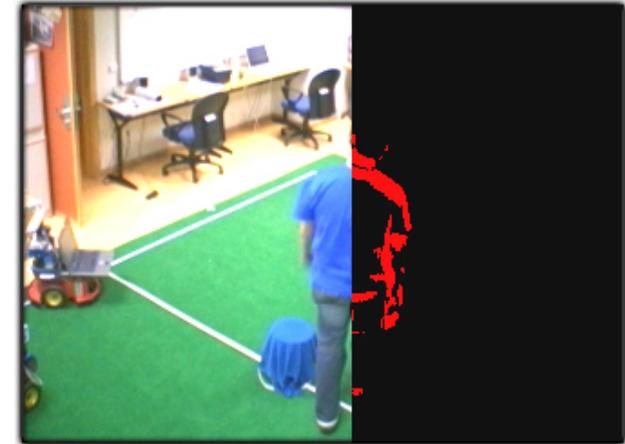
Simulated Object Control: 3d cube blue red pink
 target pink
Animation Center Lock Sp 0.05
3d cube - X:3701.2 Y:2363.6 Z:961.7
target - X:4788.0 Y:2674.0 Z:1000.0

Main Options: Simulated Real
watcher fps 26.0 fps



Probabilidad asociada Color

$$P(\text{color}_i | \text{img}_m) = \frac{\max(1, k)}{25}$$



Probabilidad asociada Movimiento

$$P(\text{mov}_i | \text{img}_m) = \frac{\max(1, k)}{25}$$

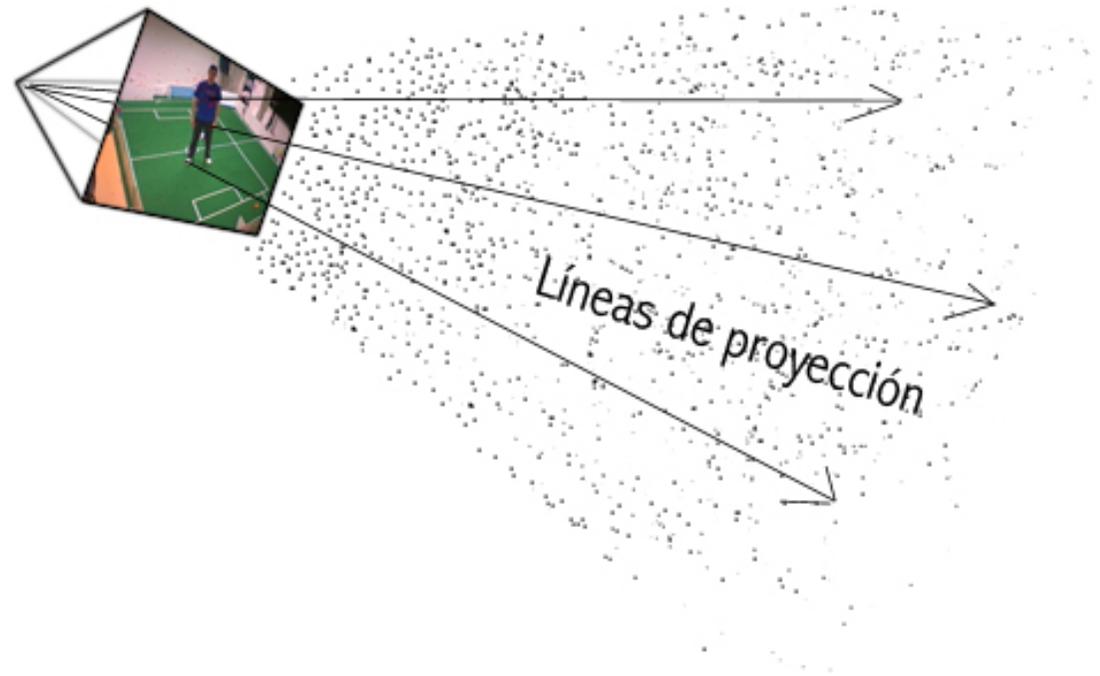
Probabilidad total

$$p_i = \prod_{m=1}^M P(\text{color}_i | \text{img}_m) * P(\text{mov}_i | \text{img}_m)$$

Con algoritmo genético de moscas

- Población de moscas se mueve por el espacio de estados
- Cada mosca puntual tiene una salud
- Salud desde visión, operadores genéticos
- Iterativo, eficiente
- Explora el color y el movimiento
- Multiobjeto se complica

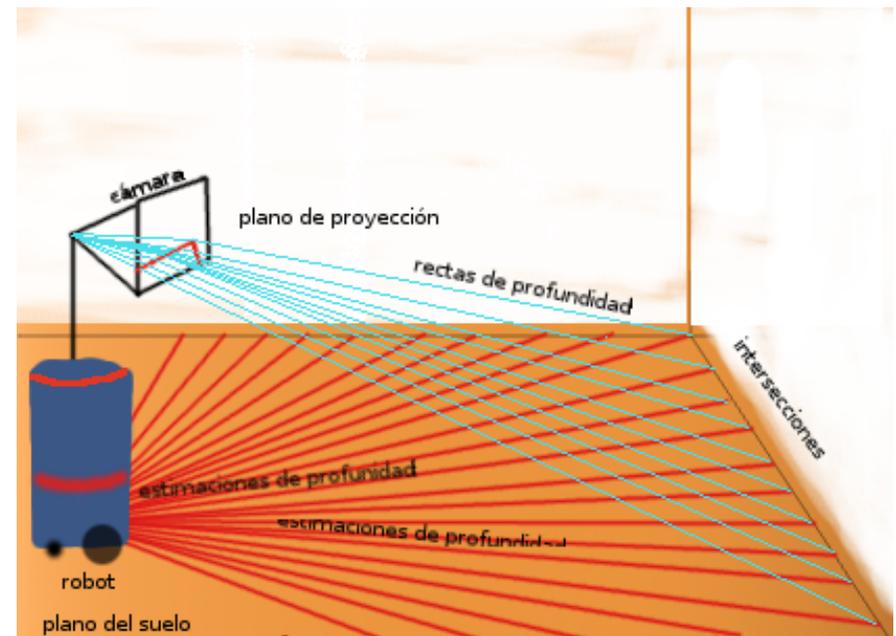
- Cada operador un porcentaje
- Generación aleatoria
- Ruido térmico
- Elitismo
- **Abducción**: Genera nuevas moscas sobre las líneas de proyección

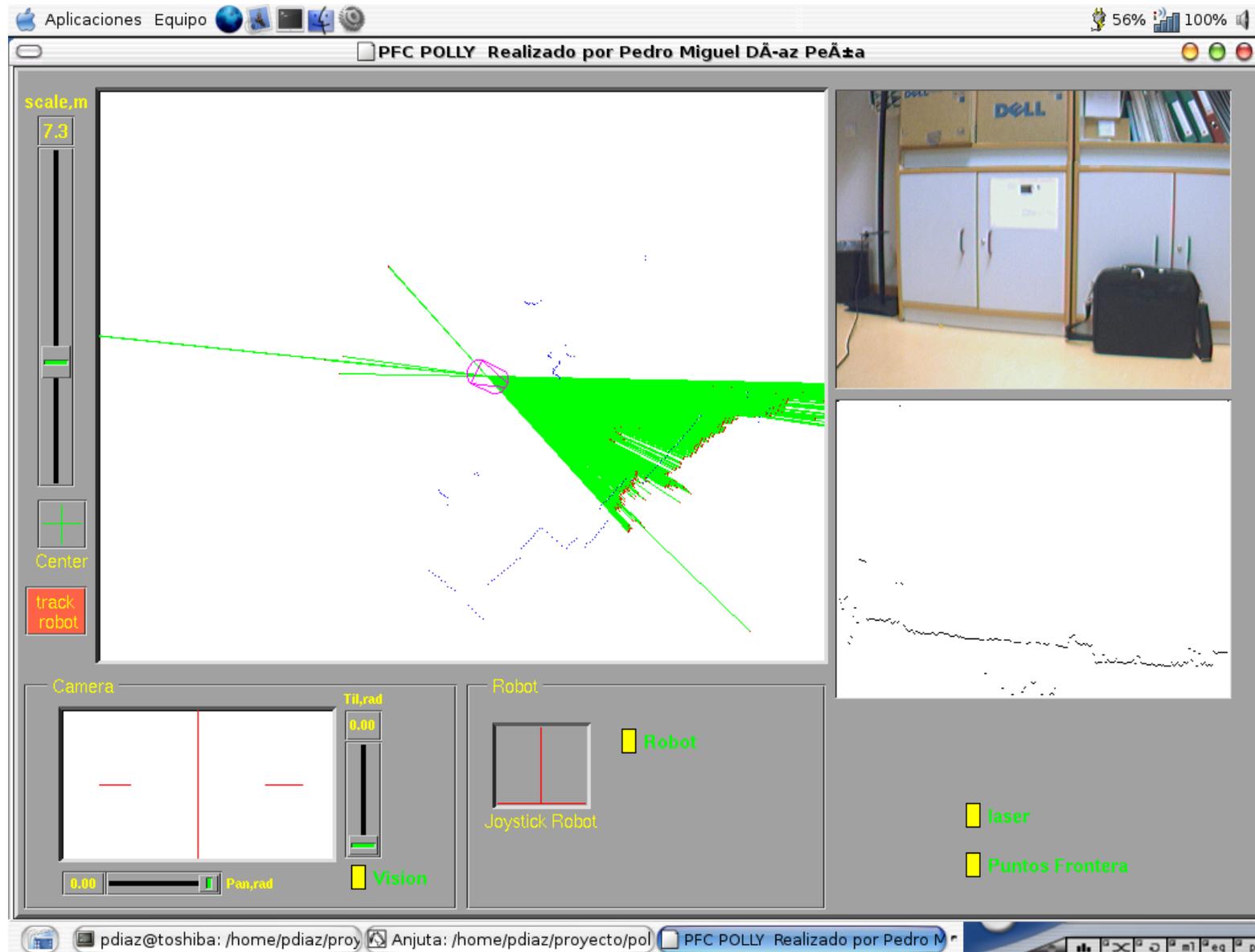


Reconstrucción 3D

Con visión monocular

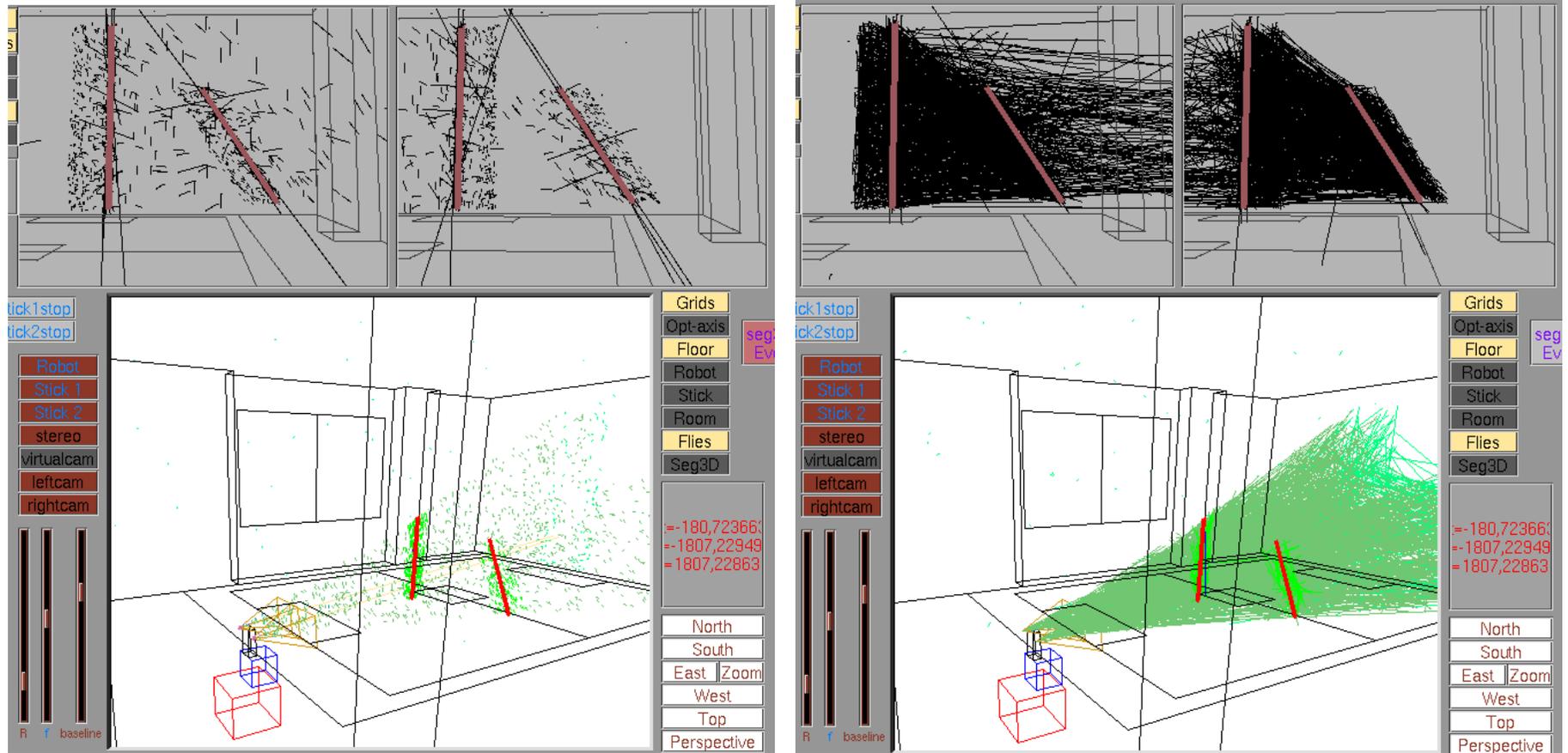
- Hipótesis suelo: bordes o color
- Con una sola cámara
- Robot Polly
- Escena moviendo la cámara con barrido horizontal sistemático





Reconstrucción 3D con algoritmo evolutivo, par estereo

- Situar en 3D los bordes de los objetos
- Primitivas: punto 3D, segmentos 3D fijos o variables
- **Salud**: proyecta en borde en ambas cámaras, parecido de proyecciones, completitud, coherencia
- Repulsión
- Reconstruir es más complicado que localizar
- Sin correspondencia ni triangulación explícitas. No construir, sino hipotetizar y comprobar



Fusión sensorial en la escena 3D

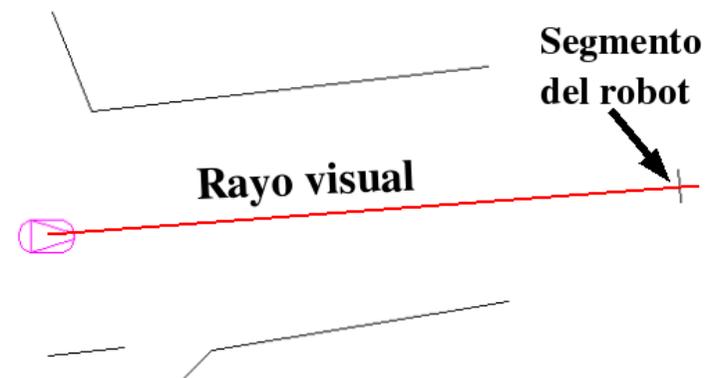
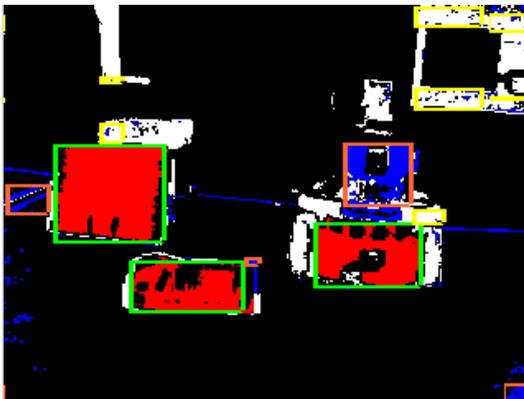
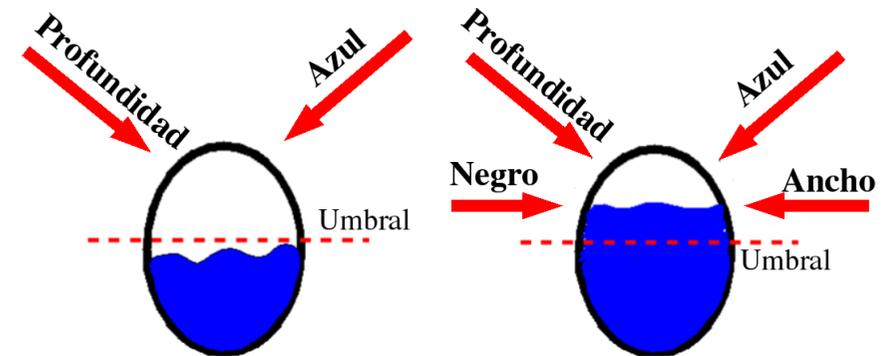
- Percibir objetos cotidianos del entorno: paredes, puertas, personas, otros robots
- Identificación, mantenimiento y olvido
- Memoria de corto plazo
- Cada objeto tiene una salud
- Correspondencia entre escena memorizada y datos instantáneos de sensores



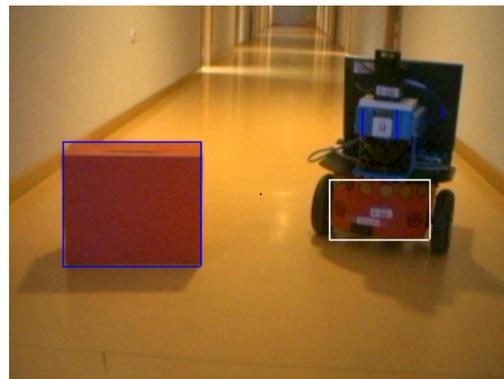
Percepción de otro robot

SUMA HETEROGÉNEA DE SUBESTÍMULOS

- Mancha visual roja
- Más ancho que alto
- Mancha negra o azul encima
- Sensación de profundidad

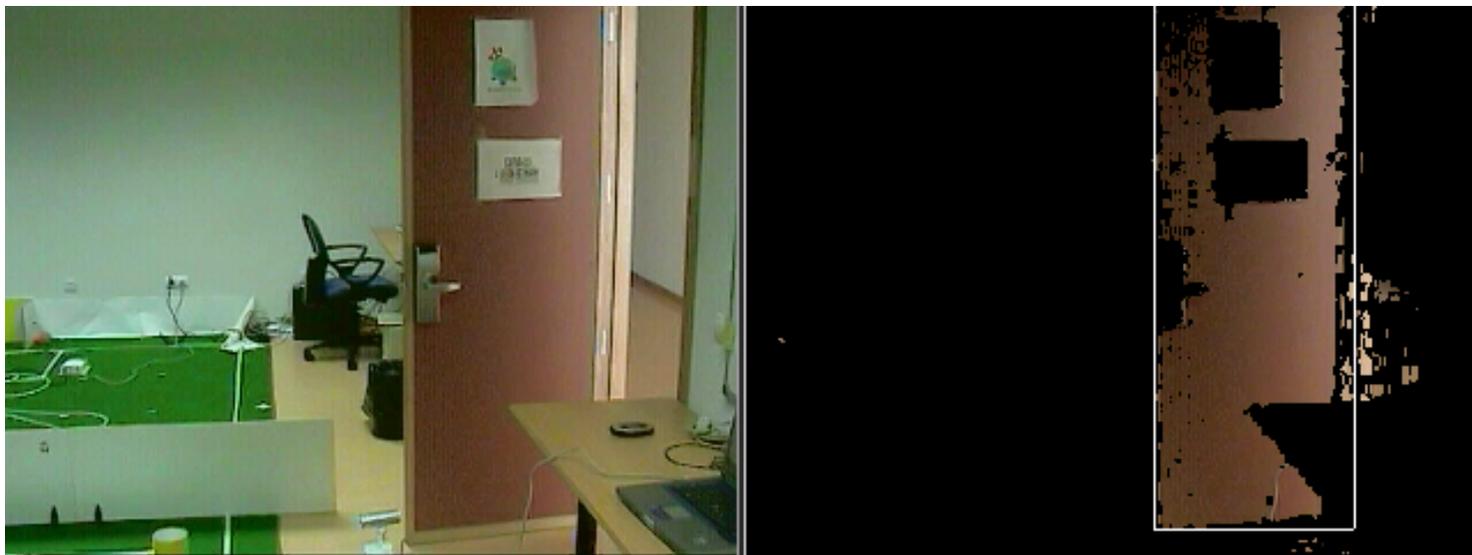


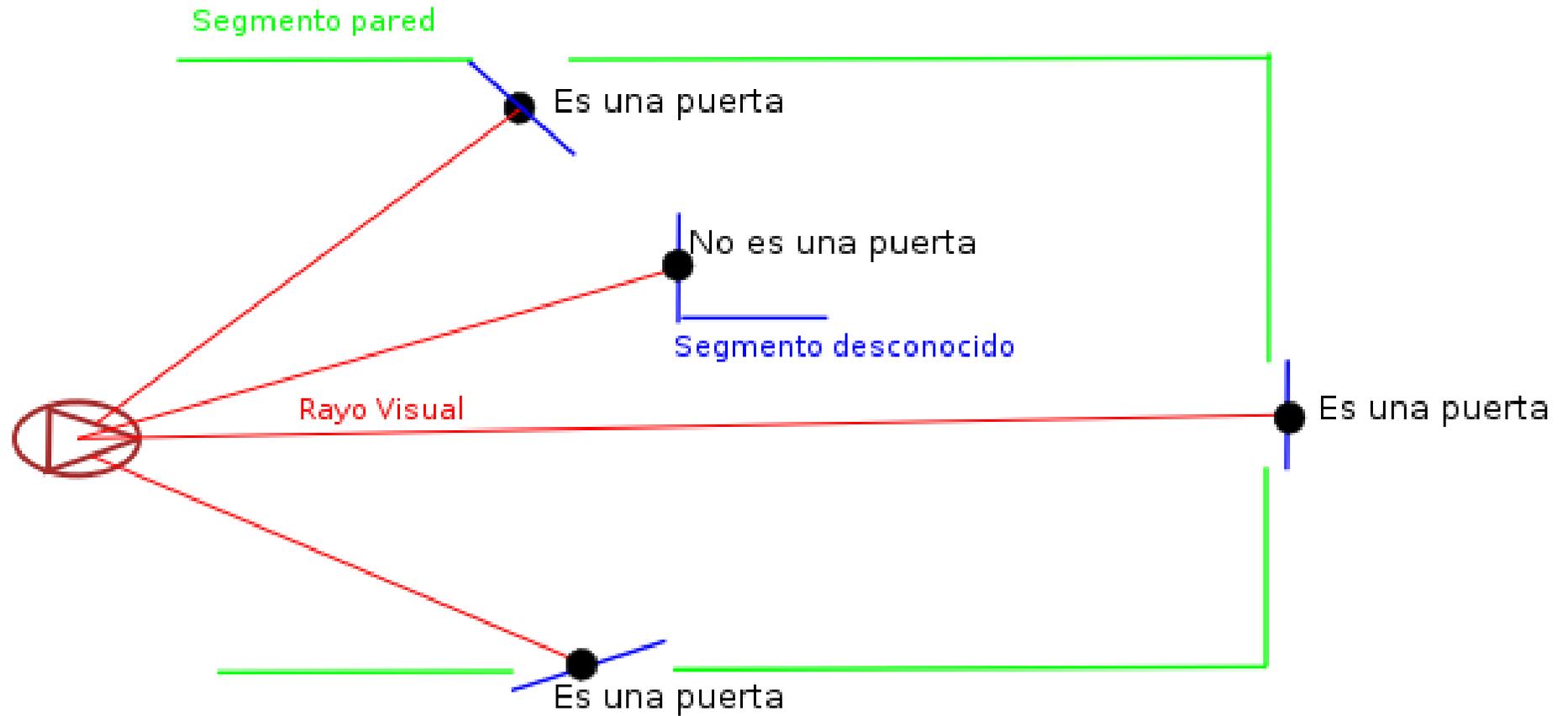
ES MUY ROBUSTO Y DISCRIMINANTE



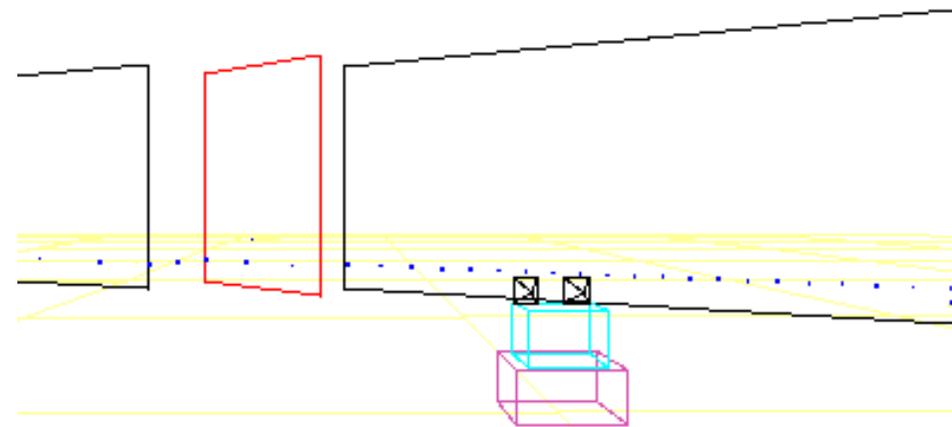
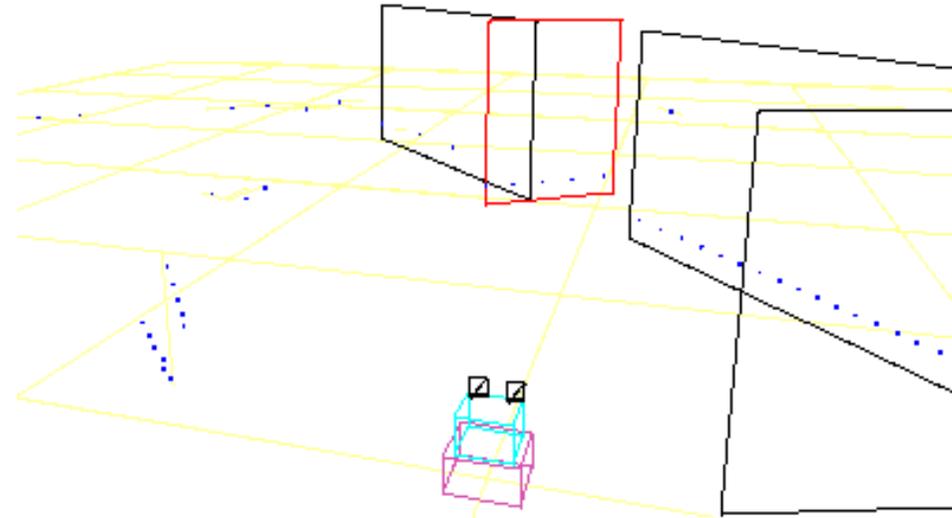
Percepción de puertas

- Filtro de color HSI
- Segmentación histográfica recursiva
- Segmento de color marrón rojizo más alto que ancho
- Segmento láser inferior a 1.5 metros y junto a una pared



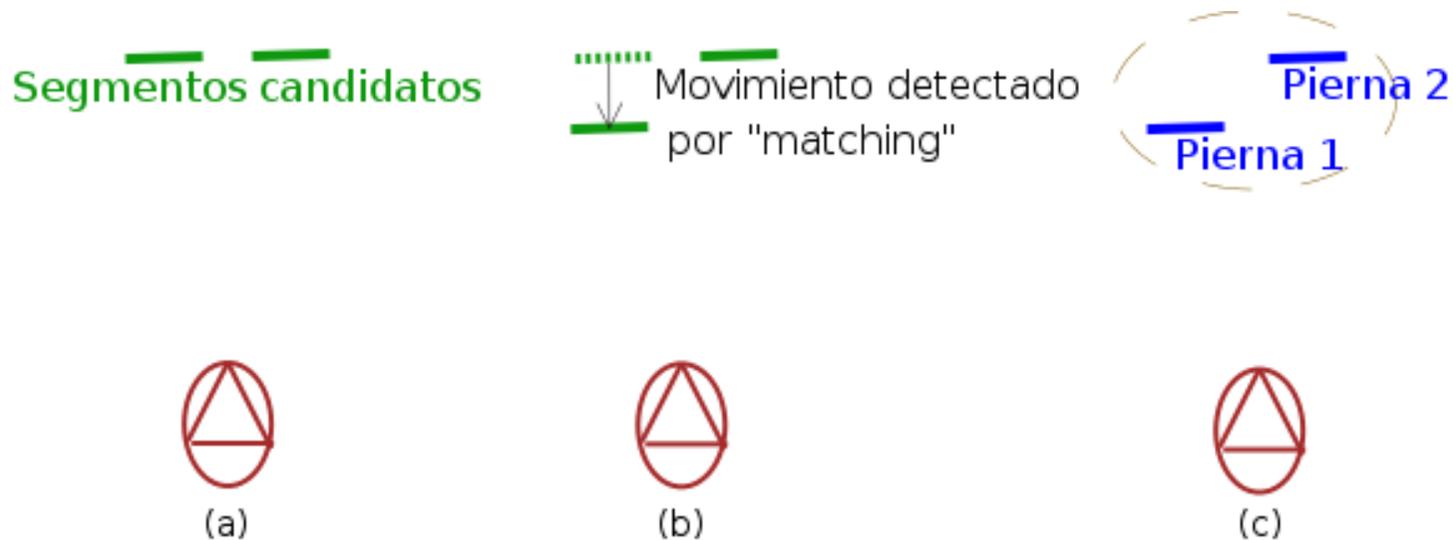


- Mantenimiento en memoria: salud y posición

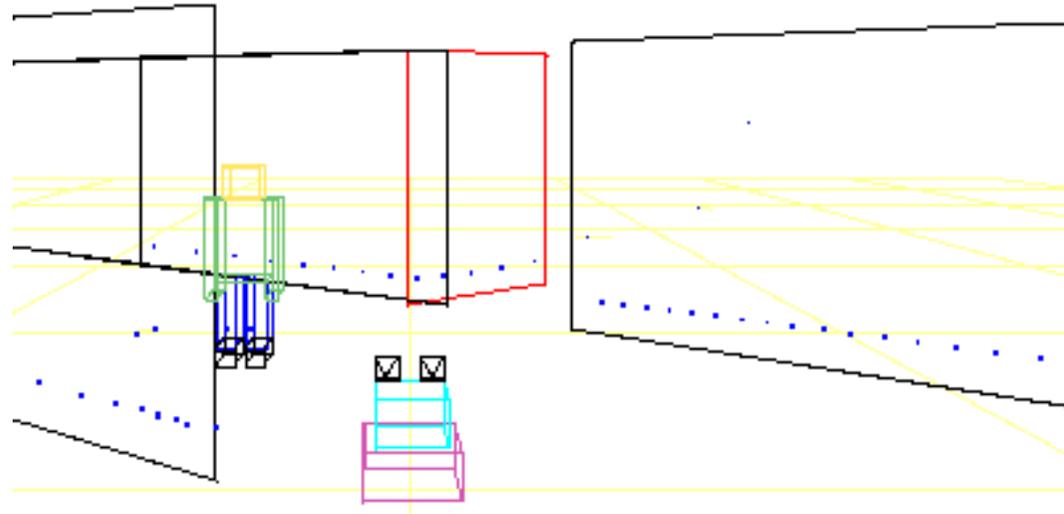


Percepción de personas

- Parejas de segmentos laser similares y pequeños (las dos piernas).
- Movimiento en algún segmento candidato, en la imagen o con laser



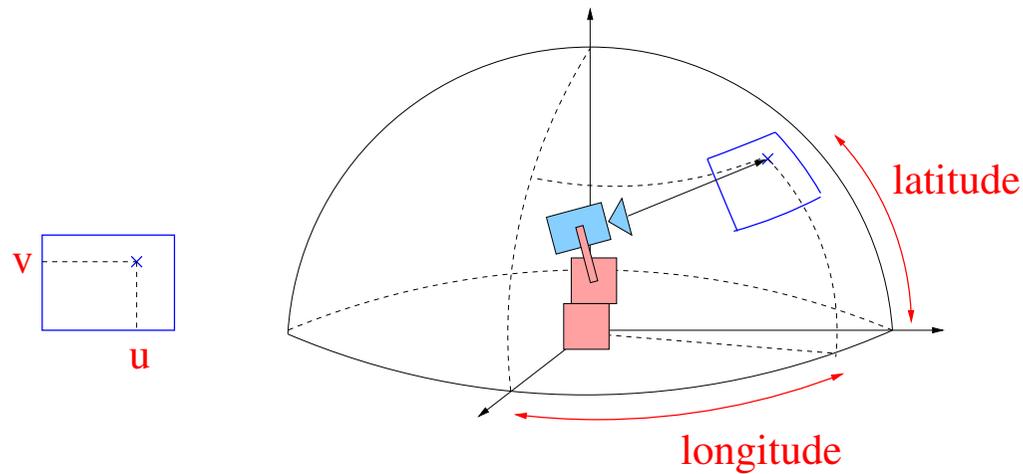
- Mantenimiento en memoria: salud y posición



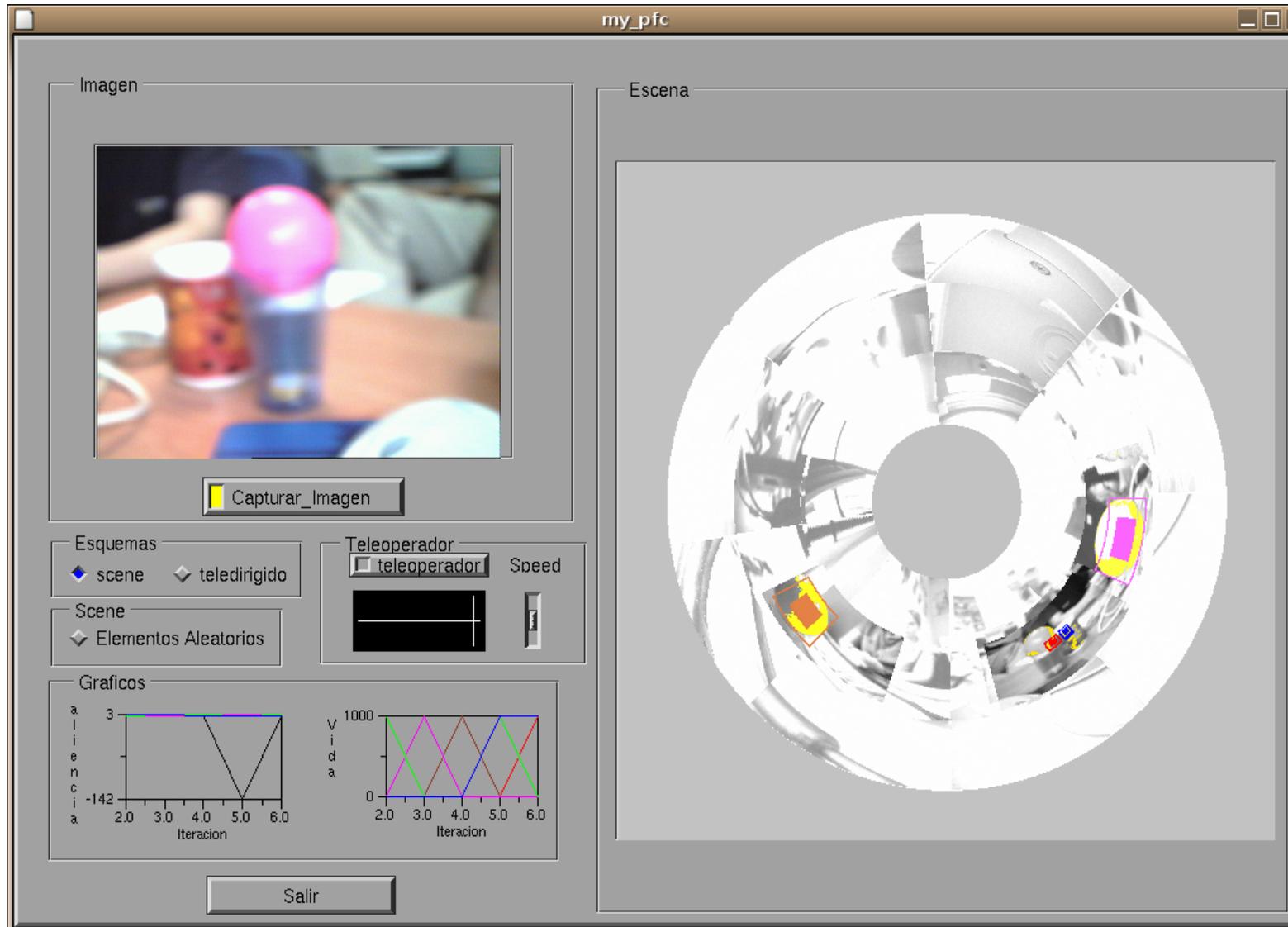
Atención visual

- La escena completa no cabe en un solo vistazo, hay que mover las cámaras
- Foco de atención: puntos de exploración (barrido y aleatorios)
- Foco de atención: los objetos relevantes de la escena ya identificados
- **Dinámicas temporales de saliencia y de vida**
 - $liv(object, t) = liv(object, t - 1) - \Delta L_{time}$
 - $liv(object, t) = liv(object, t - 1) + \Delta L_{observation}$
 - $sal(fixp, t) = sal(fixp, t - 1) + \Delta S_{time}$
 - $sal(fixp, t) = 0$

Con una sola cámara

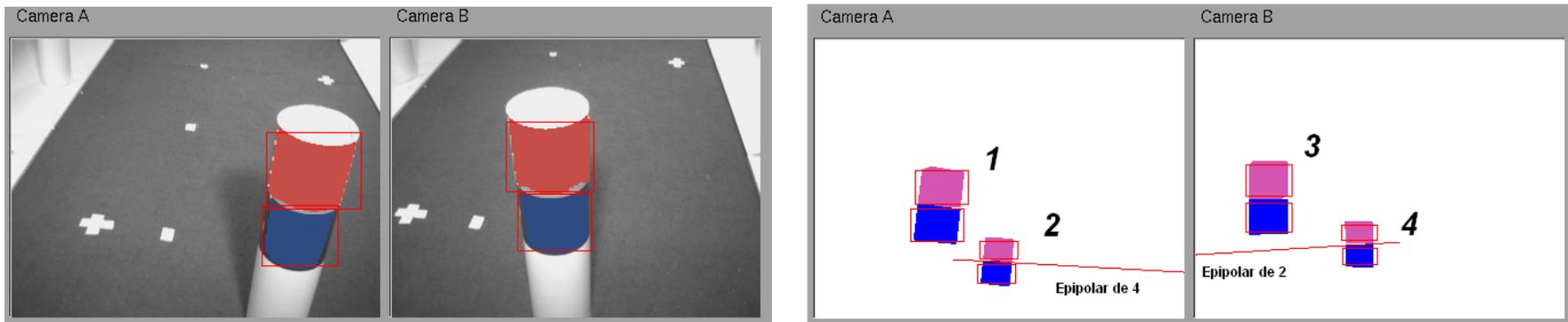


- Objetos de colores significativos
- ¿Cuántos objetos relevantes hay en la escena?
- Alternancia natural
- Prioridades
- Número máximo de objetos

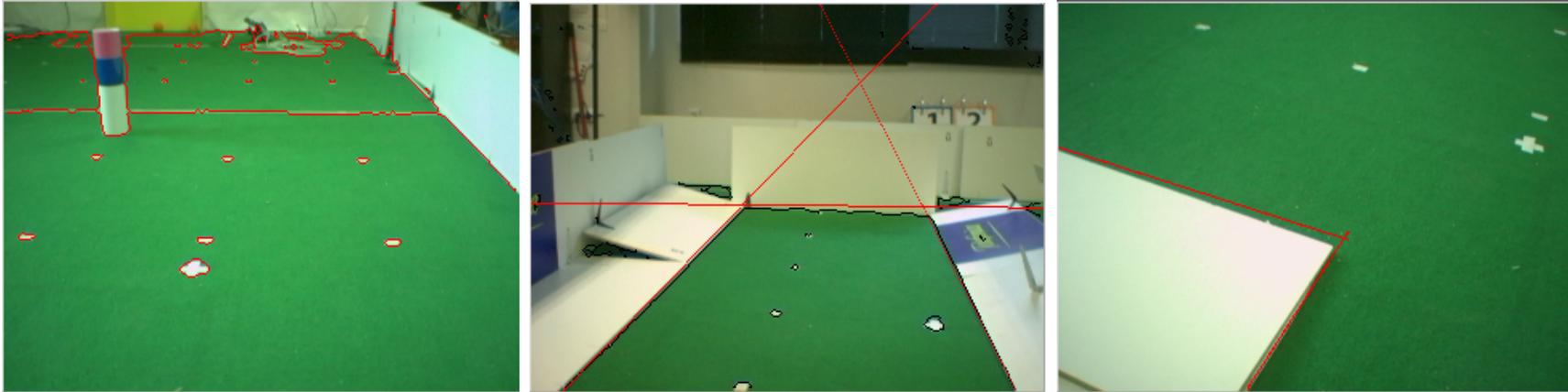


Con un par estéreo

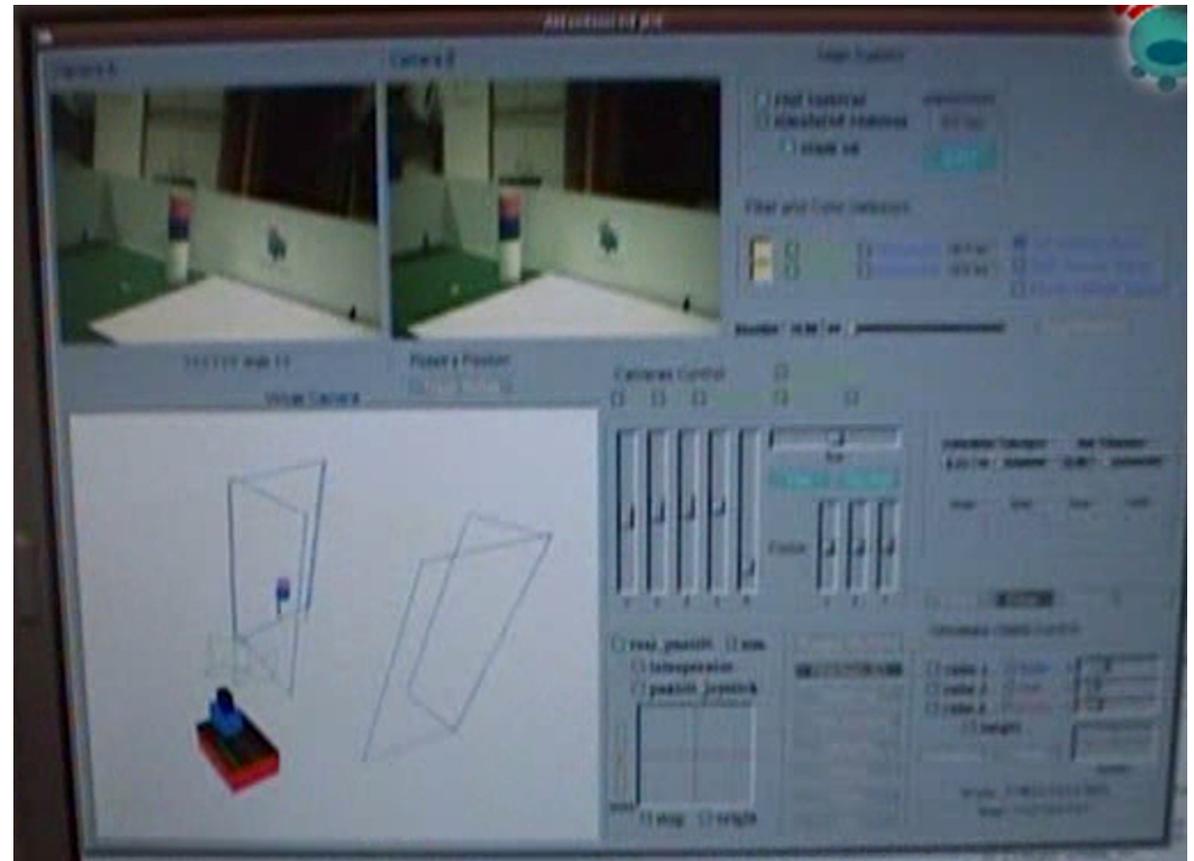
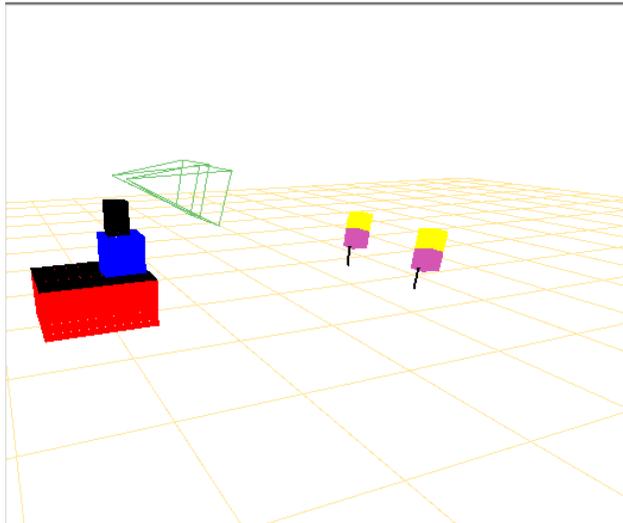
- Balizas de colores y bordes del suelo
- ¿Dónde están? Correspondencia y triangulación
- Asociación entre memoria y observación instantánea, fusión

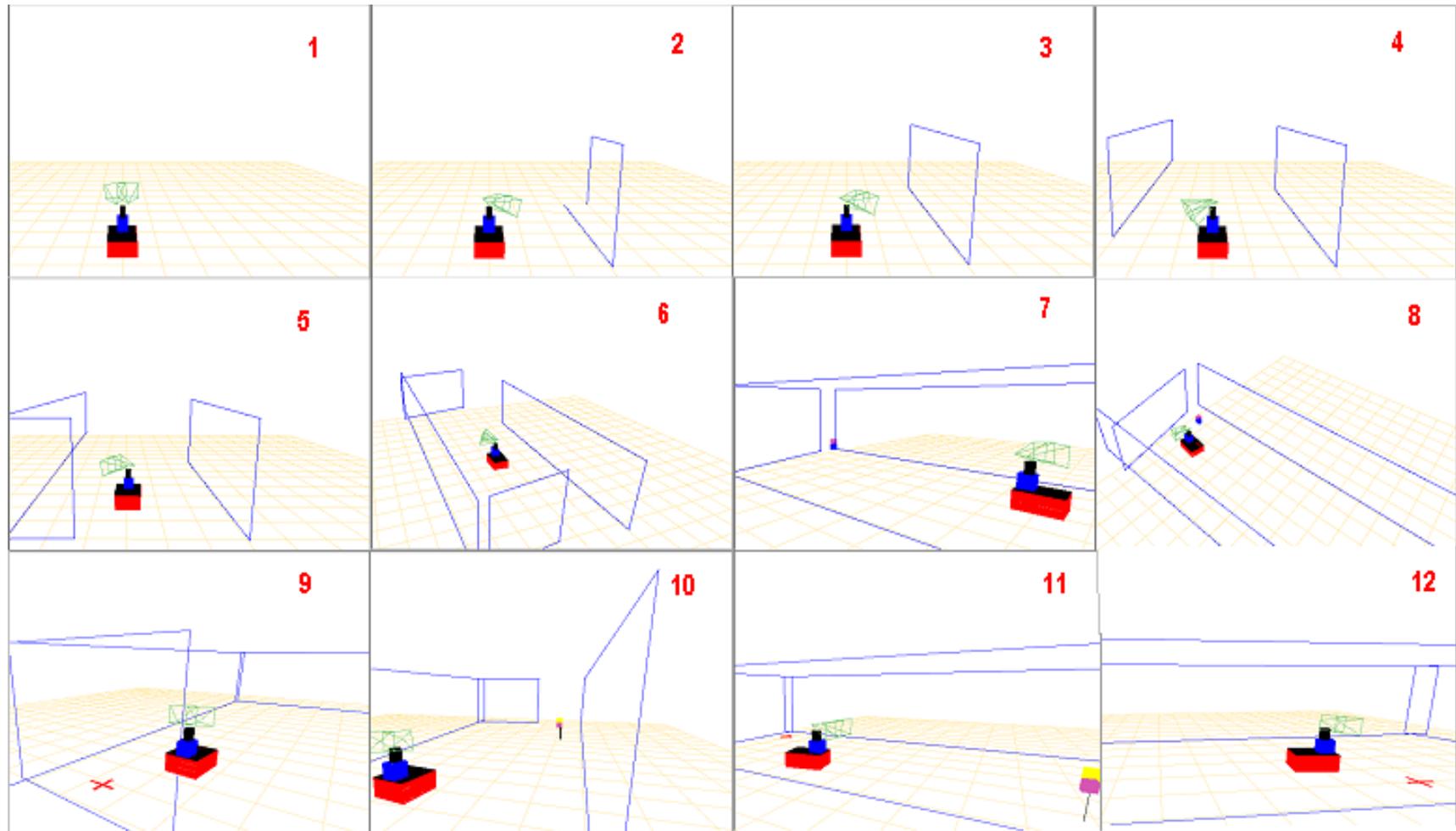


- Balizas: color, segmentación, punto central
- Balizas: asociación por distancia en 3D



- Suelo: bordes, Hough y soporte, “correspondencia”
- Suelo: líneas en 3D generan varios puntos de atención
- Suelo: asociación permite extender líneas





Conclusiones

- Escena rica 3D, dinámica y local, de los alrededores del robot
- Localización: filtro de partículas y algoritmos genéticos
- Reconstrucción: hipótesis suelo y genéticos
- Fusión sensorial para detectar objetos cotidianos
- Atención visual: dinámicas de saliencia y vida

- Líneas futuras: mejorar reconstrucción, integrarla con atención