
Sistema de atención visual para la interacción persona-robot

Julio Vega, Jose M^a Cañas
julio.vega@urjc.es, jmplaza@gsyc.es



Grupo de Robótica, 29 Septiembre 2009

Contenidos

1. Introducción
2. Sistema de atención
3. Implementación software
4. Experimentos
5. Conclusiones

1. Introducción

- **Sistemas de visión**, de los sensores más usado en robótica autónoma
- La dificultad está en **extraer información** útil de las imágenes
- **Robots sociales** interactuando con humanos



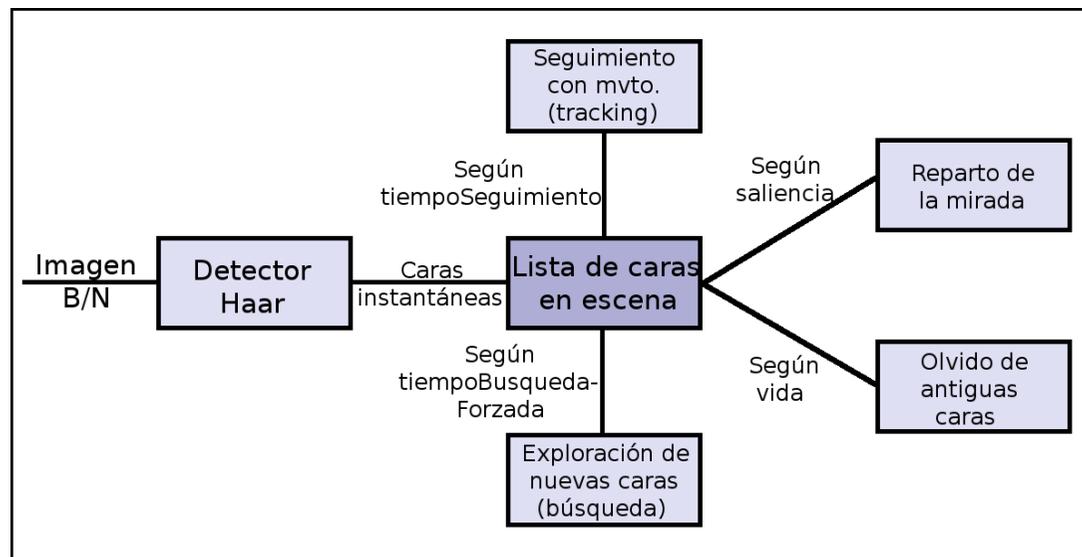
- La interacción básica entre humanos es *mirarse a la cara*

Atención visual en robots

- **Fijación** en uno o varios aspectos interesantes de la realidad y prescindir de los restantes
- Las cámaras de los robots proveen de un **amplio flujo de datos**
- Dos **vertientes** de atención:
 - Local** Seleccionar *dentro de una imagen* lo que interesa
 - Global** Seleccionar *del entorno* lo que interesa y dirigir la mirada hacia ellos

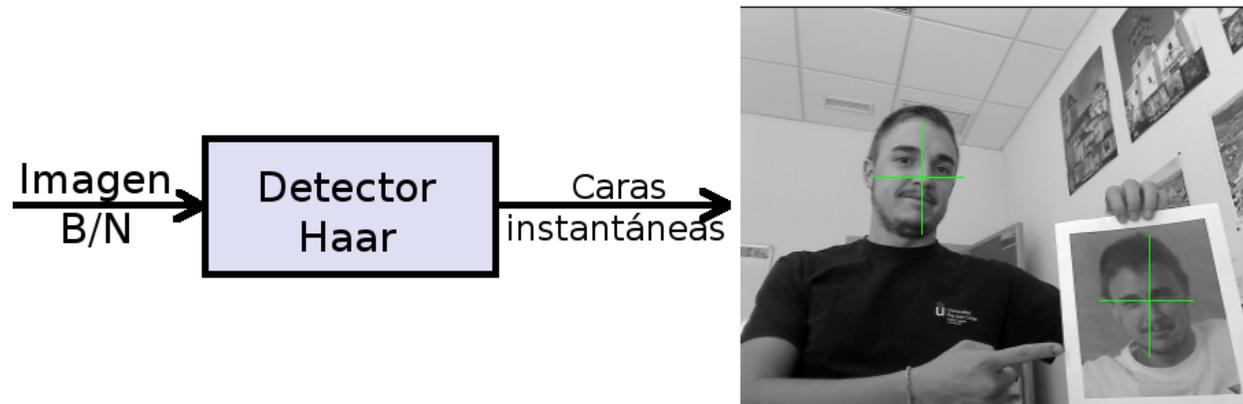
2. Sistema de atención

- **Objetivo:** realizar un seguimiento visual de las caras de las múltiples personas en escena



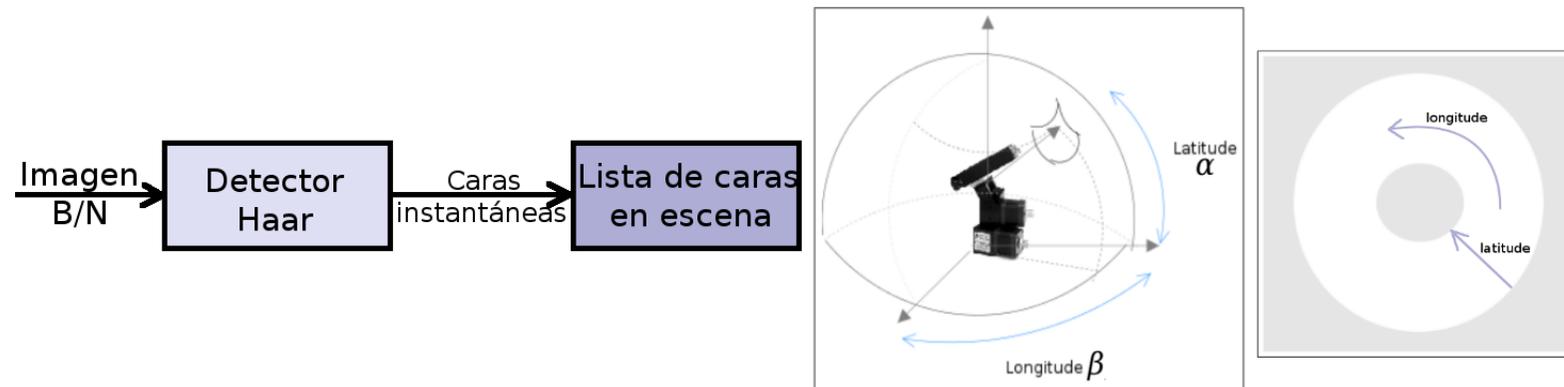
- Mecanismo *bottom-up*: los objetos guían los movimientos de la cámara
- Mecanismo *top-down*: los objetos relevantes tienen apariencia humana

1. Detección de caras



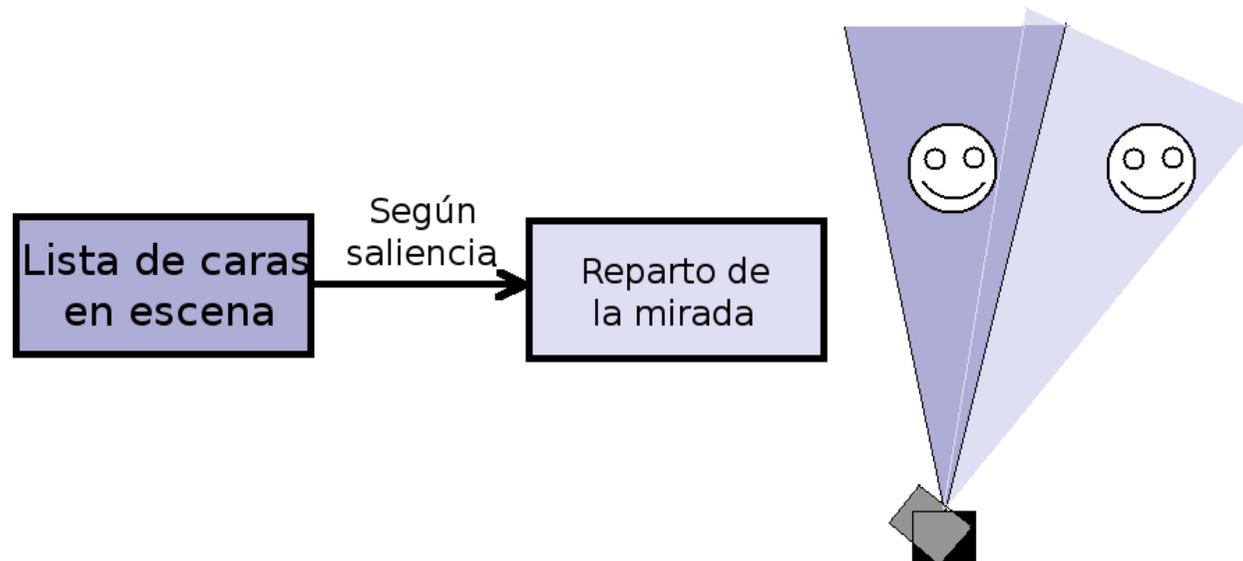
- Detecta las caras humanas existentes en la imagen actual
- Da como salida las posiciones de las caras dentro de la imagen
- Basado en filtros Haar en cascada *AdaBoost*
- Biblioteca *OpenCV*

2. Representación de la escena



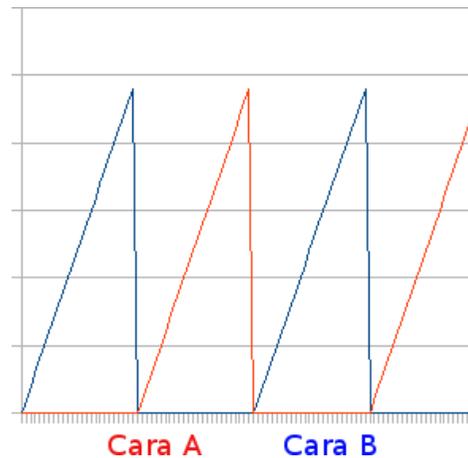
- Ampliado el **campo de visión** de la cámara
- El cuello mecánico sobre el que está permite moverla a voluntad
- **Transformación de coordenadas** de la imagen a coordenadas en la escena
- Y de coordenadas en la escena a **comandos** en el cuello mecánico

3. Reparto de la mirada



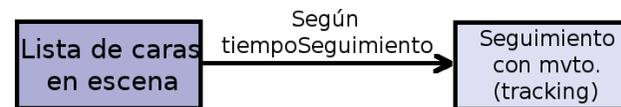
- Movimiento del cuello mecánico para **dirigir** la atención
- Con varias caras hay que decidir a **dónde mirar** después
- Cada cara tiene asociado un foco de atención y una saliencia
- El **punto de atención** es la posición de la cara en escena

Saliencia

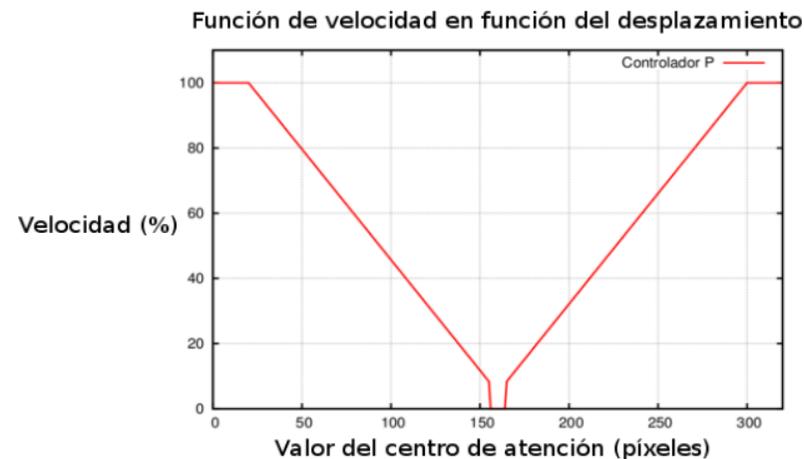


- Es todo aquello que *llama la atención* según el instante
- Valor que *augmenta* con el paso del tiempo
 - El sistema dirige la atención hacia la cara con **máxima saliencia**
- Se anula cada vez que se visita tal cara (**inhibición de retorno**)

4. Seguimiento con movimiento



- Cuando el sistema elige una cara la *sigue* espacialmente un tiempo
- **Controlador P** para conseguir un movimiento suave del cuello mecánico



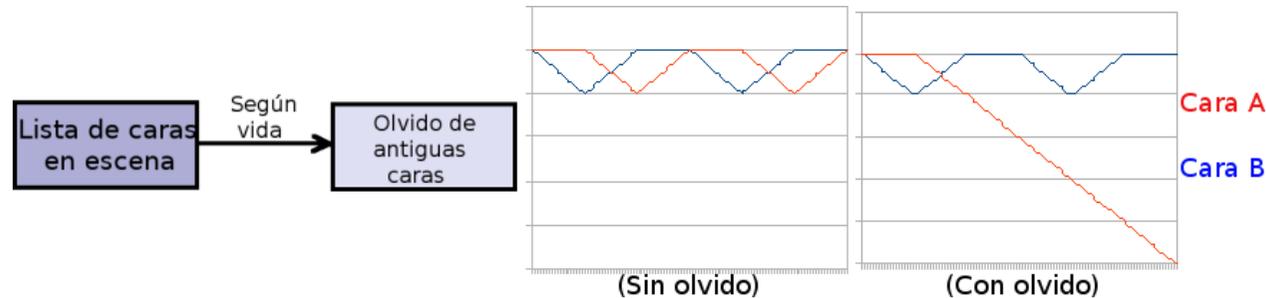
- Mantiene la cara objetivo en el centro de la imagen

5. Exploración de nuevas caras



- En cualquier momento puede interesar la **búsqueda** de nuevas caras
- Inserta en memoria local **puntos de exploración** con alta saliencia
- Dos tipos: aleatorios y de recorrido
- Son visitados rápidamente para ver si existen caras
 - Si es así, entran en memoria para el reparto de mirada
- El tiempo entre búsquedas es *proporcional* al número de caras

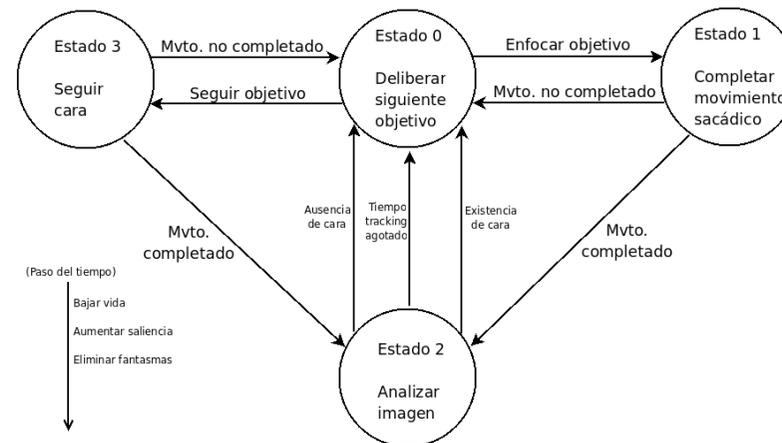
6. Representación interna del entorno. Vida



- Las personas se mueven y eventualmente **desaparecen** de la escena
- El sistema ha de *olvidar* antiguas caras que ya no están
- **Vida**: dinámica inversa a la saliencia
- Cada vez que se visita un objeto su vida se incrementa un poco
 Un objeto frecuentemente visitado tiene mayor vida
 Límite máximo para evitar saturación
- Si la vida de una cara es inferior a un umbral es que ha *desaparecido*

3. Implementación software

- Componente en JdeRobot (*iFollowFace*) con máquina de estados



- Estado 0:** el sistema se pregunta si hay algún objetivo a mirar
- Estado 1:** se completa el movimiento a la posición indicada
- Estado 2:** analizar imagen para ver si hay caras o no
- Estado 3:** seguir cara detectada

4. Experimentos



- Seguimiento de una cara (*Ver vídeo de seguimiento*)
- Seguimiento de varias caras (*Ver vídeo de reparto de mirada*)

5. Conclusiones

- Sistema **atención visual global** busca y sigue **caras humanas**
 - Campo visual de la escena mayor que el instantáneo
 - Reparte la mirada si hay varios interlocutores
 - Representación coherente *olvidando* las que ya no están
- Mecanismo de dinámica concurrente entre **vida** y **saliencia**
- **Memoria local** de corto plazo como campo de visión de toda la escena

Trabajos futuros

- Detectar **emociones**
- **Obstáculos** de navegación también como focos de atención