

Tesis: PERCEPCIÓN VISUAL ATENTIVA EN ROBOTS MÓVILES AUTÓNOMOS

Doctorando: **Julio Manuel Vega Pérez**

Director: **Jose María Cañas Plaza**

1. Introducción

La *evolución biológica* ha producido organismos que son capaces de extraer un conocimiento muy preciso del entorno exterior, en tiempo real, a partir de secuencias de imágenes. Desde el punto de vista de procesamiento de información en máquinas este problema resulta de una gran complejidad. Así, la visión artificial es todo un reto para la ciencia actual. No hay máquina, por potente que sea, que se acerque a la capacidad visual de los organismos que existen en la naturaleza. La *percepción* es una de las facultades más sorprendentes de los seres vivos: el sistema nervioso es capaz de interpretar información “en crudo” y convertirla en conceptos abstractos.

La *visión computacional* y la *atención visual* son hoy en día uno de los campos más atractivos e interesantes en los que la comunidad de investigación robótica dedica tiempo, esfuerzo y dinero. Las cámaras, cual ojos en el ser humano, proveen de información sensorial importantísima sobre el entorno a los robots; el problema está en que el hecho de extraer información de una imagen no es trivial, ya que hay que distinguir las partes importantes e interesantes de la imagen dependiendo de la tarea en curso. Para ello se utilizan sistemas de atención visual que nos ayudan a filtrar lo relevante. Esta capacidad nos permite desarrollar aplicaciones prácticas que, consumiendo menos recursos, procesan inteligentemente las imágenes percibidas. De hecho, los robots más avanzados hoy en día usan sólo la visión como sensor externo.

2. Objetivo

El objetivo de la tesis propuesta es diseñar, programar y validar en un robot móvil un sistema de *percepción visual* que le permita desarrollar tareas complejas como explorar el entorno que le rodea, buscar zonas de interés, reconocer objetos, construir mapas cognitivos o interactuar con humanos. Las plataformas robóticas reales en las que validaremos nuestra investigación son el robot humanoide *Nao* (de *Aldebaran Robotics*) y la base *Pioneer-2DX*, utilizando cámaras móviles como sensor principal.

Este objetivo general lo articulamos en varios puntos concretos que se explorarán en la tesis doctoral cuyo proyecto se propone.

2.1. Atención visual y reparto de mirada

Con objeto de dotar al sistema de capacidad multitarea, debemos desarrollar soluciones para compartir el control de la mirada entre las tareas concurrentes, ya que el único sensor que empleará será la visión. Así, el sistema de atención visual se basa en el concepto de *saliencia* para el reparto de mirada sobre los objetos relevantes en la imagen. Este sistema atento analiza la imagen de una forma similar a como lo hace un ojo humano, fijándose en los objetos que suelen llamar la atención (colores llamativos, bordes, movimiento o conocimiento previo de los objetos en la realidad) mediante *movimientos sacádicos* de las cámaras.

Desarrollaremos mecanismos de atención ascendente (o *bottom-up*), en el que los objetos del entorno del robot guían los movimientos de la cámara; y descendente (o *top-down*), en el que los objetos considerados relevantes dependerán de la tarea a realizar. Por otro lado, exploraremos las dos vertientes de atención visual, la global (*overt attention*) y la local (*covert attention*). La atención local (*Itti, 01*), (*Marocco, 02*) consiste en seleccionar dentro de una imagen aquellos datos que nos interesan. Y la atención global consiste en seleccionar del entorno que rodea al robot, más allá del campo visual actual, aquellos objetos que interesan, y dirigir la mirada hacia ellos.

2.2. Memorial visual

Otro punto esencial del sistema será la *memoria visual*, inicialmente de corto plazo, que le permitirá al robot organizar y estructurar la información que vaya recogiendo del entorno, más allá del campo visual instantáneo.

Esta memoria es la noción idónea para integrar observaciones parciales de la escena, y fusionarlas de cara a localizar e identificar objetos estructurados (cuya representación 3D selectiva estará guiada según la tarea a realizar), generar hipótesis perceptivas del mundo en 3D, así como definir puntos de atención para validarlas.

La representación visual de los objetos interesantes en los alrededores del robot puede mejorar la calidad del comportamiento del robot, así como la posibilidad de manejar más información a la hora de tomar sus decisiones. Esto plantea un problema cuando esos objetos no se encuentran en el campo de visión inmediato. Para solventar este inconveniente, en algunos trabajos se emplea visión omnidireccional; en otros, se utiliza una cámara normal y un mecanismo de atención global (Itti, 01), (Zaharescu, 05), que permite -de forma rápida-tomar muestras de un área de interés muy amplio.

2.3. Interacción persona-robot

Los objetivos anteriores aunan sus resultados para crear y mantener mapas cognitivos, reconocer objetos estructurados y no tan estructurados. Así el robot puede navegar a la vez que incorpora a su memoria visual el estímulo *persona*. Esta última habilidad es esencial para introducir al robot en el ámbito más cercano del hombre, permitiendo una interacción más fluida con las personas, pues éstas perciben más natural al interlocutor robótico. A tales robots se les puede denominar *sociales*. Este acercamiento implica que los robots tengan determinadas características que el hombre considera necesarias para aceptarlo.

Es, por tanto, interesante desarrollar técnicas que permitan al robot conocer la posición de las personas que están alrededor de él, y seguir las en todo momento. La primitiva más básica de interacción entre humanos es *mirarse a la cara*. Así, para una interacción *robot-humano* aceptable resulta indispensable un mecanismo de detección y seguimiento de caras (Lang, 03) que se podría traducir, por ejemplo, en el seguimiento de ancianos para monitorizarlos.

2.4. Arquitectura distribuida

El sistema perceptivo visual descrito se incorporará a la arquitectura de comportamientos que integra al resto de subsistemas del robot y sus funcionalidades: navegación, autolocalización, etc.

El diseño de este tipo de sistemas es un proceso software complejo, puesto que opera de forma distribuida y además en un contexto de tiempo real. Por todo ello, en este proyecto nos basamos en la utilización del paradigma de programación orientado a componentes que facilite el diseño, programación y validación del sistema. Para nuestro propósito emplearemos la plataforma *JdeRobot* (Cañas, 03), como gestor de la jerarquía dinámica de esquemas presentes en el comportamiento del robot.

Referencias

- Cañas, J.M., “Jerarquía dinámica de esquemas para la generación de comportamiento autónomo”, PhD, Universidad Politécnica de Madrid, 2003.
- Itti, L., Koch, C., “Computational Modelling of Visual Attention”, in Nature Reviews Neuroscience 2, pp. 194-203, 2001.
- Lang, S., Kleinehagenbrock, M. et. al. 2003. Providing the Basis for Human-Robot-Interaction: A Multi-Modal Attention System for a Mobile Robot. Bielefeld University, Faculty of Technology, Bielefeld, Germany.
- Marocco, D., Floreano, D. 2002. Active vision and feature selection in evolutionary behavioral systems. In Proc. of Int. Conf. on Simulation of Adaptive Behavior (SAB-7), pp. 247-255.
- Yoshikawa, Y. et. al. 2006. Responsive robot gaze to interaction partner. Intelligent Robotics and Communication Laboratories. Advanced Telecommunication Research Institute International. Keihanna Science City, Kyoto, 619-0288 Japan.
- Zaharescu, A., Rothenstein, A.L., Tsotsos, J.K. 2005. Towards a biologically plausible active visual search model. In Proc. of Int. Workshop on Attention and Performance in Computational Vision, WAPCV-2004, Springer LNCS 3368, pp. 133-147.